INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Empirical Tests of Lerdahl and Jackendoff's (1983) Low-Level Group Preference Rules for the Parsing of Melody

by

Bradley W. Frankland

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July 27, 1998

© Copyright by Bradley W. Frankland, 1998

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-36580-8

Canada

DALHOUSIE UNIVERSITY

FACULTY OF GRADUATE STUDIES

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Empirical Assessment of Lerdahl and Jackendoff's Low-Level Group Preference Rules"

by Bradley W. Frankland

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: August 20, 1998

External Examiner

Research Supervisor

Examining Committee

[redacted]

Dalhousie University

Date: July 27. 1998

Author: Bradley W. Frankland

Title: Empirical Tests of Lerdahl and Jackendoff's (1983) Low-Level Group Preference Rules for the Parsing of Melody

Department or School: Psychology

Degree: Doctor of Philosphy Convocation: Fall (October 17) Year: 1998

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

[redacted]

Signature of Author

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed of otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in this thesis (other than brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

Dedication

This work is dedicated to those individuals, who by dint of bottom-dollar economics, political expediencies or social strata, never had the chance to attempt their potential.

Table of Contents

List of Tables

List of Figures

Abstract

The mechanisms that listeners employ to parse melody are not yet well understood. Lerdahl and Jackendoff (1983) provided a comprehensive theory of music parsing that has received much attention and some empirical support. In the present work, two eight-stage experiments, using a new on-line assessment, were conducted to assess how listeners parse melody. In Part 1 of Stages 2, 4, 6 and 8 (Stage 2 served as practice), subjects listened to a melody and pressed a key any time they detected the end of one unit (or the start of a new unit). In each stage, subjects heard the same melody three times and produced a boundary profile for each repetition; Stage 4 used a familar nursery rhyme, Stages 6 and 8 used the same unfamilar tonal piece. In all stages (both experiments), boundary profiles were highly consistent across repetitions (familar: r=.77; unfamilar: r=.65). Subjects was also highly similar on the third repetition of each stage (familar: r=.67; unfamilar: r=.60). Subjects placed boundaries at the same points in both experiments (r>.90). To model the data, Group Preference Rules 2 and 3 of Lerdarhl and Jackendoff (1983) were quantified and then compared to the empirically determined boundaries. Results indicated that Rules 2b (Attack-Point), 3a (Register) and 3d (Length) had predictive validity. Rule 2b was the most important in both the familar and unfamilar melodies. The theory of Lerdahl and Jackendoff was extended to Rule 4 (Intensification) and Rule 6 (Parallelism). Rule 4 was the optimal combination of Rules 2 and 3. A new method of analysis, based on asymmetric error terms and non-linear regression, was developed to properly assess the contributions of the different rules. Rule 4 demonstrated that only Rule 2b was necessary to predict boundaries; Rule 3a and 3d played a much less important role. Rule 6, Parallelism, was quantified as two forms of Pitch Pattern Parallelism (pitch pattern including duration and pitch pattern ignoring duration) and two forms of Time Pattern Parallelism (time pattern for the veridical timing of breaks between notes and time pattern for the encoded representation of durations). These constructions go beyond those defined by Lerdahl & Jackendoff. All versions of parallelism added some predictability for at least one melody, but only the first Pitch Pattern and the second Time Pattern Parallelisms do so in both. In Experiment 1, Part 2 of Stages 2, 4, 6, and 8 consisted of a memory recognition task for extracts from the melody. In Experiment 2, Part 2 consisted of a non-memory, click-detection task. Both secondary tasks (boundary efficacy) indicated that boundaries did affect subsequent processing, although the effect was not very pronounced. Finally, Stages 1, 3, 5 and 7 assessed the sensitivity of subject to tonality using a modified probe-tone task. Several detailed analysis relating the tonality profile to the boundary profile, and to the the boundary efficacy tasks, indicated that the tonality profile was only minimally related to boundary profiles or the boundary efficacy tasks. Overall, the present work indicated the effectiveness of the new on-line boundary assessment task. The present work also revealed that the model of Lerdahl and Jackendoff's could partially predict melody parsing. However, the model of Lerdahl and Jackendoff could be refined and extended, particularly in the implimentations of Rules 4 and 6, possibly using the extensions that were developed in this work. In addition, boundary efficacy tests need to be refined if they are to be used within naturalist studies.

List of Abbreviations and Symbols

EEG: Electroencephalogram

ERP: Electric Related Potential

fMRI: Functional Magnetic Resonance Imaging

GPR: Group Preference Rule

MIDI: Musical Instrument Digital Interface

MRI: Magnetic Resonance Imaging

PET: Positron Emission Tomography

RCM: Royal College of Music

SI: Systeme Internationale

SPSS: Statistical Package for the Social Sciences

Acknowledgements

Dr. Annabel J. Cohen	Supervisor
Dr. Raymond M. Klein	Internal Supervisor
Dr. John C. Fentress	Reader
Dr. Dennis P. Phillips	Reader
Dr. Isabelle Peretz	External Examiner
Dr. Micheal Sullivan	Departmental Representative
Betty Bailey	Informal Reader
Carol Thompson	Friend

CHAPTER 1

Introduction

This work is concerned with the question of how listeners parse the stream of acoustic events that is called music. Although the goal is easily stated, it was not easily achieved. Implicit within that first statement are two fundamental questions that must be addressed before progress can be made towards the goal. The first fundamental question is, what is music? This is not as trivial as it might seem. One research endeavor cannot possibly address the breadth of that which is called music. Conversely, analysis, interpretations and conclusions will necessarily represent the sample chosen. As such, a representative sample must be selected because conclusions will necessarily reflect the sample chosen. The second fundamental question is, what is parsing? Parsing is usually considered as the process of dividing a complex pattern into meaningful units (e.g., parsing a sentence, parsing a visual scene, parsing a song). The parsing produced determines and reflects the structure of available information, subsequent processing and the content of memory. Parsing, in turn, depends on the current structure of long-term information (musical knowledge) and the limits of perceptual (auditory) and working memory. Implicit, but often forgotten within this concept of parsing is the concept of meaning. That is, parsing is intended to create meaningful units, but what is meaning in music? Of course, parsing, and therefore meaning, interact with the samples chosen to be representative.

Broadly speaking, the goal of this work is the empirical assessment of the parsing of music (or the assessment of boundary locations between units). Essentially, subjects were asked to identify the subunits of music by identifying the locations between units. However, this work is not limited to a simple description of that parsing (boundary locations). This work also models the parsing of music, relating the observed boundaries to theoretical notions of where boundaries should occur. In fact, in a field as complex as music, it is possibly fruitless to try to separate the description of parsing from the modeling of parsing because the terms used to describe empirical data will necessarily evoke theoretical notions of the causes, even the existence, of boundaries. That is, an attempt to describe empirical boundary locations in terms of music-theoretic notions of tonality, harmony, rhythm, meter and the like implicitly assumes that the concepts represented by those terms are valid[1]. Conversely, an attempt to describe empirical data in terms of conceptually neutral terms might easily become bogged down in the details of finding and defining the important concepts and terms -- concepts and terms, one could argue, that have already been provided by music theory. Hence, before experiments can be designed, it is important to consider the global, theoretical framework of those experiments.

The goal of this introduction is to explain the framework that has guided the design of the empirical tests. This introduction can be considered as three main sections:

1) Defining music
2) Defining meaning or information in music
    - A discussion of possible approaches to the analysis of meaning or information in music
    - Choosing (or defining) the theoretical framework that will be used in this thesis.
3) A general discussion of theoretic concepts and empirical findings relevant to the question of parsing of music.

This introduction is intended as a prelude to the more detailed analysis of the basic theoretical model that will appear in the next chapter. As such, this introduction explains why that particular model was chosen. This introduction also discusses the basic conceptual constructs that underlie the model and the relationship between the use of those constructs within the model and the use of similar constructs by others in the field (music theory and music cognition).

Defining Music for the Purpose of this Thesis

That which is called music demonstrates considerable variation across cultures and subcultures, across historical periods, in complexity, and in production. Given this

---

[1] In this work, I assume some familiarity with the basic terms of music. To do otherwise would expand this work beyond all reason. However, I acknowledge that many terms in music are ambiguously defined. When a single working definition is required, it will be provided.

breadth, what is it that distinguishes music from other acoustic events? This is important for the study of music because when selecting a sample, one must "know" that the sample is music (i.e., that the stimuli contain the features of music). Not only must one know that the sample is music, but one must also know that the sample is actually heard as music (and not simply as a stream of acoustic events; cf., Kendall, 1986). On the other hand, given the exploratory nature of this work, it was also important to strive for simplicity. Hence, for the purpose of this thesis, music was operationally defined as a single-line melody -- that which is often called a tune -- within a corpus defined as western-tonal music.

The single-line melody was considered fundamental because all music involves the sequential presentation of notes. Admittedly, some music involves the simultaneous presentation of notes or other acoustic events. Even though all music consists of the sequential presentation of notes (again, allowing for the presentation of other note-events, such as rests), not every sequential stream of notes or sounds can be considered music. The single-line melody seems to offer the desired balance between necessary musicality and sufficient simplicity[2]. While it is true that many of the perceptual and/or cognitive procedures inherent in the processing of much of western-tonal music are not realized in such a basic stimulus (e.g., simultaneous harmony), such a reduction is not reductio ad absurdum. A simple melody contains many of the cognitive processing dimensions normally applied to more complex music. Processing a single line of notes involves the processing of meter, rhythm, tonality, harmony (e.g., harmonic motion), and contour pattern. As such, the basic melody can be conceived as being safely across the border between that which is music and that which is just a sequence of sounds. Basic melodies contain the structures of music, while constructions below the melody (e.g., random excerpts, experimental idealizations of structure, strings of notes, chords in isolation) *may or may not* contain the structures of music (e.g., even an excerpt may fail to provide a necessary tonal center; cf., Watkins & Dyson, 1985).

[2] The importance of simplicity will become more apparent in later chapters when the modeling of the parsing of music is outlined.

Operationally defining music as a melody has other bases for validity. Simple melodies have primacy in learning (nursery rhymes, etc.) and simple melodies are what individuals can hum, whistle or sing, alone or in company. In addition, some instruments are limited to a single melody line and the learning of instruments often initially focuses upon the learning of simple melody lines.

Although it is necessary, it is not sufficient that the experimental stimuli contain the structures of music. Research on the parsing of music must also insure that the stimuli are processed as music. Again, the single line melody seems to meet such a criteria. That is, if one presents a melody to subjects and then assesses musical cognition on the basis of that processing, one may be reasonably certain that a melody was processed as music. However, if one uses stimuli that are "below" the melody (e.g., random sequences of notes, sequence designed for a particular experimental purpose), one should divert resources to insure that those stimuli are processed as music (e.g., providing musical context surrounding the experimental task as in Deliege, 1987) before one accepts any inferences about music cognition. Shepard (1982a, 1982b) makes similar comments pertaining to the delineation between music cognition from psychophysics (cf., Kendall, 1986). That is, stimuli below the melody may or may not be processed as music[3]. An alternative view of the same issue would argue that the definition (or general acceptance) of a melody as music is external to the propose of any study: However, the definition (or acceptance) of stimuli below the melody as music is subject to the constraints of the study using those stimuli.

It is important that the stimuli both contain the elements of music and be processed as music because parsing implies meaning and meaning is defined, at least to some extent, by the context in which the stimulus is processed: Music-like stimuli may be processed as music, or music-like stimuli may be processed as another type of acoustic event (i.e., a door chime can be processed as music or as a reference to external events;

---

[3] There is a similar problem when presenting music across cultures, or even across subcultures. In addition, there may be melody-like structures that are not processed as music, structures that lie "beside" the melody.

advertising jingles can be music or a reference/label to a product). The reverse may also be true (e.g., whale songs or other natural sounds presented as music). Of course, the more the stimulus diverges from the notion of music, the more likely it is that the stimulus will not be processed as music. It is interesting to note that Brown and Dempster (1989, pp. 77-79, see also Note 47) consider the ability to make such distinctions (between music and non-music) to be a necessary prerequisite for all music theories, but it is a prerequisite that is not met by many theories[4].

The use of simple melodies should help to achieve the dual goal of stimuli that contain the elements of music and stimuli that are processed as music, while not overly complicating any subsequent analysis.

Defining Information for the Purpose of this Thesis

Using melodies as the basic stimulus alleviates the need to further define music, but it does not alleviate the need to define information. Parsing is the process of dividing a complex stimulus into meaningful units (also known as chunking, also known as boundary formation). In the discussion of parsing in music, meaning is often evoked. For example, when defining the phrase as a unit of music, Christ, DeLone and Winold (1975) identified the unit as a "relatively complete musical utterance, roughly comparable to a phrase or short sentence in language" (1975, p. 56; cf., Christ, DeLone, Kliewer et al., 1966). According to Green (1965):

> The phrase is the shortest passage of music, which having reached a point of relative repose, has expressed a more or less complete musical thought.
>
> 1965, p. 7

For Anders et al. (1975) the definition is that:

> The phrase is heard as a whole group of notes that belong together in one unit, expressing one or more musical ideas, with a break or pause at the end.
>
> 1975, p. 42

According to *The New Groves Dictionary of Music and Musicians* (Drabkin, 1980, p.

---

[4] Their critique was actually directed at Boretz' "Meta-variations", a music theory cognition based on the philosophical principles of logical positivism, but they cited others who had also erred on this point, particularly Rahn in "Logic, Set Theory, Music Theory".

648, vol 15), the motive (or motif) is "A short musical idea..." As the examples illustrate, parsing identifies meaningful units within a more complex stimulus, or equivalently, parsing identifies the boundaries between meaningful units; the distinction is really only one of focus. Dowling has commented, "What we perceive [in music] are meaningful, interpreted musical events, just as in speech we hear meaningful words and sentences and not a meaningless stream of sounds. . ." (1993, p. 8). A study of parsing in music implicitly involves the meaning -- the information content -- of music and its subunits.

The definition of meaning and/or information in music in western-tonal music, is equivocal. In the early stages of this work, I examined a large number of works to find a definition of meaning that could serve as the basis of subsequent theoretical models of musical parsing, but could find no consensus (e.g., Adorno, 1990; Agmon, 1993; Anders, et al., 1975; Brown and Dempster, 1989 [with replies by Boretz, 1989; Cook, 1989; Rahn, 1989 & Taruskin, 1989]; Christ, DeLone, Kliewer, et al., 1966; Christ, DeLone & Winold, 1975; Green, 1965; Handel, 1993; Kraut, 1992; Lerdahl & Jackendoff, 1983; Lloyd, 1968; Meyer, 1956, Raffman, 1992; Sloboda, 1992). Cross has commented that "the psychology of music has relied on music theory to provide an initial musical morphology which may, in some instances, cloud rather than clarify a cognitive view of the constituents of music" (1985, p. 3). The lack of consensus makes the study of music difficult because, although there is a lot of knowledge, insight, intuition and speculation, what defines meaning (information) in western-tonal music cannot be stated with certainty. Note that this is not a difficulty associated only with that which has been labeled as referential music (music that is about something): Absolute music (music that is not a reference to the natural world) has the same problem, possibly to a greater degree (cf., Christ, DeLone & Winold, 1975, p. 16; Kuhn, 1978; Meyer, 1956, pp. 1-6). Also note that this is different than the study of parsing in speech and language (cf., Handel, 1993, pp. 322-326); in language parsing, we do know what words, phrases, sentences and intonation mean, if not absolutely, then at least to a much higher degree than in music (cf., Kraut, p. 16; Lerdahl & Jackendoff, pp. 5-6, 112-113; for an alternative viewpoint, see Raffman or Sloboda; see also Clark & Clark, 1977). As such, the study of parsing in speech is not hampered by a lack of consensus as to the theoretical locations of the boundaries between the units of speech. That is, research on speech parsing can focus on the process of parsing (e.g., what acoustic features drive boundary detection) with relatively full knowledge of the location of boundaries. Research on music parsing must deal with the more ambiguous problem of a lack of knowledge about where boundaries are in conjunction with a lack knowledge about how boundaries are detected. (In this regard, the history or study of how music is studied might be most useful if humans should ever need to develop a true universal translator for encounters with alien races.)

The ambiguities have several repercussions for the psychological study of music and melody. Firstly, the differences between speech and language, on one hand, and music, on the other hand (hemispheric specialization is not implied), suggest that transference from psychological and/or philosophical studies of speech and language to music must be done judiciously (cf., Agmon, 1993). Secondly, ambiguities of what is music (what is musical meaning or information) will cloud issues pertaining to parsing. These ambiguities will also hamper the choice of methods to study this information and parsing. Since music theorists have not yet provided a consensus (cf., Meyer, 1956; Brown & Dempster, 1989), empirical psychologists do not have a single theoretical framework to provide clear directions for research. Lacking a clear concept of meaning or information in music, the researcher is forced into one of two camps: the researcher can adopt a purely atheoretical approach, or the researcher can adopt a particular theoretical view with full knowledge that it may be incorrect (or only partially correct).

It is possible to study the parsing of music without knowing the meaning of music (i.e., atheoretically). In this scenario, one must be certain that subjects process the stimulus as music, and then one uses the results to establish meaning (explicitly or implicitly). The weakness of this approach is that as soon as one tries to explain the results, one must move into a theoretically laden explanation (even if only for the terminology). Even neural network models do not escape the need for theoretically laden constructs (i.e., see Bharucha, 1996 and Katz, 1995, for examples that apply to cadences). All such models require the designer to choose the stimulus dimensions a priori (i.e., either as inputs to layers or as layers themselves). That decision reflects the beliefs of the designer and not necessarily the realities of perception or cognition. For example. Bharucha's (1987a, see also 1987b. 1992) MUSACT model assumes that listeners perceive notes. chords and keys (many would accept this). and each of these percepts is instantiated in a layer in the network (fewer might accept). To be tractable. such models often impose other constraints that further reflect the belief system of the designer: that is. the types of connections. which layers are interconnected and equally important. which layers are not connected. For example. in Bharucha's model. keys are connected to chords and chords are connected to tones (many would accept). but there are no interconnections within a layer. between keys and between tones (few might accept; cf.. Krumhansl & Shepard. 1978; Brown. 1988). As such. neural network models are interesting examples of what might be. but they do not stand as statements or proofs of what is. either neurologically (the link between the underlying physiology and various neural nets is problematic at best; the link is at a level of abstraction). perceptually or cognitively.

As the example of neural network modeling illustrates. it is not possible to model music without evoking some theoretically-laden constructs. These constructs represent. at the very least. some rudimentary statement of the conception of the meaning of music (broadly defined). One cannot predict a boundary without a concept of why there should be a boundary (hence, the meaning of the boundary) and therefore, one cannot test predictions without meaning.

A Theoretical Framework for the Analysis of Melody

It seems clear that some theoretical framework must be adopted. but ideally. one that is not too limiting with respect to both meaning and the goals of this work (the experimental assessment of parsing and the modeling of parsing). Because the basic question concerns the information content of a melody (i.e.. what makes a sequence of notes a melody), the simplest perspective is. of course. that a melody is a rule-governed sequence of sounds. This of course, begs the question, what does rule-governed mean? The answer is far more complex than the typical answers one might have to such a question[5].

The conceptual framework that will be adopted here is that rule-governed means that information-processing algorithms can be applied to a stimulus (the sequence of notes or acoustic events used in music) and the information that results from that application can be related to a corpus or body of similar information. That is, there is a set of rules that define the allowable musical events and the typical relationships between those musical events. These rules are taken from (apply to) a particular corpus of music (western-tonal music). Broadly speaking, all music within that domain (western-tonal) follows the same set of rules. For a sequence to be heard as music within that corpus, it must follow those rules (allowing, of course, for exceptions and the evolution of the corpus). Sequences that do not follow these rules are not heard as music within that corpus. Ultimately, the various structures (or subunits) of that corpus, and therefore sequences from that corpus, have meaning, possibly referential meaning, and that meaning will be a determinant of behavior. Behavior may be limited to affective arousal or cognitive contemplation (i.e., even a behaviorist could assess the physiological changes that are associated with arousal or monitor changes in brain functioning using advanced neural imaging techniques like MRI, fMRI, and PET).

Several simple rules of western tonal music have already been empirically demonstrated. The simplest example is the limitation of selection of notes (pitches) to logarithmically spaced frequencies (12 to an octave). Furthermore, within a single piece of music, there is some variation in the relative use (frequency or duration) of each of the 12 notes within an octave. For the corpus of western tonal music (for a particular piece),

---

[5] An infinite number of conceptual frameworks is likely possible and several have been defined (cf., Brown & Dempster, 1989; Cook, 1989). The chosen framework is clearly within the information processing perspective of cognitive science while leaving open the possibility or definition of "meaning" in music (even allowing that music is a form of affective processing) and retaining the uniqueness of individual pieces. As such, it is an attempt to address and account for all the different perspectives, but it is certainly not a "new" perspective (cf., Agmon, 1990).

that variation can be predicted statistically (cf., Krumhansl, 1990)[6]. Similar rules have been shown to apply to the note-to-note transitions within a piece. For example Vos and Troost (1989) have shown that, in western music, such transitions obey statistical rules: Small intervals are predominately associated with descending steps and large intervals are predominately associated with ascending steps. Probabilistic rules are implied in many theories of music (cf., Lerdahl & Jackendoff, 1983; Meyer, 1956; Narmour, 1991, 1992; Schenker, as cited in Jonas, 1982; Watkins & Dyson, 1985), though often rules are couched in terms like "preference", "expects", "may", "imply". In these examples, one can think of the rules as reducing the uncertainty associated with the selection of notes at each point in a melody. Herein, the search is for rules that define the unitization of a piece (the locations of the boundaries between units).

This conception has the advantage of separating internal meaning (how sequences of notes relate to each other within the corpus of western-tonal music) from external meaning (consequences for human behavior). This separation is likely somewhat artificial or imperfect or incomplete (different critiques use different terms) but it allows parsing to be studied without complete knowledge of the meaning of music. That is, the rules can be used to define a "meaningful" parsing for any sequence that belongs to the corpus even if one does not know, in an absolute sense, what the resultant units "mean" in terms of behavior. Of course, this scheme assumes that behavior and meaning are somehow reflected in regular (rule-governed) structures for all the stimuli within that domain. Fortunately, this does seem to be the case. For example, a common phrase construction within western tonal music involves an harmonic motion from the area of the tonic, to the area of the fifth and then back to the tonic. This construction is associated with the

---

[6] However, the statistical rules should not be applied blindly. For example, the statistical use of the 12 notes within an octave has drifted over time. In the Baroque era, the probabilistic rules governing the selection of notes were heavily weighted to the tonic triad notes and the remaining diatonic notes. More recent classical music has de-emphasized the roles of these notes (cf., Christ, DeLone, & Winold, 1975; Cross, 1985).

building and release of tension[7] (among other things) and it is the return to the tonic that signals the end of the phrase (release of tension). This structure and its associated affect is generally considered a unit within music. The meaning of the unit is tension, but the rule that instantiates it within western-tonal music is the motion between the tonic and fifth (which is different from, for example, tension in an Indian Raga). The motion from the tonic to the fifth and back to the tonic might be useful as a rule that defines a unit without a precise definition of tension (hence, the meaning of the unit).

Adopting the framework that rule-governed structures define individual units -- units which, in turn, have meaning -- is essentially an assumption that common structures in music have common meanings. This framework neither implies nor requires a one-to-one mapping between structure and meaning. This framework does not even imply nor require a one-to-many or a many-to-one mapping between structure and meaning. Of course, analysis within this framework will be successful to the extent that meaning is systematically related to structure. That is, experimental techniques founded on the assumption that there is a relationship between rule-governed structure and meaning will be more successful when there exists a one-to-one mapping. Unfortunately for the study of music parsing, a one-to-one mapping between structure and meaning does not seem to be true in music: For example, many structures are associated with the building and release of tension (e.g., tempo changes, instrumentation changes), and many such structures do not result in the perception of tension. Conversely, many "rules" seem to define the phrase (relative repose [see above], rests [absence of sound], longer notes, terminal or partial cadences, harmonic motion). Although the lack of a one-to-one mapping between meaning and structure complicates the picture, it is also true that, as stated by Christ, DeLone and Winold (1975, p. 56), "The fact that it is frequently possible to have more than one analysis of a portion of music is one of the reasons why musical analysis is such a fascinating study." The fact that many structures may have the same meaning (and vice versa) makes subsequent explorations more difficult but not

---

[7] Tension is one of the affective terms that is not clearly defined within music, but it has been assessed behaviorally (e.g., Krumhansl, 1996).

impossible if one is careful to understand the limitations of the approach. On the other hand, the only alternative to the chosen framework requires a clear definition of meaning in music. The chosen framework seems to be the lesser of two evils.

In summary, the chosen framework is the analysis of melody as rule-governed relationships that define both the possible set of musical events and the relationships between those musical events. It is assumed that the rules can reliably define structural subunits of a melody. It is assumed that the subunits so defined can be related to the meaning of music, even though that meaning is not defined herein.

Rule-Governed Structure

If one defines a melody as a rule-governed arrangement of notes, there are three other advantages relevant to parsing. Firstly, given the rule-governed structure definition of information, the analysis is not affected by the source of the rules. That is, the analysis is the same if the rules of music are innate or learned or a mixture of both. Secondly, rule-governed implies an algorithm. An algorithm can be employed by any information processing system and as such, it should be possible to render it in a computer program. This was a major goal of the present work. Thirdly and most critically, the processing of stimuli (that which can be rendered in an algorithm) is only relevant within the defined corpus of information. Hence, algorithms/rules do not need to extract all the information that is available within the stimulus. Algorithms/rules only need to extract the information that is relevant to the defined corpus *presupposing one knows what is called information within the corpus.*

The problem is that defining information (hence, meaning) as rule-governed relationships within a defined corpus is not the final step. To study the parsing of melody by reference to a set of rules, one must be able to analyze (at some level) the structure of the melody. That is, one must define the actual rules. As stated previously, this can be done empirically (post hoc) by examination of the structures commonly implicated by empirically determined boundaries or this can be done theoretically (a priori) by experimental tests of music-theoretic notions of parsing. It seems simple, but it is not because the processing of even a simple melody implicitly involves the simultaneous processing of information along several dimensions (e.g., pitch, intensity, timbre, time, meter, rhythm, tonality, harmony, contour). Furthermore, some dimensions may be irrelevant; some dimensions may be critical. Dowling has commented that the:

> cognition involved in a basic listening situation suggests a multiplicity of cognitive processes occurring simultaneously. Sometimes these processes are conscious, but at other times they are not. . . . some of those processes depend more on the knowledge the listener brings to the situation., and other processes depend more on what happens in the musical pattern itself.
>
> 1993, p. 6

Cuddy has echoed these thoughts:

> The perceptual structures and strategies underlying melody recognition may involve a number of independent, or, alternatively, interacting, features. Melodies convey a great deal of information on several perceptual dimensions -- notably, pitch, time, and loudness -- and this information is structured in an elaborate manner -- yielding, for example, perceptual structures for tonality, rhythm, and dynamics.
>
> 1993, p. 21

To add yet another layer of complication, complete consensus has not yet been achieved as to the nature of the (important) underlying dimensions or, in many cases, the definitions of the associated terms. Cuddy (1993) has further commented that:

> Despite the seeming effortlessness of melody recognition, analysis of the mental processes involved is surprisingly difficult. To begin the inquiry, we must first decide on the perceptual dimensions and the perceptual units to be studied. Here it may seem appropriate to look at the notes contained in the melody, and to conduct the analysis in terms of the physical characteristics of the notes (e.g., the frequency spectrum of each note). This approach, however, immediately runs into problems at the level of psychological analysis.
>
> 1993, pp. 19-20

Jones has made similar observations on a number of occasions (1976a, 1981a, 1981b, 1982, 1985). Some writers (e.g., Christ, DeLone, Kliewer, et al., 1966) have chosen to delineate those dimensions in terms of basic acoustic properties like frequency, intensity, spectral composition and time, or their associated psychophysical scalings of pitch, loudness, timbre, duration (or meter and, possibly, rhythm). Others prefer to analyze the melody in terms of higher, more abstract properties like rhythm and meter, tonality, harmony, and melodic pattern[2] (e.g., Anders, et al., 1975; Christ, DeLone & Winold, 1975; Green, 1965; Lerdahl & Jackendoff, 1983). True reductionists have tried to reduce analysis to the basics of frequency and time (e.g., Boretz, 1969, 1970 and Rahn, 1979 cited in Brown & Dempster, 1989), while still others have argued that one must consider all levels (e.g., Agmon, 190; Bharucha, 1996; Lerdahl & Jackendoff, 1983). Most writers seem to use whatever seems to work at the moment, moving between the various viewpoints as needed (cf., Lerdahl & Jackendoff, 1983; Meyer, 1956; Palmer & Krumhansl, 1987a; Schenker, as presented by Jonas, 1982), even if they emphasize one level. Many theories defy an absolute classification: For example, Schenker's (presented by Jonas, 1982) theory is loosely linked to natural forms, as in the overtone series, but only as an initial point while Lerdahl and Jackendoff (1983) explicitly attempt to move from the basic acoustic properties to the more abstract constructions of music.

To consider the problem from another perspective, it is the same sequence of note events that defines frequency (pitch), intensity (loudness), timbre (frequency composition), time, meter, rhythm, tonality, harmony and melodic pattern. It can be argued that frequency (or frequency composition), intensity and time are the fundamental acoustic properties, but it can also be argued that tonality, harmony, meter and rhythm are such radical reconfigurations of frequency, intensity and time that any discussion of music in terms of frequency, intensity and time is a ridiculous reduction (see Handel, 1993, p. 265 for similar comments). This is the premise behind the notion of conceptually-driven processing (cf., Shepard, 1982b). For example, even at the most basic level of the perception of notes, it has been argued that most individuals do not hear notes as particular acoustic frequencies. Rather, when listening to music, individuals have categorical perception such that all frequencies within a given range are heard as the same note (cf., Dowling, 1993; Handel, 1993, pp. 280-284, Shepard, 1982b, pp. 349-350).

Conceptually-driven processing adds other complications: The currently available information affects or constrains subsequent processing (i.e., expectancies), but unexpected events may cause a retrospective reinterpretation of prior information. To add a final layer of complication, it would seem that in music, the important information is often, but not always, carried by changes along the various dimensions (e.g., changes in pitch define contour or melodic pattern; the relative numbers or weights of each pitch defines tonality; changes in the proportions of pitches define key changes). Important information is also carried by the interactions between the dimensions: For example, the interaction of a final cadence, a V-I chord with meter can alter the perception from masculine to feminine ending (cf., Green, 1965, p. 9, 27; Lerdahl & Jackendoff, 1983, p. 29); syncopation can result from the misalignment of metrical stress and harmonic stress. Interactions between more basic dimensions may lead to the creation of higher dimensions: For example, pitch and meter can combine to create rhythm (cf., Palmer & Krumhansl, 1987a) or changes in time, intensity, pitch, dynamics and articulation can combine to create meter (cf., Lerdahl & Jackendoff, 1983). These complications can be contrasted with speech. In speech, each morpheme has a unique meaning (more accurately, a limited number of meanings). It is true the individual morphemes may serve simply to colour the meaning of other temporally close morphemes, but generally, the meanings of morphemes sum, and most researchers agree on the fundamental meanings of morphemes. (Admittedly, this is a simplification, but it is intended to highlight the differences between music and speech. See Lerdahl and Jackendoff, 1983, for similar comments throughout their text [e.g., p. 112-114]) Music does not work this way. For example, the chord c-e-g may serve as the final point of a song if the key is c-major, as an intermediate point of temporary repose, if the key is f-major, or as nothing in particular if the key is b’-major, or as a completely ridiculous, erroneous, appalling or astonishing point of intrusive captivation if the key is f$^{\#}$-major (cf., Brown & Dempster, 1989, p. 88). The interpretation will depend upon the analyst or the listener.

What is clear from this discussion is that the basic dimensions of music are not obvious. This makes the choice of dimensions for analysis difficult. To place this in context, the goal of this work is an analysis of the parsing of simple melodies and the creation of algorithms to predict the parsing of simple melodies. It is assumed that parsing is directed at the creation of meaningful units within those melodies. It is assumed that parsing can be analyzed by reference to the rule-governed relationships between the individual (acoustic) events that define the melody. However, the lack of a consensus as to the important dimensions of music (or any consensus as to the irrelevant dimensions) stymies further analysis. Lacking a clear statement of the important dimensions, an empirical researcher has only a few options.

Firstly, the researcher can attempt to analyze "all" dimensions. This presupposes all dimensions are known. More importantly, this would complicate any analysis and possibly lead to irreconcilable differences between interpretations. Most critically however, this approach assumes that all dimensions are equally well known and equally amenable to analysis (i.e., quantifiable). For example, suppose a researcher decides to analyze parsing in terms of harmonic relationships and in terms of durations of notes. Further suppose that the results support the duration of notes interpretation and the results fail to support the harmonic relationships interpretation. The researcher might then conclude that the notion of harmonic relations should be dropped. However, it is equally likely that the results reflect the ease of accurate analysis in terms of duration and the difficulty of accurate analysis in terms of harmonic relations. That is, a different analysis of harmonic relations (i.e., a more appropriate music-theoretic analysis) might be a better predictor of the data than the duration analysis.

On the other hand, the researcher can choose particular dimensions (the "best") to analyze. This can be based on inertia, intuitions, ability and availability, a review of the literature or all of the above. The danger of this approach is that the important dimensions might be missed. This danger is magnified to the extent that the researcher tends to focus on a single level of analysis. That is, a researcher who focuses on global analysis may fail to consider the acoustic information that is available to the listener (e.g., assuming that the listener can remember all individual note events that occurred previously with equal clarity; failing to consider conceptually-driven processing during the encoding of stimuli). Conversely, a researcher who focuses on short spans of notes may fail to consider the goals of parsing (e.g., the units should be meaningful within the defined corpus of music; the stimulus must be analyzed as music, not as a sequence of acoustic events). The second danger of choosing a particular dimension to analyze is that regardless of whether or not the analysis is correct. some reviewers will automatically think that it is not.

Defining the Rules of Music for the Purpose of this Thesis

The solution to this dilemma was to choose a well developed theory that meets the previous constraints (concept of parsing related to rules and sufficient breadth of analysis), that has the general regard of the researchers in the field and that is sufficiently delineated that it can be used as a basis for an empirical test (i.e.. it can be quantified). There were two further constraints not mentioned previously but implied by the theoretical framework. The theory had to model the process of listening (not composing). or at least. the theory had to claim to model or address the process of listening. After all. the goal of the thesis was music cognition. not music analysis. The theory also had to minimize the amount of interpretation that I had to make when quantifying its predictions for subsequent tests (i.e.. the theory had to lend itself to conversion to an algorithm). West. Howell and Cross (1985) have provided a general discussion of these concerns.

There were numerous candidates. but the selection narrowed on Lerdahl and Jackendoff's *A Generative Theory of Tonal Music* (1983). Lerdahl and Jackendoff (see clarifications and/or refinements and/or extensions in Jackendoff. 1992; Lerdahl. 1988a. 1988b. 1993. 1996[8]) proposed to model the perception of music in a manner that mimics the successful modeling of the perception of speech. Their work continues to generate a lot of discussion and various empirical tests (cf.. Cross. Melen. Stammers. & Deliege. 1996; Deliege. 1987; Krumhansl. 1996; Lerdahl. 1996; Peretz. 1989). Explicit. detailed discussion of the model (and its relationship to the empirical literature) will be delayed until subsequent chapters (a review is available in West. Howell & Cross. 1985).

[8] Note that the theoretical framework for the current work had been pretty much determined by about 1993 or 1994 (algorithms written. experiments designed. and started). The data is approximate because there were retrospective alterations to the original algorithms as new data accrued. However. works later than about 1994 were not really a factor in the conceptualization. Some later works are included in this introduction. These later works provide possible solutions to problems raised within the present work: Sometimes the possible solutions were the same as those that were ultimately chosen. sometimes they were not.

Discussion of the empirical literature directed at the theory requires an extensive discussion of the theory itself and it seems parsimonious to discuss aspects of the theory when the quantification of those aspects is discussed.

Essentially, this theory proposes a set of rules that define the low-level parsing of the stream of musical sounds into various units. These units are then combined, on the basis of other rules, into a hierarchical representation of the unified perception of an entire (complex) piece of music. The process creates a hierarchical unitization of the piece of music with different rules guiding the selection of units at each level. At the lowest levels of the hierarchy, unitization (boundary creation between units) is driven by acoustic events that are contained within the music stream (changes in timing, pitch, intensity, etc.). Boundaries are formed on the basis of phenomenal accents[9] which are:

> any event at the musical surface that gives emphasis or stress to a moment in the musical flow. Included in this category are attack-points of pitch-events, local stress like sforzandi, sudden changes in dynamics or timbre, long notes, leaps to relatively high or low notes, harmonic changes and so forth.
>
> 1983, p. 17

These low-level rules focus on the acoustic features of the stimulus, though it seems evident that Lerdahl and Jackendoff assume that the perception of these features activates musical structures (i.e., the events are heard as, or for, their relevance to music). At higher levels, the various low-level boundaries are retained or dropped on the basis of progressively more complex musical structures (i.e., musical notions of parallelism, symmetry, tonality etc.). The theory creates a hierarchical parsing of a melody because every boundary at a higher level must be a boundary at every lower level (though the reverse is not true). The theory provides a fairly explicit qualitative account of the basis for parsing at the lowest levels of the hierarchy (explicit discussion can be found in Chapter 2 of this work) that can be easily quantified, but the theory requires an

---

[9] Lerdahl and Jackendoff (1983, p. 17) also define metrical accents ("any beat that is relatively strong within its metrical context") and structural accents ("melodic/harmonic points of gravity within a phrase"), but these are not the focus of this work. Metrical accents arise from the phenomenal accents, while structural accents emerge from the time span analysis of the piece.

unqualified leap of faith when moving from the low levels of the hierarchy (low-level rules) through the intermediate levels of the hierarchy to the high levels of the hierarchy. This leap does not explain how the low-level rules interact to create intermediate levels and the leap leaves two critical intermediate level concepts -- parallelism and symmetry -- undefined (Lerdahl & Jackendoff. 1983. p. 52. acknowledge this critical weakness but it was not addressed in their subsequent works). The importance of this lack of specification cannot be overemphasized. It leaves one wondering how the high level groups are really defined and the definition of these high level groups is critical to the remainder of the theory. For example. time-span reduction. which encompasses all the intuitions of tonal structure, is based on the groups that are identified at the intermediate levels. which in turn, are based on the groups that are identified at the low levels. In addition. the theory is not very quantitative (it was not intended to be; Lerdahl & Jackendoff, 1983. pp. 47. 53-55). As such. conversion into an algorithm required many more subjective decisions (from me) than I would have liked. At the lower levels. these decisions are not too important for subsequent modeling. but at the intermediate and higher levels. these decisions are much more subjective (the present thesis did not actually progress beyond the intermediate levels). In addition. the theory is somewhat ambivalent in its treatment of tonality (pitch or pitch class).

These difficulties notwithstanding. because it seems to encompass a number of the concerns raised in the previous sections. Lerdahl and Jackendoff's (1983) theory was chosen as a basis for subsequent work (rather than starting from scratch, using basic concepts of music and basic results in experimental music cognition. to produce. ultimately, something that would look very similar). However. it was necessary to modify the theory in those places where it was not sufficiently explicit. These modifications were minimized wherever possible (so to remain true to the original theory) and the rationale and methodology of each modification is extensively documented.

The relations between Lerdahl and Jackendoff's (1983) theory and other music theories will not be emphasized because Lerdahl and Jackendoff discuss these relationships extensively. Only the general relationships between Lerdahl and

Jackendoff's theory. music-theoretic concepts relevant to parsing and psychological investigations relevant to parsing will be discussed at this point. As stated earlier. links between the theory and specific psychological tests of the theory will be delayed until the discussion of the development of the algorithm.

Lerdahl and Jackendoff (1983) was chosen over other theories because. as noted by Agmon (1990) "Yet. only with Lerdahl and Jackendoff's *A Generative Theory of Tonal Music* (1983) do we have a large-scale music-theoretic work explicitly set within a cognitive-scientific framework". Jones and Holleran have declared it to "One of the most influential current theories about tonal music. . . . explicitly rooted in psychological principles" (1992. p. 3). Other large-scale theories seemed to be focussed on the global assessment of structure (e.g.. Meyer. 1956; Schenker as presented by Jonas. 1982) while remaining much more vague with respect to details at the lower levels of analysis (i.e.. how boundaries are formed). Even though the focus was Lerdahl and Jackendoff's work. it must be remembered that Lerdahl and Jackendoff's theory is not independent of other large scale theories. Lerdahl and Jackendoff see their work to complement rather than to compete with Meyer and Schenker (as well as Epstein. 1979. cited in Lerdahl & Jackendoff. 1983. p. 8). For example. Meyer uses the same Gestalt philosophy as the basis of his work. though Lerdahl and Jackendoff focus on similarity and proximity while Meyer focuses on good continuation. Lerdahl and Jackendoff explicitly acknowledge the link to Schenkerian analysis in their Reduction Hypothesis (1983. p. 106). They also acknowledge the overlap of their work with Tenney and Polansky (1980) and there is more than a passing similarity with that of Lindblom and Sundberg (1970: they focus comparisons on the work of Sundberg & Lindblom. 1976).

Interestingly. although there are similarities between the various global theories. Lerdahl and Jackendoff do not claim any similarities with the notions of Narmour (1977. cited in Lerdahl & Jackendoff. 1983. p. 106. 333). This is interesting because Narmour's (1991. 1992) implication realization model is an expectancy model based on the work of Meyer (1956): Future musical events may either adhere or contradict the expectancies generated by low and high level rules that link individual note events. Narmour's theory is empirically testable and has generated a number of studies (Cuddy & Lunney. 1995: Narmour, 1996; Russo & Cuddy. 1996: Schellenberg. 1996: Schellenberg & McKinnon. 1996). However. as a model of expectancy. it is not really a theory of parsing. As such. it could not provide an effective basis for this work.

Much of music theory (cf.. Anders et al.. 1975: Christ. DeLone. Kliewer. et al.. 1966; Green. 1965) centers its discussion of parsing (or related to parsing) on units like the theme. phrase and motive that are delineated by structures like the cadence and caesura. Lerdahl and Jackendoff (1983) clearly link to these structures though the terms are not often used explicitly (F. Lerdahl is a composer. R. Jackendoff is a linguist and performer). They prefer the use the generic term "group" for collections of notes (cf.. Lerdahl & Jackendoff. 1983. p. 12) while cadences are discussed throughout. Although I found no explicit mention of caesura. the concept of a light break exists throughout the work and is particularly important at the lower levels of the hierarchy. The reason that Lerdahl and Jackendoff were chosen was in part due to the extremely vague treatment that is typically given to such concepts as the cadence within music theory: It is almost as if one must have the knowledge of what a cadence is before the concept is defined. As was mentioned earlier. discussion of the link between the theory of Lerdahl and Jackendoff and other notions of musical theory will not be emphasized because Lerdahl and Jackendoff have devoted a great deal of their work to establishing these links.

Lerdahl and Jackendoff's (1983) theory was chosen above other theories. found more often within psychology. that have focussed on local detail (e.g.. Boltz & Jones. 1986; Deutsch & Feroe. 1981; Jones. 1976b. 1981a. 1982). Such theories discuss relationships within a short span of notes but were found to be difficult to extend to more complex stimulus (i.e.. algorithms to analyze a melody). Space does not permit a critical examination of each and every theory of parsing that has arisen in the last few years. but suffice it to say that if I had thought any of these had offered more than that of Lerdahl and Jackendoff's theory. then I would have used it. Many of these theories seemed to work in the constrained examples used to demonstrate them but the extension to other stimuli (music) did not seem reasonable. In addition, many of these theories focus on a single aspect of music ignoring the more global issues that would naturally confound the analysis if applied to more complex pieces. On the other hand, Lerdahl and Jackendoff often acknowledged the contributions of psychological literature, even if it seems that the choice is somewhat selective of those findings that support their theoretical framework (i.e., there are citations for Deutsch and coworkers, and for Dowling and coworkers, and Shepard, but none for Cohen, Cuddy, Jones, Krumhansl, Sloboda and their coworkers, to name only a few; of course, the date of the work precludes reference to much interesting psychology that has appeared since, and the style is not that of scientific writing within psychology). Lerdahl and Jackendoff stated (1983, p. 112) that they want their theory to be testable by experimental methods within cognitive psychology (concurrently, they also claim that it would not be possible to similarly test Schenker's theory).

General Considerations for the Parsing of Melody

Although the decision was made to focus on the theory of Lerdahl and Jackendoff (1983), it was also noted that the theory tends to be qualitative and not quantitative. This in turn, implies that the quantification of the theory will require some basis that is not contained within the theory itself. What follows is a general review of the empirical literature (mainly from psychology, music-theoretic concepts are mentioned briefly[10]) of factors that are implicated (or may be implicated) in parsing and boundary formation within melody (focusing on simple melodies as much as is possible). The review is general in the sense that no attempt is made at this point to tie this review directly to the specific constructs of Lerdahl and Jackendoff: The point of this review is to sort through the mass of literature to see what findings might apply to the parsing of melodies so that any gaps within Lerdahl and Jackendoff's theory can be filled in the most reasonable manner (when quantification proceeds). Reviewing the literature with the goal of supporting or refuting a particular theory might be faster and more efficient, but given the

---

[10] It is not possible in a work of this size (or a work that has this focus) to consider all the concepts of music theory that have been and might be implicated within the parsing of music. There is not a lot of cohesion in that literature. An attempt is made to link and discuss findings in terms of the most commonly cited structures: the phrase, the cadence, the motive and the caesura.

frailties of human cognitive processing (particularly selection bias), such a review has the real danger of ignoring relevant material. This review also has the secondary goal of demonstrating why it was necessary to resort to a single theory, rather than trying to define parsing on the basis of music-theoretic notions of musical structure.

The approach is to take a concept from music theory[11] that seems to be relevant to parsing or boundary formation and then to determine what support there is for that concept within the empirical literature. However, as was noted previously, many concepts within music-theory are ambiguously defined, or more rarely, suffer from multiple partially overlapping versions. In this review, an attempt was made to demonstrate the ambiguity thereby providing a rationale for the working definition of each term. Where possible the definitions of Lerdahl and Jackendoff (1983) have been provided, since ultimately, the purpose is to relate their theory to the empirical data. However, in this review, it must be remembered that Lerdahl and Jackendoff would possibly argue that some of these concepts are dysfunctional and that some of these concepts can be supplemented by the constructs of their theory. As such, *the commonalities and the distinctions between Lerdahl and Jackendoff's approach and the approaches of others in the literature need to be highlighted particularly when those distinctions may colour the interpretation or application of empirical results to Lerdahl and Jackendoff's theory.*

Different writers have different views because melody[12] or melodic pattern is a most difficult concept to define, for it encompasses all aspects of music that may be present in the serial order of tones. Music theory has typically argued that the melodic pattern can be treated as a set of hierarchically related units: motives at the most basic level, phrases at the intermediate levels, and sequences or themes at the highest level (cf., Anders et al., 1975; Christ, DeLone, Kliewer, et al., 1966; Christ, DeLone &Winold, 1975; Green, 1965; Lerdahl & Jackendoff, 1983; Lloyd, 1968; Meyer, 1956),. Listeners

[11] I am assuming some familiarity with the most basic concepts of music theory. To do otherwise would extend this work beyond reason.

[12] I have found the definition of melody (and rhythm) to be among the most capricious within music.

parse the stream of music, extracting motives, phrases and themes, thereby discovering or developing the relationships inherent in the music. Events within a melody -- notes, rests and other musical events (this work will limit the discussion of notes and rests, but in principle, other events can be treated similarly) -- can vary acoustically in frequency (pitch height and pitch chroma), intensity, duration (the duration of notes and rests, the duration between events, better known in psychology as interstimulus interval, and the duration from onset to onset of event [the sum of the 2 previous times]) and timbre (spectral composition). The perception and processing of these musical events, particularly note events, will be affected by the importance of that note within the tonal structure, as well as more global concerns of harmonic progression. The metrical emphasis given to notes will also affect the processing.

The Phrase

For most of the cited works, it seems clear that the basic unit of analysis is the phrase, much like the unit of analysis for linguistics is the sentence (or clause: cf., Clark & Clark, 1977; Miller & Johnson-Laird, 1976; Millward & Flick, 1985). What defines a phrase is not clear, but it is generally associated with the concept of a single musical thought in music (see previous comments pertaining to meaning). Lerdahl and Jackendoff (1983) do not actually define the phrase explicitly. However, it is clear that they view the phrase as a group of notes that has cohesion (cf., pp. 12, 30-31, 112-113, 118-119) delineated by structurally important events at the start and finish. In fact, Lerdahl and Jackendoff seemed to eschew the use of the term phrase, as well as related terms like motive, for the more general term "group" (cf., 1983, pp. 12, 119).

For the purposes of this exploration, the phrase can be considered to consist of two separable (but not independent) structures: the internal structure and the phrase delineating cadence (most works to discuss these constructs concurrently). Since the cadence seems to be more relevant for the question of parsing it will be discussed first. It must be emphasized, however, that even though the phrase terminating structure and the internal phrase structure are being discussed independently, they are not truly separable. What defines a cadence depends to some extent on the structure of the earlier part of the phrase.

The cadence is the prototypical phrase ending, but like the phrase, the definition of the cadence is somewhat ambiguous. Generally, the cadence is associated with a

> ...point in melody that provides momentary pause in the onward flow of a musical pattern....it separates one melodic unit from another. Like the written commas and colons and periods of speech, the cadence is a *heard* signal that helps us organize our world of tones...
>
> Christ, DeLone, Kliewer et al., 1966, p. 50

In *The New Groves Dictionary of Music and Musicians*, the cadence is:

> The conclusion to a phrase, movement or piece based on a recognizable melodic formula, harmonic progression or dissonance resolution: the formula on which such a conclusion is based.
>
> Rockstro, Dyson, Drabkin, Powers, 1980, vol 3, p. 582-586.

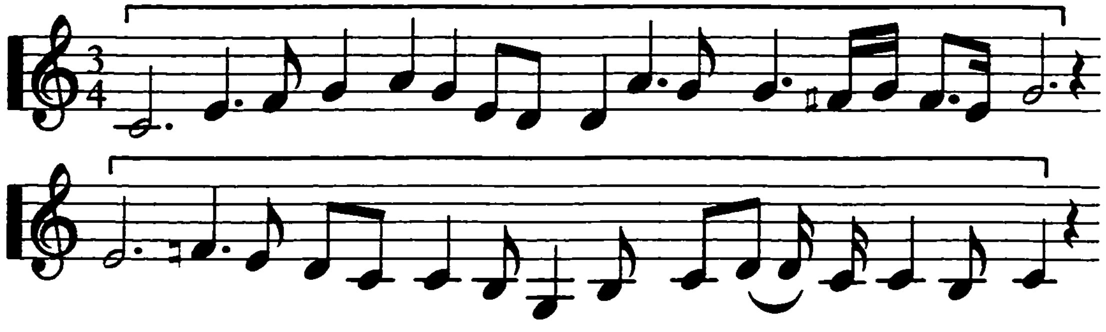
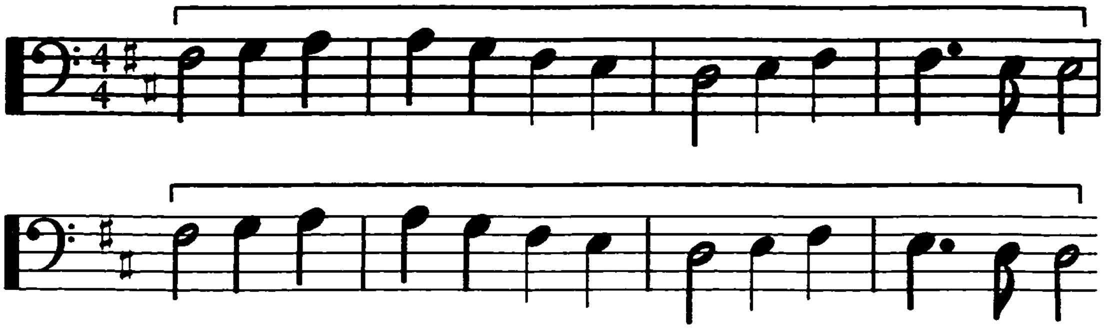
A cadence is a pause marked by harmonic, rhythmic and pitch effects. Specifically, cadences usually involve notes that are relatively longer, notes that are relatively more important within the tonal/harmonic structure of a piece (i.e., the tonic, dominant and possibly the mediant or subdominant) along with actual breaks in the flow of music (rests) and a definite slowing of the pace of music (rhythmic structure). Harmonically, cadences are often marked by specific chord progressions such as the V-I (authentic), or IV-I (plagal). However, this is not always the case. Although Lerdahl and Jackendoff (1983) often refer to the cadence as a phrase marker, the closest that they came to a definition was:

> ...the cadence, the goal of tonal motion.
>
> 1983, p. 17

> ...consider the simple V-I progression. If it occurs at the beginning or in the middle of a group, it is not heard as a cadence, since a cadence by definition articulates the end of a group. If the progression occurs at the end of a group it is heard as a full cadence -- either "feminine" or "masculine," depending on whether the V or the I is metrically more accented. If a grouping boundary intervenes between the two chords, the V does not resolve into the I; instead the V ends a group and is heard as a half cadence, and the I is heard as launching a new phrase.
>
> 1983, p. 28-29

For Lerdahl and Jackendoff, cadences represent structural accents that mark the end of the phrase (and all higher level structures). However, although a cadence must occur at the end of a group, the structures that mark the cadence do not necessarily indicate the end of a group. That is, a structural accent (a point of harmonic or melodic gravity) that occurs at the end of a group is a cadence. The group ending is defined before the cadence is identified. Although this seems similar to other conceptions of the cadence, it is somewhat reversed.

Typically, cadences are thought to mark the end of a phrase. Like punctuation in language, cadences come in many levels. The terminal cadence (final or full cadence) marks a total or partial cessation of musical activity, much like the period in a sentence. The progressive (half or semicadence) marks a break in the flow but with the hint of continuance, much like the comma in a sentence (analogies are provided by Christ, DeLone, Kliewer et al., 1966). Example A in Figure 1.1 (taken from Christ, DeLone, Kliewer et al., 1966) provides an example in which phrase markings are aligned to a semicadence on a lengthened dominant and a full cadence on a tonic. However, note that Example B (from the same analysts) places a phrase boundary at a longer note (Note 13) which is only the supertonic (not generally considered harmonically important). Observe that there is no cadence on Note 8 (or Note 21) even though both are tonics (harmonically, the most important). How does one know that the cadence is after the thirteenth note and not the fourteenth (which have the same duration). In Example C, the lengthened mediant (Note 4) does not constitute a cadence (too soon?), but the somewhat shorter, though still lengthened, submediant at Note 10 does. Note 10 would be the tonic in the minor mode (D minor), but then the previous Note 4 that lacked a phrase ending would be the dominant (which, admittedly, is slightly less important in minor keys than major keys). The lengthened supertonic (subdominant in the minor mode) does not warrant a phase ending. Rhythmic considerations and pitch changes are implicated but it is not clear which are important in any one instance. The difficulty is not so much that one cannot understand the rationale for the phrases chosen: Rather, the difficulty is understanding why other choices were not made.

Example A

Example B

Example C

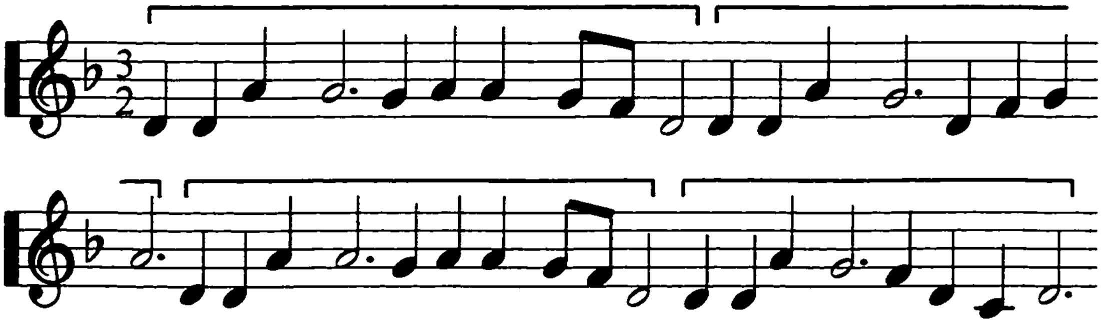
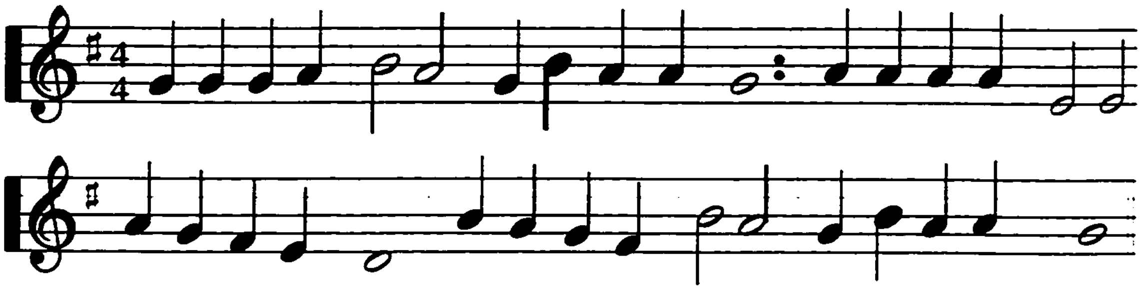
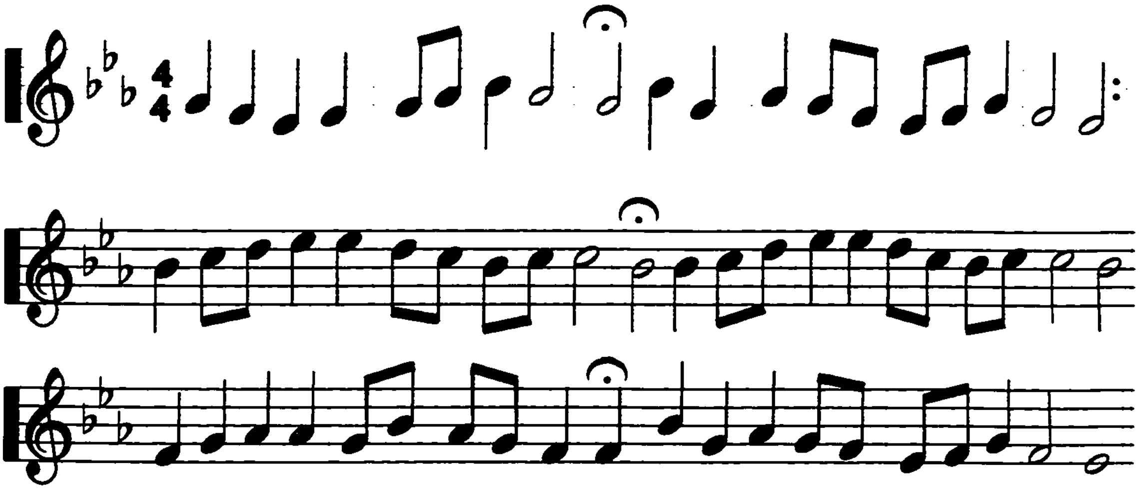



Figure 1.1 Examples of phrase delineation taken from Christ, DeLone, Kliewer, et al. (1966, p. 57 [John Dunstable.: Sancta Maria"], 51 ["Beethoven: Symphony No. 9, IV"], 53 ["American folksong"]). Note that the definition of phrase endings is somwhat idosyncratic.

Example A

Example B

Example C

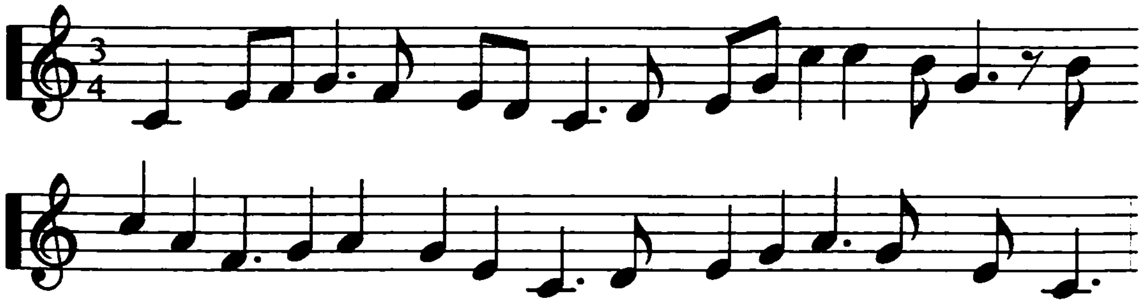



Figure 1.2 Examples of phrase delineation taken from Lloyd (1968, p. 424 ["Oh Clair de la lune"]), Green (1966, p. 79-80 ["Bach: Schmucke dich, o liebe Seele, Corale No. 22"]) and Anders et al. (1975, p. 43 ["Barbara Allen"]). Notice that the definition of phrase endings is somewhat variable, though each can be rationalized. It is the lack of phrase endings that is more difficult to explain.

That is, the rationale for a boundary at Note 13 in Example B (Figure 1.1) is clear, but the lack of a boundary at the very similar Note 8 in Example B is not clear. For the purposes of an algorithm, all phrase endings and only phrase endings must be identified by the application of a set of rules. The uncertainty shown in the examples in Figure 1.1 are not unique to those writers: Figure 1.2 provides several examples from different authors. Again, one can see the rationale for the phrase boundary constructions, but one can also question the lack of phrase boundaries at other locations. Regardless of the cause, this uncertainty makes using a cadence (however defined) as a boundary marker within an algorithm very difficult.

These examples also raise an issue that will reappear throughout this work. It has been argued that music-theoretic analyses of phrase structure is guided more by the visual stimulus (i.e., the score) than by any associated auditory process. For example, in Example A of Figure 1.2, while listening to the melody, one might delineate a phrase on Note 4, a lengthened dominant, or possibly after Note 5, the second lengthened note. While hearing the melody, the listener would not know, at the time that Note 4 was played, that a better cadence marker was due to arrive soon. The same can be said in Example B, where the lengthened subdominant is not regarded as a phrase boundary, but the following lengthened (with fermata) mediant is. Later the lengthened (with fermata) dominant is not grounds for a phrase boundary (it is notated as a repetition of the previous phrase). I raise this issue because, although Lerdahl and Jackendoff claim to model the listening process (admittedly the listening process of an experienced listener, 1983), their theory has been critiqued as modeling the "intuitions of a sophisticated listener who has the score of the piece" (Rosner, 1984, cited in Cook, 1989, p. 119). It is important to always keep this caveat in the back of one's mind while thinking of these findings, and particularly while thinking of the link between theory (musical analysis) and empirical data.

From these examples (and others not cited), one may tentatively conclude:

1) Lengthened notes *may* be a basis for a cadence,
2) Tonally important notes *may* be a basis for a cadence,
3) Rests are *very often* the basis for a cadence,

but

4) A tonally-important and lengthened note *may or may not* be the basis of a cadence.

Based solely on music-theoretic notions of parsing, for the design of an algorithm, it would seem that lengthened notes and tonally important notes are neither necessary nor sufficient conditions for the generation of a phrase boundary.

Regardless of the knowledge that can be gleaned from theory, it is important to consider the empirical data pertaining to the phrase delineating cadence; that is, based on empirical studies, which theoretical aspects of the cadence are actually perceived (with some regularity)? The notion of a phrase delineating cadence has been tested both indirectly and directly, although direct methods are relatively rare. In the direct approaches, behavioral measures of boundary formation or unitization are taken, and these are compared with the analysis of the music. The strength of the approach is that one knows the actual listener's unitization of the piece (assuming the design is valid), but one must be careful to consider all possible causes of the unitization (i.e., responses that may be due to factors other than phrase boundaries), not just those defined by the experimenter. In the indirect approaches, studies assume the music-theoretic notion of a cadence as a starting point, and then examine various behavioral measures that should be affected by that cadence. The cadence is the assumed cause of behavioral responses, but given the complexity of music (and, often, the actual stimuli) this is not always a safe assumption. Hereafter, the terms direct and indirect are used as a shorthand: Direct means that subjects indicated phrase endings and indirect means that the researchers assumed a music-theoretic (or similar) idea of phrase endings and tested subjects responses based on this assumption.

Using a direct approach, Krumhansl (1996, Experiment 1) asked subjects to indicate, by a button press, the end of a major section while listening to Mozart's Piano Sonata in E♭ Major (K. 282, First Movement; note that this is a fairly complex piece, well beyond the notion of a simple melody). Results coincided nicely with the notion of the cadence. That is, subjects indicated, with high consensus, boundaries at a point where the tempo slowed (which corresponded to the use of relatively longer notes). By inspection of the score, these relatively longer notes were also tonally more important, at least relative to the preceding notes (i.e., the first boundary occurs on a variation of a I-V cadence and the second on a different variation of the I-V cadence, see Lerdahl, 1996 for a more detailed analysis). Although the lengthening of notes seemed to be relatively more important than tonality, without a quantitative analysis, the role of these two factors cannot be separated. In another phase of the experiment, subjects were asked to indicate the beginning of each new musical idea by a button press. Although the introduction of new ideas was associated, as expected, with the beginnings of sections, the introduction of new ideas was not generally associated with longer notes (i.e., in Krumhansl's terms, new ideas began on a neutral tempo). New ideas were not associated with tonal centers either. That is, Krumhansl did not make any such association, and no such associations could be discerned from the score (however, temporal resolution of subjects' responding was not high). A qualitative analysis indicated that new ideas were associated with changes in register, contour or pitch pattern, rhythm dynamics or accompaniment. Most importantly, the introduction of new ideas was associated with positions of relatively low tension (defined empirically), implying higher tension within the musical idea. Experiments 2 and 3 of the set explored the same issues at a lower level of analysis. That is, subjects were presented with progressively smaller sections of the music, but performed the same analysis. Although the units identified were smaller, the basic results were the same. Sections were identified by notes of relatively longer durations, but the introduction of new ideas was not. The general implication is that in the parsing of music, boundaries can be formed on the basis of cadences, but these usually mark large-level structures and not low-level structures. It did not appear that even the mini-version of the cadence (i.e., a caesura) was used as the marker of units smaller than the section.

Using a more indirect approach, Palmer and Krumhansl (1987a) asked subjects (all had some musical training: 2.5 to 14 years) to rate the goodness or completeness of a short musical segment (based on Bach's Fugue XX in A minor (from the *Well Tempered Clavier*, Book 1). The segments either ended on a phrase boundary or on some point within a phrase (phrase analysis by Lerdahl, based on Lerdahl & Jackendoff's, 1983). Higher ratings of completion were produced by segments ending on a phrase boundary (similar, but smaller variations were noted for the smaller units within a phrase). It is important to note that Palmer and Krumhansl found that simple notion of tonality, as defined by the quantified tonal hierarchy (see Krumhansl, 1990), was also associated with the sense of completion. This is to be expected since phrases are supposed to end on tonal centers. The second experiment of the study replicated the results with a slightly different method. Palmer and Krumhansl (1987b) repeated the experiment using a more complex (harmonically) piece of music from Mozart's A Major Piano Sonata (K. 331). The results were essentially the same. They also explicitly examined training and found no effect in this task.

Using an indirect procedure, Sloboda and Gregory (1980) presented subjects with tonal sequences that contained phrases delineated by either a lengthened note (called a Physical Marker), or a tonal center (called a Structural Marker), or Both, or Neither. One can see that the Both condition mimics the notion of a cadence (i.e., a relatively longer, tonally important note). Subjects were required to detect the presence of clicks inserted into the melody. The perception of clicks tended to migrate to the boundary, with the additional effect that subjects seemed to be able to use structural, but not physical, information in an anticipatory manner.

Although not directed at phrases within a song, Boltz (1989, Experiment 1) presented both musically trained and untrained subjects with unfamiliar folk tunes and asked subjects to rate the degree of completion for each melody. The endings of tunes were altered to study the effect of the tonality of the last few notes. Using an indirect test, some melodies were designed to have clear indications of a phrase ending (leading tone to tonic; dominant to tonic), some had less clear phrase endings (submediant to tonic; tonic to dominant) and some had unclear phrase endings (tonic to leading tone; tonic to submediant). The ratings of subjects demonstrated that the least complete ending was tonic to leading tone while the most complete ending was leading tone to tonic. The second most complete ending was submediant to tonic (not dominant to tonic), but one must remember that these were folk tunes and not orchestral pieces. All other endings were approximately equivalent. Musical training was not a factor in the results. These ratings of melodic completion were highly correlated with ratings of tonal appropriateness. Boltz concluded (p. 753) melodic completeness was determined by the tonality (however, this was the only dimension manipulated). The second experiment confirmed these results with slightly different endings in conjunction with rhythmic alterations and the presence or absence of key-defining, tritone relationships.

In a continuation of the work, Boltz (1991a) asked musically experienced subjects to notate unfamiliar folk songs. All songs contained theoretically unambiguous cadence markings, but in the presentation of some songs, the tonal markers of the phrase endings (i.e., notes of the tonic triad marking the cadence) were replaced with notes that were contour preserving and diatonic, but not of the tonic triad. In addition, songs were presented either with meters (actually, rhythms was the term used) that highlighted the phrase ending (i.e., meters that were compatible with the harmonic structure) or with meters that obscured the phrase endings: All meters were musically meaningful and none changed the specific timings of the notes. Accuracy of notation was highest when the tonic triad phrase endings existed: the presence of a compatible meter was also important. Interestingly, the absence of the triad tone or an incompatible meter or both were all detrimental, to more-or-less the same degree. A more detailed analysis at the phrase markers indicated that when phrases contained the tonic triad member and the meter was compatible, the most common error was to substitute a different tonic triad member. If the meter was incompatible, subjects rarely notated the proper tonic triad, nor any other member of the tonic triad, but rather, preferred some other note (no pattern in the errors was discerned in the phrases marked by non-triad notes). The implication is that meter defines when an (important or otherwise) event should occur and without appropriate meter, important events can be missed. Alternatively, one can argue that subjects thought that any note that did not occur on the important beat could not have been important. The important point is that a cadence is more than just a lengthened, tonally-important, note: It must match the metrical structure as well.

Many other studies have implicated the role of cadences in phrase boundaries. though more indirectly. Palmer (1989) noted that musically trained subjects slowed their performance of piano at the points that they had indicated as phrase boundaries (which were associated with cadences). However. subjects indicated their perception of phrase boundaries by visual inspection of the score which may or may not correspond to the auditory processing of the same music. Sloboda (1977) asked musicians to continue playing after the score had been removed from sight. He noted that the eye-hand span tended to reflect the presence of phases marked by the music-theoretic notion of cadences. Again, performers may be taught to read a score in visually-delineated phrase units. but they may not hear the music that way. Sims (1988) observed that when instructed to move in response to music. both adults and young children reflected phrase structure. Krumhansl and Jusczyk (1990) have demonstrated that infants preferred to listen to Mozart's minuets that had been segmented at the "natural" phrase boundaries as indicated by the score. rather than at some arbitrary position. However. the determination of natural was made by the authors (based on the score). although it was verified by untrained adult listeners. These listeners were asked to indicate which version sounded more natural in a forced-choice paradigm: 79% of the time the natural parsing did sound more natural (the lack of perfection was attributed to task demands). Jusczyk and Krumhansl (1993) used a more extensive design and analysis to confirm their observation (in the previous work) that infants "identified" increased durations. large pitch drops and possibly, the harmonic interval of the notes (the base and melody line formed an octave) as cues to phrase endings (note the similarity to the notion of a cadence).

The empirical data indicate that phrases can be and are marked by cadences (as defined by lengthened. tonally important notes or breaks in the music). However. both the theory and the empirical studies indicate that this cannot be the whole story (e.g.. metrical emphasis is important). However. as alluded to previously, phrases have structure besides that of the terminating cadence.

The idea is that a phrase should start from a point of relative stability. move to a point of relative instability, and then return to a point of relative stability: A key word here is "relative". For example, a phrase could move from the tonic (e.g., the tonic itself or a note of the tonic triad), to the dominant and back to the tonic, or a phrase could move from the tonic, to the mediant, and then to the dominant: Other combinations are possible. In this construct, phrases are only delineated by cadences that mark a return to a point of tonal stability (cf., Lerdahl & Jackendoff's, 1983, previously cited definition). Anders et al. (1975, p. 43) presented the phrase parsing of a simple melody (Figure 1.2, Example C). Phrase 1 is delineated by a lengthened tonic, Phrase 2 by a lengthened dominant plus a rest, Phrase 3 by a lengthened tonic, and finally, Phrase 4 by a lengthened tonic and no rest. However, a lengthened dominant in Phrase 1 is not a phrase boundary, and neither is a lengthened subdominant in Phrase 3. It seems reasonable that these were not considered phrase boundaries because these do not express a complete musical thought. That is, in each, the phrase had not moved from stability, through a second lesser stability to a final stability. However closer inspection reveals some rather quirky applications of the "rules". In the example, the first half of the first phrase moves from the tonic to the dominant but this is not a phrase because it does not move through a point of stability (relative or otherwise). However, this is unclear, since the second note is the mediant (considered a point of relative stability) and this note is longer than the preceding tonic. Perhaps the sequence is too short. The second phrase does not really start on a point of stability (though it starts as the first step away from stability) and more strangely, moves through the tonic (the most stable tonal point) and "returns" to the dominant (not as stable as the tonic). Is it possible that a single dotted quarter note on the dominant is more stable than two quarter notes on the tonic? The third phrase does not really match predictions either, starting on the leading tone, moving quickly through the tonic, pausing on the subdominant and finishing on the tonic. The final phrase seems more prototypical, moving through the subdominant to the tonic. None of the phrases, except the first (but it was an eighth note), truly started on a point of relative stability (contrast with Phrase 2 of Example B of Figure 1.1). It is true that for all the phrases (except the first), the preceding phrase ended on a point of stability, but the first note of the phrase was not one typically called stable within the key. To return to the previous Example B of Figure 1.1, one can easily argue for a tonal arc from the mediant (Note 1) through two instantiations of the dominant (Notes 3 and 4) to the tonic (Note 8), but this is not the phrasing chosen. Is it because this would leave no role for Notes 9 though 13 (too short to be phrase)? A listener would not know that there were four notes to come that did not form a proper group while listening: Once again, the distinction between having a visual representation (the score) and having an auditory representation is critical when reviewing these issues. The repose created by two half notes (e-f; Notes 13 and 14) does seem like a strong resting point. It seems natural (in some sense) to split this into two, and to subsequently use the dominant to begin a phrase, but why not use the dominant to end the phrase (or the tonic at Note 21)?. Is it because such a scheme renders the role of Notes 15-21 less clear? Is it because the choice of Note 14 as the beginning of the second phrase parallels the beginning of the extract?

There is a second point to this notion of phrasing (phrasing marked by cadences and harmonic progression) that must be mentioned, if only briefly. If one is to treat phrases as motion from tonic to dominant to tonic, then one must know that listeners can establish key (or key-neighbourhood). Fortunately there is a good deal of evidence to suggest that people can establish key quickly (cf., Cohen, 1991) and that people can use internalized representations of tonal hierarchies as a basis for processing music (Krumhansl, 1990). Note that Lerdahl and Jackendoff's (1983) conception of a cadence as the structural accent that occurs at the end of a predefined group does not actually suffer from this problem. That is, groups are defined on the basis of acoustic feature and higher-level rules define what is musically important within a group (time-span reduction).

What does this mean for parsing? Obviously melodies are structured around the important notes (this is the basis of Schenkerian analysis and it is also a part of the reduction analysis of Lerdahl and Jackendoff, 1983). The classic structure is one of beginning on the tonic, followed by motion to the dominant (or subdominant) followed by a return to the tonic. Some tones are only important as steps between these points and some tones are only important as pointers to these elements (e.g., leading tones).

Unfortunately, the implications for parsing are not clear at this point. For example, a cadence is usually aligned with structurally important notes, but as has been shown, this is not a necessary nor sufficient condition for a cadence, either theoretically or empirically. It is possible that the important notes of a phrase may be anywhere within a phrase, and as such, cannot be used as a phrase delineating structure (i.e., as a "rule" for parsing melodies into phrases).

Analogies with semantic language can be fruitful (cf., Lerdahl & Jackendoff, 1983, pp. 112-113; see also Clark & Clark, 1977; Miller & Johnson-Laird, 1976; Millward & Flick, 1985 for discussions of grammatical syntax, linguistics and language [speech] perception). Semantic phrases are generally structured around one important word (i.e., noun, verb, infinitive, participle, gerund and prepositional phrases), but that word need not be the first, central or last of the phrase: It can appear at any point in the phrase. Hence, one cannot use the *position* of the word to mark the boundaries of the phrase even though the relative positions of words can be used as an aid to parsing (e.g., articles precede nouns, not vice versa). If one knows the meanings of words, it is a trivial matter to pick the important word from a phrase because words are hierarchically ordered in importance: nouns and verbs, then adjectives and adverbs, then articles, prepositions, et cetera. A phrase contains only one noun or verb (that is its defining feature). Other words of the phrase lead into or modify that central word in some way (e.g., most easily seen in the prepositional phrase) and minor variations on the other words can colour the meaning of a phrase (consider: "at the ballpark", "at a ballpark", "at that ballpark", "at their ballpark"). Phrases can consist of a single word (e.g., usual for noun and verb phrases) and furthermore, the important "word" may actually be two words (e.g., infinitive phrases). However, the point is that word position is no indication of importance even though word order is constrained (grammatical syntax). The same may be true for phrases in music. It may be easy to pick out the important notes in a phrase (via tonality, harmonic progression), but the positions of those notes are not fixed for all phrases, even though musical syntax, like grammatical syntax, will restrict possible note orders (e.g., important notes often occur at the end of a phrase -- the cadence).

Regardless of theoretical concerns. there is evidence that phrase structure (important notes) is a determinant of music perception. Jones and Ralston (1991) have indirectly demonstrated that harmonic structure is important for recognition of a phrase. They asked subjects to remember a number of sequences that modeled the prototypical phrase structure: starting on the tonic. moving to the dominant (there were actually two instantiations of the dominant) and returning to the tonic in the usual arch motion (see below). These phrases could be presented with different rhythmic structures (these will be defined more concisely later). For present purposes. in a subsequent recognition task. subjects more often confused distracters that had the same harmonic structure than those distracters that had a different harmonic structure. Those distracters with the same harmonic structure placed the dominants in the same temporal positions. with or without the same durational accent. Those distracters having a different harmonic structure temporally altered the positions of the dominants. The phrase delineaters were never altered. Unfortunately. more detailed analysis of these structures (particularly the relationship of the dominants to the durational accents) was not provided.

In a continuation on her previous study of melodic completeness, Boltz (1989. Experiment 2) asked subjects to rate the completion of various melodies. The different melodies contained rhythms that emphasized different harmonic progressions repeatedly within melody: tonic to tonic. tonic to dominant. tonic to mediant and tonic to leading tone (these points corresponded to relatively longer notes). In conjunction with this. the melodies altered endings to reflect various types of cadences: leading tone to tonic. leading tone to dominant. leading tone to mediant. tonic to leading tone. dominant to leading tone and mediant to leading tone. Finally some melodies contained the tritone as a marker for tonality. while others did not. The basic result implied that all three factors were significant. but the results clearly demonstrated that, compared to the type of ending. the different harmonic progressions (and the presence of the tritone) were not a major factor in the perception of completeness.

The structure of a phrase has also been associated with the pitch contour of a phrase. Pitch contour is basically the pattern of up and downs in the frequencies of notes.

It is often depicted as a wavy line denoting the motion in frequency. Many have discussed the general pitch contour of phrases as an arch from the initial stability to the final stability. In the usual arch (the "arch of tonal motion" in Lerdahl & Jackendoff's terms. 1983, pp. 30-31), the central point of relative repose occurs at the highest point (in frequency) of the arch (Christ. DeLone. Kliewer. et al., 1966, p. 68, 70-71). The situation can be reversed, so that the central point of relative repose is at the lowest point (in frequency) of the arch. Such a phrase structure is typically referred to as an inverted arch. A rarer structure is that of the axis contour in which there does not seem to be an arch, but rather motion that "dances" around a single pitch (Christ. DeLone. Kliewer. et al., p. 50). Analogies to the notion of inflection in speech are often made. That is, an arch is a statement, the finality of which depends on the type of cadence. The inverted arch is a question, usually terminating with a progressive cadence to imply that an answer will be forth coming (Christ. DeLone. Kliewer. et al., p. 72). Pitch contour is also critical for the theory of Lerdahl and Jackendoff (1983) but in a different manner: Abrupt changes in the pitch contour are a basis for creating a boundary. That is, in the model of Lerdahl and Jackendoff, a break in the (smooth) pitch contour is considered a basis for parsing. The term smooth was added because what constitutes a break depends upon the structure of the contour (details are presented in Chapter 2).

At this point, I would like to introduce the term pitch pattern to replace the term pitch contour. The reasons are three-fold. Firstly, ultimately the term "pitch pattern" will be more representative of the subsequent discussions than the term "pitch contour". That is, pitch pattern is intended as a term that encompasses all relationships between the pitches of adjacent notes within a melody, not just the contour per se. Secondly, the term serves to remind that this discussion will ultimately lead beyond that which is normally associated with pitch contour (i.e., the pattern of up and down), and it is more in accordance with the discussion of Lerdahl and Jackendoff. (This also avoids creating another potentially confusing definition of pitch contour.) Thirdly, I need to introduce the concept of time pattern, a construct to parallel that of pitch pattern, based on the timing of notes and other note events (to be explained in detail later).

The two constructs of pitch pattern and time pattern will be critical to later discussions of parsing within music: Both are critical elements of Lerdahl and Jackendoff's (1983) theory, particularly at the lowest levels of their hierarchy. That is, the theory of Lerdahl and Jackendoff implicates pitch pattern and time pattern as the cause of boundary formation at the most basic level of their hierarchy, although they do not use these terms to define these effects. Moreover, pitch and time pattern are concepts that are necessary for discussion of parallelism within music (most of which will appear in Chapter 4). *Parallelism is critical for parsing within the theory of Lerdahl and Jackendoff, but parallelism is not defined in the theory of Lerdahl and Jackendoff* (1983, p. 52). Hence, to properly convert the theory into a testable model, it is necessary to consider parallelism. Hence, it is necessary to define the concepts that will be used later for parallelism.

The two concepts of pitch and time pattern are introduced here (i.e., not later in Chapter 4) because the goal is to develop a general framework of the essential elements of these two concepts. This framework is best developed within the context of a general discussion of the structure of music and within the context of other specific features of music that may be relevant to the parsing of music. As with other aspects of Lerdahl and Jackendoff's theory, explicit discussion of these concepts as they apply to the algorithm (i.e., details of quantification) will be delayed until the algorithm is discussed.

Pitch contour is typically construed as the motion between different pitches or frequencies, that is, the pattern of up and downs in the frequencies of notes, sometimes without regard to actual interval size. However, this is too basic. Pitch perception in music is categorical (cf., Handel, 1993, pp. 266-289, particularly, pp. 280-286; Dowling, 1993; Shepard, 1982b). Hence, frequency is not the same as pitch. In western-tonal music, categorical perception means that frequencies are heard as belonging to 1 of 12 defined notes (pitch chroma) equally spaced within an octave. The spacing is equal in the logarithmic sense. Hence, pitch contour (hereafter, the term pitch pattern will become more appropriate) could be thought to represent categorical or step motion in between different frequencies that are equally spaced on a logarithmic scale. This is still not plausible, because most western-tonal music restricts itself to only 7 of the possible 12 pitches within an octave. This restriction probably reflects a basic limitation of human information processing in its ability to classify (i.e., not discriminate) within a single acoustic dimension (Handel, 1993, pp. 268-270, 327; see also Agmon, 1990). These pitches are not equally spaced, even on a logarithmic scale, but it is arguable that the steps between adjacent pitches are perceived as equivalent (cf., Shepard, 1982a, 1982b). There is a large number of ways to chose seven semitone-spaced notes from an octave (792) but only seven combinations have been developed historically within music. These seven "diatonic" scales are the Ionian, Phrygian, Mixolydian, Locrian, Dorian, Lydian and Aeolian. Of these, most western tonal music uses the Ionian scale, which is better known as the major scale, while some uses the Dorian or Aeolian scales (melodic and natural minor scales). All western tonal music is written in a particular scale, but the music itself need not be restricted to only those notes of a scale (i.e., the music may include notes that are not from the set of 7 defined notes, although only in small proportions). Music that is restricted to the set of seven tones is called diatonic. Much of the simpler western-tonal music is diatonic (nursery rhymes, Christmas carols and folk songs; Lloyd, 1968, p. 141).

This leads to a critical question, one which Lerdahl and Jackendoff (1983) did not explicitly answer (it is a frustrating fact that one cannot prove by reference that the authors did not say something on a particular topic). Should pitch pattern be construed in terms of equal steps along a diatonic scale? It seems reasonable, but this creates a problem of how to deal with the non-diatonic tones that are common in much of western-tonal music (i.e., non-diatonic pitches within a melody are still part of the structure of western-tonal music). Are these to be treated as exceptions? or half steps? or are these perceived categorically the same as their diatonic neighbour?. For example, when listening to a piece of tonal music within the key of C (notes: c-d-e-f-g-a-b), the occurrence of the note c# is clearly off-key, and yet, is still an appropriate part of the music. The step from c to d is not qualitatively the same as the step from c to c# or c# to d. Furthermore, the step from e to f is not the same as the step from c to c# even though both are semitone steps. It is difficult to know how to treat these differences algorithmically: An algorithm needs a precise definition of such concepts (since Lerdahl & Jackendoff have proposed a qualitative theory, this is not an issue for them; however, they do say that their theory can be quantified; see Lerdahl & Jackendoff, 1983, pp. 53-55). There is an additional complication. Within the diatonic scale, tonality implies that some pitches -- the tonic, dominant, subdominant, mediant -- are more important than other pitches (this is the basis of, or linked to, the notion of harmony). How can these differences be accounted for algorithmically within the notions of pitch pattern or pitch contour (see Handel, 1993, pp. 344-362 for similar comments)?

As stated, Lerdahl and Jackendoff (1983) were less than explicit about the scale to be used for the processing of pitch information. In most examples, they seem to be using an implicitly diatonic system (i.e., steps on the diatonic scale are considered equivalent). That is, the early examples of the application of the low-level rules were restricted to diatonic melodies (no accidentals). Later more complex examples tended to focus on the high-level rules, which by their global nature, tended to ignore the issue of steps between individual pitches. Tonality is clearly involved in these higher level rules. Lerdahl (1992) developed an extension of Lerdahl and Jackendoff to deal with pitch relations (the pitch-space model) but that model was still focussed on the role tonality at the higher levels. This model did not seem to apply at the low levels of the hierarchy. Jackendoff (1992) implied that pitches were immediately heard within a particular key; that is, the listener tries to place the key and does so within the first few measures of a complex piece. This key can then retrospectively affect the role of prior notes. Unfortunately, in the example provided, the key fixation depended on the existence of a cadence, which in turn depends on the key; recall that the cadence is the "goal of tonal motion" (1983, p. 17). Although Lerdahl implicated tonality as an immediate construct, he did not indicate how it was to be rendered for use within the low-level rules (i.e., the rules that pertain to pitch pattern).

Other authors have tried to define a scale for the analysis of pitch pattern. For example, Deutsch and Feroe (1981) have constructed a three-level model of melodic structure that analyses short melodies in terms of non-diatonic, diatonic and triad tones (note that Lerdahl's pitch-space model was based, in part, on Deutsch and Feroe). However, the model has not been the subject of many tests. Many others have tried to model such structure (e.g., Krumhansl, 1979; Shepard, 1982a ,1982b; see also Narmour, 1991 for a different scheme that separates interval information from pitch height). The point is that, in a review of the literature, one must always judge results pertaining to pitch pattern in light of the pitch scale[13] being used within the study. In fact, it is arguable that studies based on non-diatonic pitch scales do not actually sound like music, and as such may not be processed as music (see previous comments pertaining to the choice of stimuli).

In the following review, the role of pitch pattern (defined as the relationships between the pitches of notes which implicitly includes pitch contour) in parsing will be examined. Because Lerdahl and Jackendoff (1983) used pitch pattern as a basis for parsing, but did not define the scale for discussions of pitch pattern (or pitch contour), it is important to explore the literature in this area so that when the algorithm is constructed, decisions can be made as to the scale to be used. Also, as stated previously, pitch pattern (or pitch contour) has been more generally implicated in the parsing of melodies. In addition, the analysis of parallelism will require a rigorous definition of pitch pattern. Hence, the following review examines the role of pitch pattern as a mechanism of phrase delineation with due regard for the issue of what is the most appropriate alphabet for

---

[13] In this work, an attempt has been made to restrict the use of the word "diatonic" to those sequences which limit their component tones to the set of diatonic scale pitches. The term "tonal" has been used to define a more restrictive set of conditions. To be tonal, sequences must be diatonic and weigh pitch classes in accordance with the rules of musical composition (cf., Krumhansl, 1990). To this scheme, we could add "harmonic" for sequences that adhere to the rules of musical composition (cf., Cuddy, Cohen and Mewhort, 1981). Herein, the terms "atonal" and "non-diatonic" are considered equivalent for sequences that are not diatonic. An attempt has been made to correctly classify the stimuli of all experiments cited, but unfortunately, more often than one would like, the statements concerning tonality or diatonicism made by authors (if any at all) had to be taken at face value. For details one must consult the actual work. Extracts from "real" music can generally be considered tonal (harmonic), but exceptions will be noted.

discussing pitch pattern. Unfortunately, studies that directly address the role of pitch pattern in the construction of phrase endings are rare, so a number of less direct observations will have to be utilized.

Time pattern is intended as a construct to parallel that of pitch pattern. That is, time pattern is the relationships between the timings of adjacent notes (and note events, particularly rests). This includes both the durations of notes and the interstimulus times. Lerdahl and Jackendoff use such structures as the rules driving unitization in the lowest levels of their hierarchy (see Chapter 2 for details), though they never use the words "time pattern". However, the exact nature of these relations was not defined. Much of the conceptualization of time that is to be used within the present work has been influenced or borrowed from Boltz, Jones and their coworkers (e.g., Boltz, 1991b; Boltz, Marshburn, Jones & Johnson, 1985; Boltz & Jones, 1986; Jones, 1976a, 1976b, 1981a 1981b 1982; Jones & Boltz, 1989; Jones, Maser, & Kidd, (1978). As with pitch pattern, there is some question as to the most appropriate scale to analyze the temporal relationships within sequences of notes. In this case, however, it seems relatively unambiguous that humans are capable of accurate linear assessments and comparisons of acoustic events with temporal characteristics of music -- that is, events in the centisecond to second range (cf., Allan, 1979; the Steven's Law exponent for duration is 1.1 [Schiffman, 1982]). This is important because discussions of the temporal features of musical stimuli tend to assume a linear relationship between physical reality and perception.

In music theory and music cognition, discussion of the time pattern of notes has generally fallen within the constructs of meter and rhythm. However, neither of these terms captures the relationships between the temporal aspects of adjacent notes and neither of these is linked to temporal pattern as it is used by Lerdahl and Jackendoff (1983).

Meter is a fairly straightforward construct: It is the regular pattern of strong and weak beats that is notated in the time signature (e.g., 4/4, 3/4, 2/4). Meter is an abstraction that arises from systematic changes within the surface structure of a piece of music. That is, meter is not the duration of notes or rests: Meter is not the timing of note onsets. A beat is perceived when the surface structure of a piece of music provides periodic emphasis (accents is the usual term in music) to musical events. In Lerdahl and Jackendoff's terms, this emphasis is called a phenomenal accent. Accents may be any (acoustic) feature that can give emphasis, such as changes in intensity, changes in timbre, changes in the relative timing of notes, changes in the pitch region of notes (i.e., relatively larger jumps in pitch): some writers include tonal centers with acoustic accents, but as stated earlier, Lerdahl and Jackendoff distinguish phenomenal (acoustic) and structural (tonal) accents. The specific type of accent does not have to be repeated, for the beat to be perceived, though undoubtedly this makes the perception of beat easier (see Lerdahl & Jackendoff, 1983, for a more extensive discussion). Periodic is to be taken in the absolute sense of time. A beat (weak or strong) does not have a duration, though it does have a temporal location. This conception is consistent with Lerdahl and Jackendoff (1983) as well as many others (Christ, DeLone, Kliewer et al., 1966; Green, 1965; Handel, 1993). As such, meter forms the time base of the music and it is likely that meter guides attention in music (important events occur on strong beats; see Boltz, 1991a). Although meter is a temporal pattern, meter is, in fact, a static structure: Meter does not change as a function of time. This distinction is often lost in work in music. It is true that in some songs the meter changes, but this is usually explicitly notated by the insertion of a new time signature.

Rhythm is a much more difficult concept to pin down. The term "rhythm" seems to be used for all levels of music and, in fact, has been implicated as a fundamental aspect of life (cf. Handel, 1993, p. 383). Despite its ubiquitous use in music, rhythm remains one of the most poorly defined terms of that field. Handel (1993, p. 399) notes that it "seems intuitive that different levels of analysis [of rhythm] might require different perspectives." Unfortunately, in a proper consideration of the parsing of melody, one cannot ignore rhythm; therefore, some attempt must be made to define rhythm. In addition, many studies use the term rhythm to define aspects of their stimuli and design. However, the lack of a single definition of rhythm makes it very difficult to compare these studies to each other and to various theories of parsing. For example, many studies use the term rhythm when it might be more appropriate to use the term meter.

Rhythm is not the same as meter, thought the two are related. Meyer (1956, p. 105) has said, "Rhythm and meter, though obviously intimately interrelated, are nevertheless independent variables. This will be evident as soon as one considers that several different rhythms can arise within the same metric organization . . ." Lerdahl and Jackendoff (1983) also cleave meter from other aspects of music; meter is rigidly defined and used to develop a hierarchy of beats. The hierarchy of beats is separate from the hierarchy of note units (groups). The hierarchies of beats and note units (groups) are considered to be parallel structures of music. Although Lerdahl and Jackendoff used the term rhythm repeatedly, they did not provide a definition that I could find. To maintain a simple distinction, one might say that, at a local level, meter is the pattern of beats, while rhythm is the content (of note and rest durations) at each beat (see Meyer, 1956, p. 105 for similar comments).

Building on this, for this work, rhythm was considered as the pattern of change in other aspects (e.g., meter, pitch pattern, time pattern, intensity, tempo) of music, and in particular, the temporal structure of the pattern of change in other aspects of music. In this definition, one can see a clear distinction between meter and rhythm and time pattern and rhythm. Rhythm is the result of the interaction of various aspects of music including meter. This is compatible with Handel's (1993, p. 399) statement, "...the interplay between meter and grouping [unitization] determines rhythm." This notion is the foundation of Lerdahl and Jackendoff's (1983) separation of metrical and grouping (the aforementioned hierarchy of units) structures of music. It must be admitted that this view of rhythm seems to counter that which has been advocated by Jones and coworkers (e.g., Boltz, 1989, p. 755, 1991a, pp.241-243, 1991b, p. 423; Jones, Summerell & Marshburn, 1987, p. 93; see also Lee, 1985) where rhythm is considered the time invariant pattern of something[14]. Their rhythm is akin to the notion of meter herein. However, this presentation is very similar to that offered by Lerdahl and Jackendoff (1983) though some terminology has been altered. Meyer (1956, 102-127, particularly, 102-103) also advocates such a distinction between rhythm and meter. It is important to further refine the notion of rhythm so that its role within parsing can be defined. Rhythm is the perception of the temporal pattern (not necessarily periodic) of motion or change in other dimensions of music. Hence, at a global level, rhythm could be the perception of a slow, then fast, then slow tempo, or the perception of a quiet, then loud, then quiet intensity, or even the combination of slow-quiet, then fast-loud, then slow-quiet. Rhythm can also be a series of phrases each having a typical arch contour. At a local level, rhythm could be the recognition of the repetition of a 2/4 meter followed by a change to a 4/4 meter and a return to a 2/4 meter. Rhythm can arise from the recognition of rapid sudden changes or from the recognition of gradual changes in tempo or intensity (i.e., accelerando followed by ritardando or crescendo followed by decrescendo). This is also consistent with normal usage. Changes such as these were cited by Krumhansl (1996) as the basis for the perception of new musical ideas within a piece of music.

The critical point for parsing is that rhythm does not determine the content of an attribute at any moment in time. Rhythm is the manner in which an attribute has changed over time. To the extent that many attributes change in complementary fashions, one has a strong percept of rhythm. (Tacitly, one must recognize that the lack of change in an attribute is an aspect of rhythm because time moves continuously forward. However, it is arguable that the perception of no-change is less salient than the perception of change.) Critically, for the purpose of parsing a melody, *rhythm has no function.* If a melody is parsed into units, then rhythm would arise from the recognition of changes (or lack of changes) in the content of those units. However, rhythm would not specify the content of those units or for that matter, the location of boundaries between units. The unitization

---

[14] It is difficult to compare different constructs of rhythm, but Boltz and Jones seem to most often argue that the perception of rhythm requires temporally regular accents within the melody. Hence, rhythm is regular and predictable.

would be based on other attributes of music. Rhythm cannot be both the mechanism of parsing and the pattern of unitization: That is circular (i.e., saying that the rhythm demands that the song be parsed at these points and then saying that the rhythm arises from that pattern of parsing).

Hence, to summarize the considerations of time in music (as it pertains to parsing), meter is considered the static pattern of beats. Meter arises from regular accents in the surface structure of music. Meter is not a part of this review because this work does not consider melodic parsing on the basis of meter (this is because Lerdahl and Jackendoff have separated meter from other structures of music; see their discussion for more details supporting the validity of this approach). Time pattern is considered to be the relationships between the timings of adjacent notes, including the timing of notes, the durations of notes and interstimulus durations between notes. Onset to onset durations are also important, but these arise from the other features. This discussion is consistent with Lerdahl and Jackendoff's (1983) use of these features, and it allows for an examination of parallelism based on time pattern (see Chapter 4). Time pattern is critical for parsing. Rhythm is considered to be the perception of change in other aspects of music. Rhythm has no role for parsing (in this conception). This allows the subsequent review to examine time pattern for its role in parsing while ignoring meter (because it is a separate structure within Lerdahl & Jackendoff's theory) and rhythm (because it has no role). However, because rhythm and meter are ambiguously used in a manner that fits within the notion of time pattern, it is important to verify the content of all studies reviewed.

The previously discussed experiments of Palmer and Krumhansl (1987a) explicitly tested phrase boundary formation (utilization) on the basis of pitch contour and time pattern. Subjects were asked to rate the goodness or completeness of a short musical segment (based on Bach's Fugue XX in A minor (from the *Well Tempered Clavier*, Book 1). The segments either ended on a phrase or at some point within a phrase (analysis by Lerdahl, based on Lerdahl & Jackendoff's theory, 1983). High ratings of completeness were associated with phrase endings. For present purposes, the important point is that the melody was decomposed into two components: the pitch pattern and the time pattern (called pitch and temporal conditions, respectively). The pitch pattern retained the pitch of all notes but made all notes equitemporal (rests were removed). The time pattern retained the durations of all notes, but rendered all notes equtonal (rests were retained). Ratings of completeness for each of these conditions were taken. Both conditions produced higher ratings near phrase boundaries and more interestingly, the sum of the ratings of completeness for pitch and time information predicted the overall rating for the melody. The quantitative ratings of the tonality of the notes within the defined key (cf., Krumhansl, 1990) were correlated with the pitch-pattern ratings but not with the time-pattern ratings. The second experiment of the study replicated the results with a slightly different method. These results implicated both pitch and time pattern in phrase recognition. Palmer and Krumhansl (1987b) essentially replicated the findings with somewhat more complex stimuli. However, these experiments do not explain how or why pitch and time pattern relate to phrase endings. It is true that the results correlated with the analyses of Lerdahl (based on Lerdahl and Jackendoff, 1983), but those analyses were time-span reductions; as such, the analysis of Lerdahl did not explicitly separate the various effects (pitch pattern, tonality or time pattern) from each other. Krumhansl did attempt to separate the effects of pitch and time patterns, but the analysis did not identify the elements of the pitch and time patterns that were associated with boundaries (e.g., large pitch changes? longer notes?). The most important conclusion was that time pattern (rhythm) and pitch pattern seemed to be separable additive components of music.

Jones, Summerell and Marshburn (1987) explicitly compared the role of pitch and time pattern processing in music. Based on common musical themes, they constructed 12 note sequences in the key of F, using one of three time patterns. In actuality, though they used the word rhythm to describe these patterns, this is an example of a situation where the term meter might have been more appropriate (the time patterns were fixed periodic patterns, that would automatically evoke a metrical structure; see Handel, 1993, pp. 386-390). The three time patterns were SLLrSLLrSLLrSLLr, LSSrrLSSrrLSSrrLSSrr and LLSrLLSrLLSrLLSr (where S is a short note, L is a long note, and r is a rest, equivalent in duration to the short note: short notes were half the length of long notes). There were three pitch patterns defined as the simple up-down pattern of pitch motion (i.e., not the pitches per se, and no consideration of interval size, though all sequences were somewhat tonal, and completely diatonic). These patterns were ---++++----. --++-+++--- and -++--++-++-. differing in the number of contour directional changes (2, 4 and 6 respectively). Note that the combination of pitch and time patterns would create accents at particular locations. For some combinations, the accents would reinforce, and for others they would conflict. In the first phase, subjects were trained to recognize a subset of the sequences and then, in the second phase, subjects were tested for their recognition memory in the presence of distracters (subjects were asked to attend to the "melody" or pitch pattern, supposedly ignoring the time pattern). Distracters preserved pitch pattern (contour) or time pattern. Subjects erroneously confused foils that altered either the pitch pattern alone, or the time pattern alone (even though subjects were not supposed to attend time pattern). Foils that altered both pitch and time patterns were the least confusable. Jones and Raltson (1991) essentially replicated these findings with longer sequences and more strict control over the concept of pitch pattern (the function of notes within the key was addressed explicitly; the notion of the up/down pattern was de-emphasized). The stimuli of Boltz and Jones (1986) were very similar. They constructed diatonic (or tonal) melodies that exhibited different melodic patterns. Superimposed on these patterns was one of two rhythms (or meters). The quadruple rhythm used notes of 300 ms, with an interstimuls interval of 150 ms (120 ms in Experiment 2). However, after every third note, the interstimulus interval was changed to 600 ms (540 ms in Experiment 2). Effectively, groups were note, note, note, rest. The sextuple rhythm used notes of 300 ms, with interstimulus intervals of 100 ms, changed to 900 ms after each fourth note (in both experiments), creating groups of note, note, note, note, rest, rest. Subjects (all music majors) were requested to notate the melodies. The results were essentially the same: Performance depended on the interaction of time pattern and pitch pattern. Compatible time and pitch patterns were better remembered.

In a similar design, Boltz (1991a) also addressed the interaction of meter (calling this rhythm) and pitch pattern. Musicians were asked to recall unfamiliar folk tunes after hearing one presentation. In the presentation of some songs, the tonal markers of the phrase endings (i.e., notes of the tonic triad marking the cadence) were replaced with notes that were contour preserving and diatonic, but not of the tonic triad (however, it should be observed that in some songs, the mediant was replaced by the subdominant). In addition, songs were presented either with meters (rhythms was the term used) that highlighted the phrase ending (i.e., meters that were compatible with the harmonic structure) or with meters that obscured the phrase endings. None of the meters actually altered the timing of notes: metrical emphasis was agogic. Pitch pattern errors were classified into three categories: missing notes, contour preserving alterations and contour altering alterations. When meter was compatible with the tonal structure, contour-altering and missing-note errors were rare. However, contour altering errors were common. In the other conditions (incompatible meters and/or altered tonal centers), all types of errors were equally common.

Kidd, Boltz and Jones (1984) also found main effects and interactions between time patterns (called rhythm: these were not metrical) and pitch patterns. The analysis did indicate that time pattern differences were quite obvious, but further interpretations are difficult because, although the pitch patterns were diatonic (not tonal), the time patterns were distinctly odd. The regular time pattern consisted of notes uniformly 150 ms in duration, with an interstimulus interval of 50 ms (akin to a series of quarter notes each followed by something close to a dotted 16th rest); the other time patterns could only be described as highly irregular, having notes either 100 or 150 ms (this is not a simple 1:2 ratio typical of music, but more like a quarter and dotted quarter) with interstimulus intervals of 20, 50, 100 or 150 ms (almost sixteenth, eighth, quarter and dotted quarter rests, if one ignores interstimulus intervals). It would seem (though the results were not presented in a manner that allows one to be deterministic) that comparisons involving two regular time patterns would be very different from comparisons that involved either a regular and irregular time pattern or two irregular time patterns. The regular patterns would seem musical, while the irregular would seem very unmusical (the results of Experiments 2 & 3 seem to imply such domain distinct processing).

Generally, the previous findings implicate pitch and time pattern as factors in musical processing though the results do not generally speak directly of the question of parsing. Other researchers have replicated effects on pitch pattern and time pattern alone.

Jones, Summerell and Marshburn, (1987, p. 95) based their notion of pitch pattern on that of Dowling and Fujitani (1971). Much of the work in this area (pitch pattern) seems to have begun with Dowling and Fujitani's premise:

> Melodic contour (the sequence of ups and downs in a melody, regardless of interval size) expresses those aspects of a melody that are most essential to manipulation of that melody in various musical structures, e.g., folk tunes and fugues.
>
> 1971, p. 524

It is important to be clear about what is implied by this definition; Music, or at least, much of it can be reduced to an absolute judgement of up or down in pitch space regardless of the amount of up or down, regardless of the absolute pitch sizes. Note that absolute frequency, categorical pitch, diatonicism, tonality and harmony are not considered. Two phrases (or melodies) would be recognized as the same if they shared the same contour.

Contour, so defined, has been tested in many different paradigms. Most studies examine recognition memory for recently presented stimuli. Dowling & Fujitani (1971), Dowling (1972), Dowling (1978), Bartlett and Dowling (1980), and Dowling and Bartlett (1981, Experiments 3 & 4) have all demonstrated that two sequences having the same contour, defined as simply the up/down pattern regardless of interval size, are easily confused. However, the results are complicated by the variable use of the degree of tonality of the stimuli (i.e., atonal, diatonic and/or tonal stimuli). The restriction to either diatonic or tonal sequences seems to improve the discrimination of contour. Cohen and Frankland (1990) required subjects to indicate the contours of eight tone sequences on a grid (temporal-auditory to spacio-visual translation) and found that generally, more complex contours and more atonal pitch sets generated more errors. On a larger scale, Cohen, Trehub and Thorpe (1989) presented subjects with three presentations of the same five-note sequence, and then, on the fourth presentation of that sequence, asked subjects to judge whether or not that fourth presentation was the same as the previous three. When different, the fourth presentation changed the third note by one semitone (the contour of the five-note sequence was unchanged). The important parameter was the complexity of the macro contour linking the four presentations together. Complex macro contours produced lower performance, as did the use of augmented triads. Idson and Massaro (1978) and Kallman and Massaro (1979) asked subjects to recognize various distortions of familiar folk tunes. Distortions that preserved contour were better recognized.

It is important to realize that in the course of these studies there was a general historical trend away from atonal sequences (e.g., Dowling & Fujitani, 1971, Dowling, 1972) to diatonic or tonal sequences (e.g., Boltz, 1991a; Boltz & Jones, 1986; Jones, Summerell & Marshburn, 1987; Jones & Raltson, 1991). Studies that have explicitly compared levels of tonality have consistently found different effects for different degrees of tonality with contour discriminations being easier in tonal sequences (e.g., Cohen & Frankland, 1990; Dowling, 1981; Dowling & Bartlett, 1981). The implication is the interval information is critical to the definition of contour. In a more recent review, Dowling (1991) stated:

> The fact that tonality affects judgements of melodic contour indicates that contour is not an entirely separable feature of melodies but rather that a melody with its contour constitutes an integrated perceptual whole.
>
> 1991, p. 305.

Bartlett made the much stronger statement, "Despite the role of contour in aiding recognition, contour by itself is virtually useless as a cue" (1993, p. 57). This idea has important ramifications for anyone attempting to consider contour in subsequent musical analyses. Contour, it seems must be defined for changes along a diatonic scale (not any pitch scale) or possibly for changes along a tonal (triadic) scale. This is relevant to the definition of pitch pattern that is necessary for explorations of Lerdahl and Jackendoff's theory (1983).

In an extension of other work, Dowling and Bartlett (1981, Experiments 1 & 2) were surprised to observe that an excerpt from classical music could not induce a feeling of familiarity in a second excerpt that shared the same contour. Dowling and Barlett attributed the perceived difference in the two stimuli to the imposition of a 5-minute delay between the presentation of the initial stimulus and its later comparison. Dowling (1991) tested this hypothesis using the same recognition paradigm with delays of 0 and 39 s. It was found that although contour was important at the short delay, contour information was not retained at long delay. There was also an interaction with tonality (atonal, diatonic and tonal sequences). Earlier, Edworthy (1982, 1985) used a different paradigm but produced the same conclusions. After presentation of a diatonic standard, subjects were asked to detect interval or contour changes in a comparison sequence (subjects indicated the difference as soon as it was detected). With longer sequences (hence longer times), subjects were more accurate with interval changes, but less accurate with contour changes. The implication is contour information is immediately accessible, but not resistant to decay while interval information is not immediately available, but resistant to decay.

Bharucha and Pryor (1986) asked subjects to discriminate two seven-tone (isotonal) sequences in a same-different task. The critical factor was the construction of the time pattern of the sequences. Sequences could be either rhythmic (could be heard as a rhythm on a duple meter) or disrupted-rhythmic (no simple pattern). For Bharucha and Pryor, the important finding was that subjects could more reliably detect the difference when the rhythmic preceded the disrupted-rhythmic than vice versa. However, for phrasing, there are several implications. Percent correct was 0.81 when the rhythmic sequence was repeated. Accuracy dropped to 0.69 when the disrupted-rhythmic sequence that was repeated. Recognition of the difference between the two types was only 0.52 and 0.41 (rhythmic, then disrupted rhythmic was the higher value). The implication is that listeners are not particularly good at recognizing a repetition of a rhythm or noting a change in rhythms (there were no other distractions) even when the difference is fairly obvious and the time between presentations is very short. For parsing, the implication is that rhythm is not a particularly salient phrase marker (however results should not be overgeneralized on the basis of one study).

It is the idea that pitch pattern information is immediately available but not resistant to decay that has important consequences for the role of pitch pattern in the parsing of melodies. This might seem to mitigate against any role for pitch pattern in the delineation of phrase boundaries. If pitch pattern (even the simply version of the up/down pattern) is not remembered, then it cannot be important. If pitch pattern is not remembered, how can various patterns be remembered for subsequent comparison either within a song (parallel phrase structures), or within the corpus of music (typical phrase structures)? The same can be said of time pattern (rhythm). If changes in time pattern are not detectable at short time intervals, then how can time pattern serve as a structural point.

However, on closer inspection, these conclusions may not be reasonable. Time and pitch patterns might be important, at the moment of listening, for their roles in affective arousal (or other "meaning"), but after the goal has been achieved, only memory for, or the lingering effects of, arousal might be retained for future comparisons. That is, subjects remember how they felt, not what produced the feeling. Hence, one might expect composers to understand the role of time and pitch pattern in music (explicitly or implicitly), but one cannot argue that listeners understand the cause of their arousal. The analogy from language is the notion of memory for gist. Subjects rarely remember the exact words of a particular passage, but they are very good at remembering the gist (cf., Clack & Clark, 1977). Writers know how to use words to create the meaning, but readers rarely remember the text literally. Pitch pattern has been empirically associated with affect. For example, Krumhansl (1996) noted a correlation between the pitch contour of music and the empirically measured tension induced by the music. It is possible that pitch and time pattern are not a defining characteristic of the beginning or end of the phrase, but rather that the pitch and time pattern (among other things) is the content of the phrase. It is possible that many phrases demonstrate similar overall pitch or time patterns (particularly, the general contour or arch) and endings, but this is a result of the desire to create similar affects: That is, pitch and time pattern is not a mechanical design factor. Hence, when considering the parsing of melody, one can probably treat pitch and time pattern per se as secondary (at least in the initial studies). That is, the notion of the arch contour, or the inverted arch contour or the axis contour is probably not too helpful in predicting where boundaries might be located, at least, not beyond the associated predictions created by the need to return to a place of tonal stability. However, pitch and time pattern are probably critical for the affective arousal that is characteristic of music.

On the other hand, it might be premature to completely discount the role of pitch pattern (contour) per se. Pitch and time pattern simplicity are implicated in simpler songs (cf., Lerdahl & Jackendoff, 1983, p. 67), and pitch pattern complexity is associated with processing efficiency. Pitch contour complexity (number of up/down pattern changes) has also been implicated in number of empirical studies as a factor in the processing of information within of contours (Boltz & Jones, 1986; Cohen & Frankland, 1990; Cuddy, 1993; Cuddy & Cohen, 1976; Cuddy, Cohen & Mewhort, 1981; Cuddy, Cohen, & Miller, 1979; Cuddy & Lyon, 1981; Warren & Byrnes, 1975). Again most of these studies implied that the simple definition of pitch pattern as the up/down pattern was not sufficient: There were statistical interactions with degree of tonality and/or time pattern (defined as rhythm and/or meter). Polanka (1995) has noted that musically trained subjects tended to use larger saccades (hence bigger units) when reading simpler musical passages. Simplicity was related to pitch pattern complexity (subunitization and hierarchical structure) as well as tonality. The point is that pitch pattern and pitch pattern complexity is associated with affect and that pitch pattern complexity is associated with processing demands. Pitch pattern might not determine phrase boundaries, but pitch pattern complexity might have major impact on the subphrase structure: That is, pitch pattern complexity may be manifest in the parsing of the phrase into motives. It is possible that more complex pitch patterns induce more parsing boundaries (more unitization) as a means of dealing with the complexity. The same might be true for time pattern processing. This is consistent with Lerdahl and Jackendoff's approach to parsing at the lowest levels of their hierarchy.

The Motive

In the parsing of a melody, the phrase is generally considered the basic structure (i.e., like a sentence), but the phrase itself is composed of a series of motives (i.e., like the words of a sentence). Although pinning down the phrase has been difficult (and not completely successful), pinning down the motive is next to impossible. Christ, DeLone, Kliewer et al., (1965) offer the following insights:

> The motive is a short and distinctive melodic pattern, often characterized by simplicity of rhythm and pitch design. Because of its brevity, the motive is easily recognized and frequently plays an important role in the organization of a melody. Both pitch and rhythm produce the distinct qualities that characterized a particular motive. Either the pitch or the rhythmic structure can be the dominating factor, or both can be combined and equally important.
>
> 1965, p. 63.

Green (1965) provides us with:

> The motive is a short melodic fragment used as a constructional element. However, not every short melodic fragment is a motive. In order to act as a constructional element and thus constitute a motive, a melodic fragment must appear at least twice, *though reappearances need not be in the original form.*
>
> 1965, p. 31. (italics mine)

The New Grove Dictionary of Music and Musicians defines the motive as:

> A short musical idea, be it melodic, harmonic or rhythmic, or all three. A motif may be of any size, though it is most commonly regarded as the shortest subdivision of a theme or phrase that still maintains its identity as an idea.
>
> Drabkin, 1980, p. 648, vol 15

In considering this, it must be noted that in the same work, the corresponding definition of a phrase is:

> "...short musical units of various lengths; a phrase is generally regarded as longer than a motif but shorter than a period. It carries melodic connotation..."
>
> Sadie, 1980, p. 663, vol 17.

Lerdahl and Jackendoff do not discuss the motive per se (at least, not at any length). They prefer the generic term group for all collections of notes. However, it is arguable that the motive is a small group of notes. As such the motive must exist (if it exists) at the lowest levels of their hierarchy.

A motive is a short series of notes, somehow cohesive, somehow representing a complete, or unitized idea of music: It is the smallest unit of musical significance (cf., Anders et al, 1975. p. 40; Christ. DeLone. Kliewer et al., 1966, p. 63; Christ. DeLone. Winold, 1975, p. 49; Green. 1965. p. 31; Llyold. 1968, p. 338). The motive is typically between two and seven tones. having a distinct rhythmic pattern and/or melodic contour and/or harmonic structure. In the terminology developed herein, the motive has a distinct pitch and/or time pattern. The motive is recognizable across repetitions. even though that repetition may be transposed (raised or lowered in pitch), inverted (increases in pitch become decreases in pitch). retrograded (played in reverse order) or any combination. Each of these alterations may introduce minor or major alterations to the relationships between notes (time and/or pitch relationships) while retaining recognizability. That is. the actual pitches can change. the contour can change. the role of the pitches within the tonal structure of the piece (locally or globally) can change. and the timing can change but the motive is still recognizable. Such a loose definition makes the processes of creating an algorithm to extract motives very difficult.

Some work has been done on the recognizability of motives. Dowling (1972) asked subjects to compare two five-note. atonal sequences presented sequentially (2 sec. delay). The comparison sequence underwent either a retrograde. inversion or retrograde and inversion transformation. Presentation was either fast (each note was 200 ms) or slow (each note was 500 ms). In addition. different groups of subjects were requested to base their judgements on either the contour and intervals (exact match), or on the contour only (contour match) of the original sequence. Retrograde inversion sequences were only recognized as transforms when the task was contour match only at the slow presentation rate. Retrograde sequences were recognized as transforms in all conditions except fast exact matches. Inversions were recognized as such in all conditions. However. given the simplicity of the stimuli and design (only 5 notes. with only 2 s between presentations). the fact that the best performance was only 0.80 implies that subjects do not generally recognize such transformations as the same entity: In music. this would have to be an automatic process. and yet, subjects perform very poorly even when attention is directed at the task. However, it must be observed that the sequences were atonal.

In Dowling's work. subjects did not recognize transforms when instructed to do so. Somewhat paradoxically, subjects often seem to be unable to detect differences when instructed to do so. Cuddy and Cohen (1976) presented subjects with a two-alternative forced choice recognition task. Subjects heard a major triad (1 of 6 sequential orders) followed by two comparison sequences. One comparison sequence was a transposed exact match and the other was an inexact transposition in which one of the tones in the triad was changed by one semitone up or down (this would not change the contour and did not change the time pattern). Subjects were asked to detect the unaltered version. An interesting point is that performance, even for the most highly trained subjects averaged about 0.90 mean proportion correct: Untrained subjects averaged about 0.60. This is amazing given the temporal simplicity of the task (only 3 notes per unit, short delays between units) and the comparison of the major triad (theoretically the most important or stable structure given 3 notes) to something that could not be a major triad. In this case, many subjects could not ignore a similar motive (no condition or group of subjects produced perfect performance). The implications are that there is a critical level of structure for the motive. In the second option, the idea is that motives are recognized on the basis of certain features and not other features. For example, if motives are recognized solely by pitch contour (no interval information), then Cuddy and Cohen and Dowling are consistent. The same contour causes confusion. Changing the contour destroys recognition. Variations in performance (individual differences and/or training) might be due to the use of supplementary information (such as interval information; cf., Cuddy & Cohen) or the recruitment of non-musical processing strategies.

As defined, the motive need not appear in its original form, may be transposed (with attendant alteration of pitch chroma and tonal function), inverted, or retrograded. Even if motives can be recognized under distortion, how are the units delineated before any such recognition can take place? What serves to bind the notes of the different motives together? Typically, references are made to caesura or light breaks in the melodic motion. Caesura are mini-cadences: That is, they are a pause in the flow of music that is not strong enough to be called a cadence. If the cadence can be treated as a period (terminal) or comma (progressive), then the caesura is the pause between words.

Unfortunately, speech does not exhibit any such appropriately placed pauses in a spectrogram (cf., Jusczyk, 1986). What delineates the caesura from the progressive cadence? If they are the same mechanism, then what quantitative point marks the border between the two? These questions are not addressed in music theory.

Many studies of phrase construction might be better construed as studies of motive construction. For example, Gregory (1978), demonstrated that, when listening to simple music-like stimuli, the perception of a secondary event (a click) tended to migrate to the theoretically-defined position of a phrase boundary. This migration depended on how subjects were instructed to parse the sequence, implying that unitization was at least partially conceptually driven. However, given that the units involved were only two or three notes, it is more fruitful to interpret the results for their implications in the subdivision of the phrase. In this case, the implication would be that motives can be delineated on the basis of pitch pattern. However, given that subjects could interpret the sequences as either two or three note groupings depending on instructions, there is an obvious top-down component to motive parsing.

There is other support. Boltz and Jones (1986), Jones and Ralston (1991) and Jones, Summerell and Marshburn (1987) all constructed tonal sequences that used repetitions of the same basic time pattern structure: three (or four) notes followed by a rest. One could argue that the listener would perceive such a phrase as a set of motives of unvarying time pattern per motive (i.e., the rhythmic structure was invariant). Sequences were designed to resemble musical phrases, particularly by Jones and Ralston, where the sequence began on the tonic, moved to the dominant and returned to the tonic. In a recognition task for previously learned material, in all studies, subjects often failed to recognize a melody as the same if the time pattern of the motives had changed, even though subjects were told to ignore such time pattern changes. The authors used these results to argue against parsing on the basis of the location of rests (i.e., as three note units; see Jones, Summerell & Marshburn, pp. 92-95,108 and Jones & Ralston, p. 18). However, this interpretation seems to confuse two issues: the chunking of the phrase into motives and the content of those motives. It is plausible that all stimuli provided the same basic pattern for chunking. Hence, differences in performance for comparisons between stimuli reflect, not necessarily the grouping structure, but rather the differences due to content of the chunks. In this interpretation, both the pitch pattern and time pattern of each motive is an important aspect of the content of the chunk. In fact, these studies seem to show that time pattern is so fundamental to the structure of the motive that subjects find it difficult to override even when instructed to do so. This interpretation does not prove the parsing of phrases on the basis of rests (caesura), but the results are not inconsistent with such notions. This is particularly true in the case of Jones and Ralston, where the most obvious distracters changed both the time pattern within the chunks, the content of some chunks, and the serial order of the chunks. The second most obvious distracters changed the content of some chunks and the serial order of the chunks. The third most obvious distracters changed the content of some chunks or the content of some chunks with the time pattern of the chunks. The least obvious distracters changed only the time pattern of the chunks (which subjects were told to ignore). That is, performance was lowest when given the fewest changes both between and within chunks (this pattern is duplicated in Jones, Summerell and Marshburn).

Cohen, Trehub and Thorpe (1989) explored the sensitivity of subjects to hierarchical structure, particularly the macro-contour of five note sequences based on either a major or augmented triad that were presented to subjects three times. For the fourth presentation, the sequence could be repeated exactly or have a one-semitone change on the third note (this did not change the contour of the 5 note sequence). Subjects had to detect that change. The important parameter was the relationship between the three repetitions (4 counting the final test presentation) of the five-note sequences. In some cases, the first note of these repetitions formed a simple contour (ascending or descending) while in other cases, these repetitions formed a complex contour. Subjects, with and without musical training, exhibited much better performance with the simple macro contours (and major triads) implying that subjects did utilize the structural information both within and across motives. However, although this is consistent with the use of motives, it does not necessarily demonstrate the use of motives.

Further support for the notion of parsing sequences into motives can be found in work that tries to relate performance on various pitch or time patterns to structured, usually hierarchical or linear, rule systems (i.e., Deutsch & Feroe, 1981; Lerdahl & Jackendoff, 1983; Restle, 1970). All these rule systems assume that musical sequences can be segmented into groups of musical events: Generally the basis for such segmentation is the Gestalt principles (in some form) of proximity (in pitch space or time) and similarity (in pitch space, tonal function, or time). Unfortunately, many such systems are vague with respect to the details of such segmentation and as such result in a large number of substantially different interpretations for the same sequence of notes (space precludes a demonstration of the ambiguity in these system; a fair assessment would require a thesis length exposition for each, as has been done herein for Lerdahl & Jackendoff). In an exploration of the roles of hierarchical structure verses linear (non-hierarchical) structure verses no structure, Boltz and Jones (1986) found that any type of rule helped retention and recall of diatonic (arguably tonal) melodies better than no rules at all. However, performance with hierarchical rule systems were no different than linear rule systems, and contour complexity and meter seemed to be the best overall predictor of performance. This led Boltz and Jones to discard the notion of rule-based system and to replace it with a notion of the accent model. In the accent model, groups of notes are still created, but there is no (necessary) relationship between the different groups. Hence, although they discarded the notion of rule-based structure, they did not discard the notion of motives. In their model, in a sequence of notes, there are temporal accents created by changes in duration (lengthened or shortened notes and rests) and there are melodic accents created by changes in pitch contour (large jumps or contour changes). These accents can reinforce or interfere, but together they determine the parsing of the melody into motives. This system is in fact, very similar to that which has been offered by Lerdahl and Jackendoff.

The general implications from the theoretical and empirical literature is that motives *might* exist as a perceptual/cognitive structure in the listener (as opposed to, or in addition to, their use as a constructional device by composers). These motives are delineated on the basis of both pitch and time pattern. This is consistent with the model of Lerdahl and Jackendoff (1983) though it does not prove that the model is correct.

Summary

This work is concerned with the empirical assessment of boundary formation in melody, using Lerdahl and Jackendoff's (1983) *A Generative Theory of Tonal Music* as the theoretical framework for explaining and discussing the empirical observations. Melodies were chosen as the stimuli because melodies should contain the structures of western-tonal music and melodies should be processed as music. Lerdahl and Jackendoff's theory was chosen as the framework because it provides a quantifiable basis for many structures, because it is representative of music theory in general, although there are differences and because it attempts to analyse music at all levels. In addition, the basic constructs of the theory do seem to have some support within the literature of experimental music cognition.

However, the theory is not complete. It does require quantification (see Chapters 2 and 4) and there are several aspects that need considerable clarification. Because there is some ambiguity in the concepts at the intermediate levels, the present work does not try to progress beyond the intermediate levels. That is, modelling is restricted to the testable low-level rules for parsing, and modelling is extended to the intermediate levels of the hierarchy. Modelling does not attempt to address the higher levels of the hierarchy. The quantification of the low-level rules is straightforward (see Chapter 2). This quantification relies on the pitch and time patterns of note events.

The quantification of the intermediate level rules (combinations of low-level rules, parallelism and symmetry) requires that a number of constructs be defined a priori, particularly parallelism and symmetry. Since these constructs depend on the notions of pitch and time pattern, they are discussed after some empirical data has been discussed in Chapter 4 (i.e., after some issues pertaining to pitch and time pattern have been resolved).

Generally, the approach is to determine the locations of boundaries within simple melodies by a direct empirical test (Chapter 3). These empirically determined boundaries are then compared with the predictions of Lerdahl and Jackendoff (1983), after appropriate quantification (Chapter 2). As knowledge accrues, more complex aspects of Lerdahl and Jackendoff's theory can be tested: that is, Intensification and Parallelism (quantified in Chapter 4, and tested in Chapters 4 and 5). Empirical feedback and intermediate-level rules should provide increasingly accurate predictions of empirically determined boundaries.

This leaves only one general issue to be discussed: the population of choice. Since the intent is to model the perception of music, the population is all people who listen to music written in the western-tonal idiom. Hence, this study is not restricted to only those with high levels of musical training (cf., Handel, 1993, p. 381; Kraut, 1992, p. 17-18). This is somewhat at odds with the theory of Lerdahl and Jackendoff, which claims to model the listening process of a sophisticated listener. However, the use of simple melodies should negate such discrepancies. More importantly, the intent is to model the basic structures of western-tonal music. To use the analogy with language one last time, if one is interested in studying language processing (or visual processing), one should examine the use of language in the average person, not the use of language in those few individuals who are professional writers. This does create one problem, however. It is possible that musical training will affect musical processing. For example, the internalized representation of tonality is affected by training (cf., Frankland & Cohen, 1990; Krumhansl, 1990), and tonality may affect the perception of the pitch pattern. This means that these explorations must consider musical training. In the present work, an attempt is made to assess musical knowledge both by questions of background and by behavioural test that are known to correlate with knowledge, so that the parsing of melody can be assessed across the spectrum of individuals who are involved with that music.

CHAPTER 2

Quantification of the Group Preference Rules of Lerdahl and Jackendoff

Lerdahl and Jackendoff (1983) have produced one of the few theories that builds from the level of the musical surface (as an auditory sequence of sounds that change in pitch, duration, intensity and timbre) to the level of the global organization of the piece as music. Moreover, it is, as theories go, fairly explicit about the processing that occurs at each level.

In this section, the basic theory is reviewed (see also West, Howell and Cross, 1985, for a review). All concepts cited herein are taken directly from their work, except those specifically notated as such. The intent is to review the essential elements of the theory so that it can be appropriately quantified. Those who seek more detail can consult the original texts (Lerdahl & Jackendoff, 1983). In this theory, it is assumed that:

> the listener naturally organizes the sound signals into units such as motives, themes, phrases, periods, theme-groups, sections and the piece itself. . . . Our generic term for all these units is *group*. At the same time, the listener instinctively infers a regular pattern of strong and weak beats to which he relates the actual musical sounds.
>
> Lerdahl & Jackendoff 1983, p. 12

The grouping structure that arises from the organization of units is considered independent from the metrical structure that arises from regular pattern of strong and weak beats, although both are hierarchical in nature. The additive combination and the interaction of the two produces the rhythmic structure of the piece (1983, p.12) [15]. The basis for both structures is the phenomenal accent:

---

[15] Lerdahl and Jackendoff (1981, p. 486) point out that grouping structure is more universal than metrical structure and some styles such as Gregorian Chants and the alap (opening section of a North Indian raga) do not have metrical structures. Some styles have much more complex metrical structures than are found in western tonal music.

> any event at the musical surface that gives emphasis or stress to a moment in the musical flow. Included in this category are attack-points of pitch-events, local stress like sforzandi, sudden changes in dynamics or timbre, long notes, leaps to relatively high or low notes, harmonic changes and so forth.
>
> 1983, p. 17

Phenomenal accents are the basis for metrical accents which are "any beat which is relatively strong in its metrical context" (1983, p. 17). Phenomenal and metrical accents are distinct from structural accents which are "an accent caused by the melodic/harmonic points of gravity in a phrase or section -- especially by the cadence, the goal of tonal motion" (1983, p. 17). These different accents are interrelated (the same sequence of notes gives rise to all of them), but they are seen to serve or define different purposes. For this work, it is the phenomenal accent that is of most interest because it is the basis for low-level group structure (as well as metrical structure). The phenomenal accent is essentially, an analysis of the pitch and time patterns of the sequence of note events. It is a recognition of note-to-note relationships based on acoustic features of stimuli.

Because group and metrical structure are viewed as independent in principle, it is possible to focus on only one[16]: The focus of the present thesis is the grouping structure of Lerdahl and Jackendoff (1983). As stated, the grouping structure is hierarchical: The highest level (the largest units) is the entire piece and the lowest level (the smallest units) is the motive. Ideally the grouping structure is strictly hierarchical, but Lerdahl and Jackendoff (1983, p. 13) do acknowledge that there may be low level deviations (e.g., elisions) from the notion of a strict hierarchy. These deviations are limited to a short span of note events or time and such deviations can only occur under highly constrained situations. For this reason, they are treated as exceptions or special cases, rather than as a central component of the theory. Apparent deviations at higher levels involving longer spans of note events or time would be considered alternative interpretations, rather than a

---

[16] Even in western tonal music, the careful control of metre (the timing of notes) should allow one to minimize the influence of the metrical structure, thereby emphasizing the grouping structure. This is the tactic taken within the present work.

non-hierarchical organization.

Five Grouping Well-Formedness Rules clearly explicate this notion of a hierarchy, without imposing any other (musical) limits on the actual construction of that hierarchy. Intuitions of musical structure are captured within a number of Group Preference rules. The term preference is intended as a mnemonic to remind one that:

> the rules establish not inflexible decisions about structure, but relative preferences among a number of logically possible analyses; our hypothesis is that one hears the musical surface in terms of that analysis (or those analyses) that represent that highest degree of overall preference when all the preference rules are taken into account.
>
> (1983, p. 42)

On the basis of the context of the definition of their rules, it seems that Lerdahl and Jackendoff do not believe that individuals exert preferential control over the application of the rules. Rather, the rules tap natural (instinctive?; they argue that they are universal, cf., 1983 p. 36; see also Note 2 for Chapter 3, p. 336) elements of human information processing but application of those rules at any particular point, or over some span of, the musical surface may conflict with one another: the "preference" is an attempt to determine which rules will predominate, even in the face of musical ambiguity. Also, though they may be universal, it can be argued that for different listeners, the degree of development or instantiation of the rules may vary.

The first three Group Preference Rules govern the unitization of the piece at the lowest levels of the hierarchy (following Lerdahl & Jackendoff, 1983, I will use the mnemonic GPR for Group Preference Rule). These three GPRs depend on the acoustic properties of the sequence of note events (a generic term to cover notes, rests and other audible musical events). At intermediate levels, there are four more GPRs that define how the lowest level groups should be combined. At still higher levels there is time-span reduction (a Schenkerian-like [1983, p. 106] analysis of pitch and harmonic relations) and prolongation reduction (analysis of tension and relaxation).

Ultimately, the unitization of the piece is governed by the first three GPRs (Group Preference Rules): Note events are grouped in accordance with the gestalt principles of proximity and similarity (called change by Lerdahl and Jackendoff, 1983). Although it is not stated explicitly, it can be assumed these lower level groupings are related to phenomenal accents in that proximity and similarity are essentially defined by the lack of such accents (i.e., those events that are identified as proximity and change are also classified as phenomenal accents; 1983, p. 17; cf., Agmon, 1990, p. 303). At a level beyond the local detail, it is the next four GPRs that determine how the low-level units are to be combined.

Note that in the present work, I will be bouncing between the concept of a unit and the concept of boundary formation. The two are really the same thing in that boundaries are placed between units. As one moves up the hierarchy, one combines units (higher levels represent combinations of units) or one retains boundaries (a boundary from a lower unit that is not retained implies that the units it delineated were combined).

Even though it is not the focus of this work, it is important to understand the role of the metrical structure within the theory of Lerdahl and Jackendoff (1983), if only to understand why it can be, or how it should be, delineated from the grouping structure. The basis of the metrical structure is the beat, an idealized non-durational point in temporal space[17]. There is a finite amount of time between beats: the time-span. Beats are periodic: That is, they occur at relatively regular intervals of time. Hence, for any piece of music, the minimum time-span must be the shortest time between the attack points of any two adjacent notes. Metrical structure is the organization of beats into a periodic pattern of strong and weak beats. Relatively stronger beats occur at/on note events that coincide with a phenomenal accent. In essence, phenomenal accents that occur at some integer multiple of the minimum time-span are metrical accents. Metrical structure is the organization of beat and metrical accents into patterns of relatively stronger and weaker beats. At the lowest level, every minimum time span is delineated by a beat. By virtue of metrical accents, some of these beats will be relatively stronger. These stronger beats at

---

[17] Lerdahl and Jackendoff's (1983) analysis of the metrical structure is not unique to their work. For links to similar frameworks, see their text. Because such links are not central to the goals of this work, they are not discussed.

the first level become the beats of the second level. At this second level, by virtue of metrical accents, some beats will be relatively stronger. These stronger beats will become the beats of the third level. At the third level, by virtue of metrical accents, some beats will be relatively stronger. These stronger beats . . . and so on. In principle the metrical structure could extend all the way to the level of the whole melody (i.e., the last level would have a single beat that represents the entire melody), but in practice only one or two intermediate levels are perceptually important (the most important level is the tactus), and the regularities are most stringent at these levels (Lerdahl & Jackendoff, 1983, p. 70-74; interestingly, such levels are close to that of the human heart beat at 40-160 beats per minute [p. 73]) Since metrical structure is concerned with the perception of periodic patterns of strong and weak beats, the truth of metrical structure at the lowest levels (e.g., in the time-span of 32nd or 16th notes) and the highest levels (e.g., the time-span of themes, theme-groups or sections) are likely irrelevant except as an abstraction. At the large scale, it is likely that the grouping structure based on structural accents predominates. To paraphrase Lerdahl and Jackendoff (1983, p. 21), metrical structure is a relatively local phenomenon.

As with the grouping structure, the metrical structure is idealized as a rigid hierarchy; there are four Metrical Well-Formedness Rules to define this hierarchy. As with the grouping structure, there are ten Metrical Preference Rules that define preferred metrical structures. Interestingly, it is Well-Formedness Rule 3 that states that the spacing of relatively stronger beats at any one level should be two or three beats apart. That is, this is not a preference rule, though Lerdahl and Jackendoff do acknowledge that their metrical rules are, in part, western-tonal idiom-specific.

One important contribution of Lerdahl and Jackendoff (1983) is the explicit separation of metrical and grouping structures. As such, low-level groupings are not constrained by metrical structure, in turn, implying that the higher levels in the grouping structure are also not constrained by metrical stress. In general, in western-tonal music, the two structures will align: That is, the two will be in-phase. Slight phasing differences are the basis for specific musical structures such as the distinctions between feminine and masculine cadences or upbeats and afterbeats. Although large phase differences are possible, the fact that the perceived (i.e., important) metrical structure is limited to the time scale near the tactus negates the importance of such relations.

The main point for the inclusion of metrical structure has been to demonstrate that in the view of Lerdahl and Jackendoff (1983), the metrical structure can be separated from the grouping structure. As such, the grouping structure can be emphasized and the metrical structure de-emphasized so that parsing on the basis of the grouping structure can be seen. To consider the concept from another point of view, listeners (people in general) will attempt to parse a sequence of sounds on some basis. If there is no basis, then listeners might impose an arbitrary one (cf., Handel, 1993, p. 386). According to Lerdahl and Jackendoff, listeners can use considerations of the grouping structure and/or metrical structure. If a melody does contain information for creating a group structure, but does not contain information for creating a metrical structure, then the resulting parsing will more strongly reflect the grouping structure. That is, in the absence of metrical accents, listeners need not impose metrical structure, but can simply parse on the basis of grouping structure alone. Since Lerdahl and Jackendoff have argued for the separation of grouping and metrical structures, it is theoretically possible that in the absence of metrical structure (i.e., strong, periodic, phenomenal accents as defined by Lerdahl & Jackendoff), individuals will parse on the basis of group structure.

Of course, since both grouping and metrical structures rely on phenomenal accents, it is not possible to eliminate one or the other completely. However, the use of the metrical structure as a basis for parsing can be de-emphasized by the elimination of metrical accents (phenomenal accents that are periodic: i.e., removal of the beat in performance). Hence, if subjects are requested to parse music that lacks acoustic metrical stress, their parsing will be more reflective of the grouping structure (to the extent that metrical accents have been removed). This is distinct from other theories that emphasize metrical stress as the basis for low-level unitization. In these theories, in the absence of metrical stress, subjects will either fail to parse or impose some arbitrary metrical stress and then parse. In either case, the unitization will, in some musical sense, be meaningless (cf., commentaries on other theories by Lerdahl & Jackendoff, 1983, pp. 13-36).

At higher levels, considerations of the grouping and metrical structures merge into a time-span reduction (tonal/harmonic analysis) and prolongation reduction (tension/relaxation analysis) that work in parallel to define the global structure. At these higher levels, note events are grouped on the basis of structural accents. A structural accent is a melodic/harmonic point of gravity in the phrase or section, exemplified by the cadence that marks the end of the tonal motion. In the words of Lerdahl and Jackendoff, "Structural accents articulate the boundaries at the phrase level and all larger grouping levels" (1981, p. 500; 1983, p. 30) although the structural accent may be slightly delayed from the onset of the group (anacrusis) or slightly ahead of the end of a group (extension). Although the structural accent dominates at these higher levels, groups have been determined by the application of Group Preference rules at the lower levels: structural accents serve to organize the groups into a unified perception of the entire melody.

Experimental Realization of the Theory

It is important to remember that, at this point, the theory of Lerdahl and Jackendoff (1983) is just a theory. It is supported by a broad base of both musical intuitions (theory and practice) as well as a broad base of psychology (perception and cognition) but it still needs much more direct empirical verification. The simplest verification is to see if the principles work; that is, to see if (or more likely, when) the parsing of subjects corresponds to the predictions of the model. For example, as alluded to, one could present subjects with music lacking metrical accents to see if parsing can proceed in the absence of such accents (i.e., do the parsings make sense?, do the parsings seem to represent an imposed metrical hierarchy?, do subjects find the task impossible?). This simple idea (to "test the theory") requires a fair amount of background development in order to determine which of the various insights and predictions of the theory can be tested. This background development is necessary before issues of experimental methodology can be addressed (i.e., how to induce subjects to parse music and how to measure or analyze that parsing).

The first required development concerns the identification of the specific aspects of the theory that are currently well enough defined for experimental verification. As stated earlier, the interest is in the grouping structure, but within the grouping structure, what particular rules should be studied? The explicit formalism of the system is good for experimental analysis. The grouping structure must form a rigid hierarchy. Any (and only a) sequence of adjacent events may constitute a unit (Group Well-Formedness Rule 1), and any unit at a higher level may be divided into any number of units (in principle) at a lower level (Group Well-Formedness Rule 3), but that subdivision must be exhaustive (Group Well-Formedness Rule 5) and all lower level units must be completely contained within a single unit at all higher levels (Group Well-Formedness Rule 4). Ultimately, the entire melody is one unit (Group Well-Formedness Rule 2).

The Well-Formedness Rules do not explicate the actual structure: This implies that they cannot be studied directly. The actual structure is determined by the GPR (Group Preference Rules). At the lower levels, there are seven GPRs. (There are a corresponding ten Metrical Preference Rules.) At higher levels, considerations of the grouping and metrical structures merge into a time-span reduction with an associated two Segmentation Rules, four Time-Span Reduction Well-Formedness Rules and an additional nine Time-Span Reduction Preference Rules. In addition, there is a second type of reduction -- prolongation reduction -- with four well-formedness rules and six preference rules, that works in parallel with the time-span reduction. Complications arise within the general hierarchical structure because GPR 7 is a statement that the preferred (low-level) grouping is one that results in more stable time-span and/or prolongation reduction. Metrical Preference Rule 9 (not 10, the last rule [to parallel the GPRs], for some reason) is a statement that the preferred (low-level) metrical structure is one that results in the fewest conflicts in the time-span reduction. Both of these embody top down processes in the formation of structures.

It is a somewhat amusing comment on the circularity of the entire process that Time-Span Preference Rule 5 implies that the choice of time-span reduction should be one that maximizes the stability of the metrical structure. Time-Span Preference Rule 6 requires that the choice of time-span reduction should be one that maximizes the stability of the prolongation reduction and Prolongation Preference Rule 1 states that the choice of important events should reflect considerations of the time-span reduction. If nothing else, the theory is complicated. Also note that the presence of circularity can also be construed as a need for feedback from upper levels to the lower levels. It is only fair to point out that Lerdahl and Jackendoff (1983, p. 54) are aware of the circularity and ambiguity: They believe that, at this point, there is not enough information to properly disambiguate such effects (i.e., they know that top down processes will affect the perception of low-level units, but they do not know how much or when).

Ambiguity notwithstanding, to test parsing on the basis of the grouping structure, one must cut through the multiplicity of rules and considerations to find some point of attack. The first observation is that the time-span and prolongation reduction seem to depend on the grouping and metrical structures and not vice versa. That is, although top-down processes are acknowledged, such considerations do not constitute the bulk of the rules for either grouping structure or metrical structure. Time-span and prolongation reduction are directed to the organization of the units delineated by the low-level grouping and metrical rules. One cannot organize that which does not exist: The groups need to be defined before they can be organized. Once units have been defined, higher level principles can choose particular combinations of those units, thereby emphasizing some boundaries and de-emphasizing (effectively deleting) other boundaries.

There are two critical points that can be distilled from this. Firstly, given the overall hierarchical nature of the system, any grouping boundary at a higher level must also be a grouping boundary at the lowest level. Secondly, units at the lowest level are the fundamental building blocks of later structures, and these units are delineated by very simple rules that pertain to the surface structure.

What in the Theory Can Be Tested?

Assume that there is an experimental test in which subjects indicate the parsing of a piece of music (i.e., subjects indicate boundaries). Given a theory as complex as that of Lerdahl and Jackendoff (1983), what can those empirical boundaries be compared against? The theory predicts boundaries on a multiplicity of levels, but subjects would only indicate boundaries on one level. Which level of the theory corresponds to that of the subjects?

If subjects can produce parsings of a melody by indicating boundaries (somehow), then those boundaries must correspond to boundaries predicted at the lowest level. That is, every position that subjects place a boundary must be a theoretical boundary predicted by the GPRs (i.e., the rules that define how boundaries are formed) at the lowest level, assuming metrical considerations have been controlled. However, on the other hand, every theoretical boundary predicted by the rules at the lowest level need not be manifest as an empirical boundary placed by a subject. The reason for the asymmetry is that subjects may be parsing at an intermediate level of the system (i.e., at some level that applies to time-span or prolongation reduction, again assuming metrical considerations have been controlled). At these higher levels only some of the lowest level boundaries are retained.

What this means is, if we study grouping at the lowest level, then we must be careful to watch for misses (no boundary was predicted by the theory, but the subject places a boundary) and assign less importance to false alarms (a boundary was predicted by the theory but none was placed by the subject)[18]. If subjects indicate parsing at an intermediate structure, while theoretical boundaries are placed on the basis of low-level rules (i.e., GPRs 2 and 3), then false alarms of the theory could be common, but misses of the theory should be impossible.

Some might consider it more fruitful to attempt to match the empirical boundaries to some intermediate level of the hierarchy. Presumably, the parsing of subjects represents some intermediate level. Unfortunately, when working at an intermediate level of the theory, it would be difficult to know which intermediate theoretical level

---

[18] In this work, I consider the empirical data of subjects to be the "truth" (small t-truth). The theory tries to predict truth. Listener's responses do not try to match the theory. That is, the theory is the "active agent" trying to predict the occurrence of events in the world. Hence, a miss is an occasion when the theory failed to predict the event and a false alarm is an occasion when the theory did predict an event that did not occur.

corresponded to the parsing of subjects. In principle, one might be able to build from the lowest theoretical level, using the rules to then determine which level most closely associates with the parsing of subjects, but at this point there is simply not enough empirical evidence to support such an analysis. The key word in all the different rules is "preference": The rules are called "Group *Preference* Rules". The current level of understanding of musical cognition does not enable us to define (with any precision) which preferences apply (cf., Lerdahl & Jackendoff, 1983, pp. 52-55). This problem is inherent in the rules as defined by Lerdahl and Jackendoff. Every preference rule uses a phrase like, "all else being equal", "the transition may be heard", "a boundary may be placed", or "prefer groupings that" (cf., 1983, pp. 346-352). On occasion, a rule will state "strongly prefer" or "weakly prefer". Note, in contrast, that the well-formedness rules were explicit.

Without explicit guidelines, it is impossible to build from the lowest level to any particular intermediate level. One could generate a number of plausible alternatives, all based on the same set of rules. It is true that Lerdahl and Jackendoff (1983) provide numerous examples, and it is easy to follow their logic in those examples, but it is a very different order of analysis to generate an unambiguous (or relatively unambiguous) structure (cf., West, Howell & Cross, 1985, particularly pp. 39-40). It is true that Lerdahl and Jackendoff could likely produce what they would consider to be an unambiguous structure for any piece of music. It is also likely that any scholar of music could produce what that scholar considers to be an unambiguous structure for any piece of music given the rules of Lerdahl and Jackendoff. These interpretations would not necessarily agree, and in the absence of some objective statement, none could be said to be more correct than any other. Many have touted that the beauty of music lies in its ambiguity of interpretation (cf., Christ, DeLone & Winold, 1975, p. 56), but ultimately, in science one should be able to define that ambiguity, the reasons for that ambiguity and the relative preferences of various populations for each variant. One might try to use a panel of experts to choose the best representation, but it is arguable that a panel can never be more accurate than Lerdahl and Jackendoff in the interpretation of Lerdahl and Jackendoff's rules. If the interpretation of the panel of experts differs from that of Lerdahl and Jackendoff, then it can be argued that the panel had interpreted the rules in light of their own (non-objective) biases. We must accept the notion that the best "panel" to interpret the rules of Lerdahl and Jackendoff is the panel composed of Lerdahl and Jackendoff. Conversely, to the extent that others cannot use the rules of Lerdahl and Jackendoff objectively (creating the same interpretations), it must be admitted that the rules, however elegant their expression, are nothing more than the subjective interpretation of one panel of experts. As such, they would have no more validity than any other expert of equal stature. This same point has been raised by Brown and Dempster (see also 1989, pp. 94-96):

> But, how are analytical prescriptions justified? Should a naive listener simply place faith in the authority of some particular analyst? If so, which one? Sooner or later, prescriptions are only as believable as they are rationally justified. If this justification is rational, then appeals to authority are unnecessary.
>
> 1989, p. 92

The point has been to hammer home the futility of any examination of grouping structure at an intermediate or high level of the theory, at least until the various preference rules have been more clearly delineated (for an interesting start in this direction, see Deliege, 1987). Before the theory can be tested, it must be unambiguously defined so that all individuals can apply the rules to the same effect (even to the point of establishing probabilities for the alternative applications). In the presence of ambiguity, one is too often left with numerous post-hoc explanations of why something did or did not conform to predictions. That is, subjects' parsing may have coincided with the predictions of Group Preference Rule 5 (symmetry), but it could have been Group Preference Rule 4 (intensification of low-level principles), or Metrical Preference Rule 1 (parallelism) or top-down principles augmenting or overriding any of the above (remember that GPR 7 states that the preferred low-level grouping is one that results in more stable time-span and/or prolongation reduction). By similar considerations, the lack of conformity with the predictions of Group Preference Rule 5 may simply imply that other concerns were more important (e.g., time-span or prolongation reduction), even though the rule was initially applied.

The thrust of the argument is that, at this point, only the lowest levels of the theory have any real hope of an accurate test. However, in testing these rules, we must remember that when comparing the theoretical predictions of the rules to the actual parsings of subjects, we can expect to see many false alarms and hope to see no misses. The only solution to this problem is to somehow encourage subjects to demonstrate parsings at the lowest level possible.

The Rules of the Theory that Can Be Tested

The grouping structure at the lowest level is determined by the first six Group Preference Rules (the seventh relates the lowest levels to the higher levels). These are presented in Table 2.1 (the wording is taken directly from Lerdahl and Jackendoff (1983, p. 345-346); note the use of the words "prefer" and "may", throughout). Of these rules, the last three depend on the second and third. That is, it is difficult to imagine symmetry or parallelism in the absence of some boundaries based on Rule 2 or 3 (cf., 1983, pp. 48-49). As such these last three rules can be considered as the first step in the motion to the second level of analysis (an intermediate level of the hierarchy where low-level units are combined into larger units). That, in turn, implies that the first three rules define parsing at the lowest level. However, consideration of Rule 1 indicates that it is not a rule defining when to parse but rather, a rule defining when not to parse. As such one can view it in a manner like the last three rules. That is, if on the basis of Rules 2 and 3, boundaries were indicated that would result in a unit containing a single note event, then Rule 1 would be invoked to avoid that situation, but some higher level rules (such as Rules 4, 5 or 6, or possibly 7) would be needed to decided which boundary (if either) should be retained. Hence, at the lowest level, there are really only two rules to study. These two rules, in the system of Lerdahl and Jackendoff, should be sufficient to define all the group boundaries that might exist at all other levels of analysis. The point of the next section is to describe the method of quantifying these two rules so that they may be objectively related to a number of pieces of music and then related to the parsing of listeners. Examples taken from Lerdahl and Jackendoff that illustrate the meaning behind these rules are presented in Figure 2.1.

Table 2.1: The Group Preference Rules of Lerdahl and Jackendoff, 1983.

Rule 1		Avoid analyses with very small groups -- the smaller the less preferable.
Rule 2	Proximity	Consider a sequence of four notes $n_1n_2n_3n_4$. All else being equal, the transition n_2-n_3 may be heard as a group boundary if:
	a. (Slur/Rest)	the interval of time from the end of n_2 to the beginning of n_3 is greater than that from the end of n_1 to the beginning of n_2 and that from the end of n_3 to the beginning of n_4.
	b. (Attack-Point)	the interval of time between the attack points of n_2 and n_3 is greater than that between n_1 and n_2 and that between n_3 and n_4.
Rule 3	Change	Consider a sequence of four notes $n_1n_2n_3n_4$. All else being equal, the transition n_2-n_3 may be heard as a group boundary if:
	a. (Register)	the transition n_2 to n_3 involves a greater intervallic distance than both n_1 to n_2 and n_3 to n_4.
	b. (Dynamics)	the transition n_2 to n_3 involves a change in dynamics and n_1 to n_2 and n_3 to n_4 do not.
	c. (Articulation)	the transition n_2 to n_3 involves a change in articulation and n_1 to n_2 and n_3 to n_4 do not.
	d. (Length)	n_2 and n_3 are of different lengths, and both pairs n_1,n_2 and n_3,n_4 do not differ in length.
Rule 4	Intensification	Where the effects of Group Preference Rules 2 and 3 are relatively more pronounced, a larger-level group boundary may be placed
Rule 5	Symmetry	Prefer grouping analyses that most closely approach the ideal subdivision of groups into two parts of equal length.
Rule 6	Parallelism	Where two or more segments of the music can be construed as parallel, they preferably form parallel parts of groups.

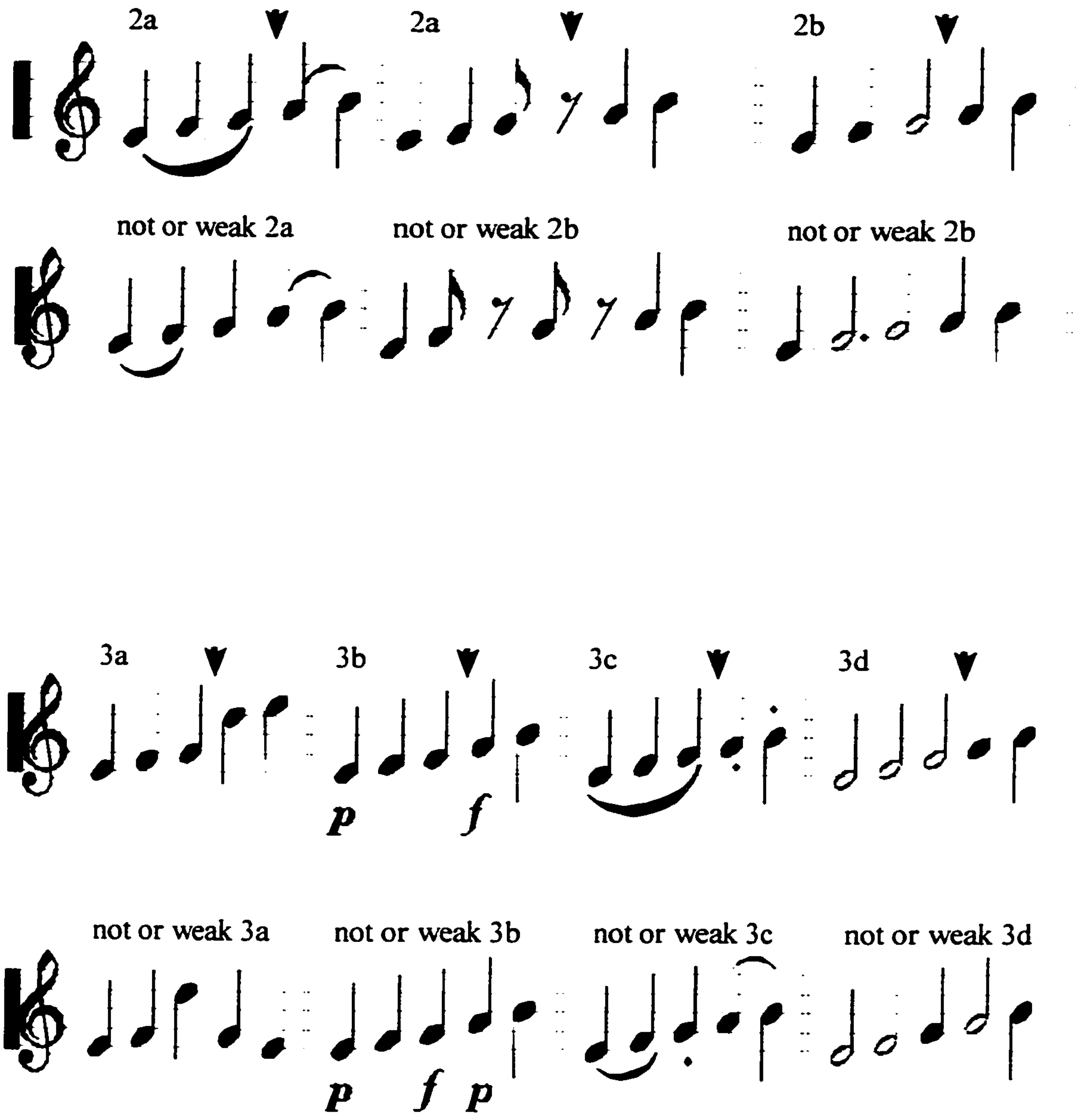
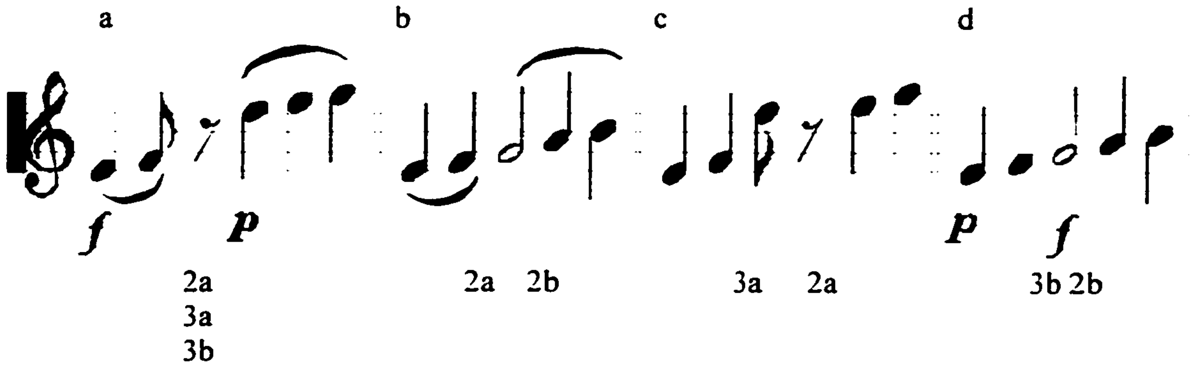



Figure 2.1: Example for Rules 2 and 3 of Lerdahl and Jackendoff's Group Preference Rules. These examples show what would be a basis for a boundary and what would not be a basis for a boundary (note that the second measure of the first staff is not a boundary based on 2b, though this is ambiguous). Note that boundary formation in the "not or weak" cases does not preclude boundary formation, particularly if other rules dictate the need for one. Figures adapted from Lerdahl and Jackendoff, 1983, pp. 44-46. ▼ indicates the boundary location.



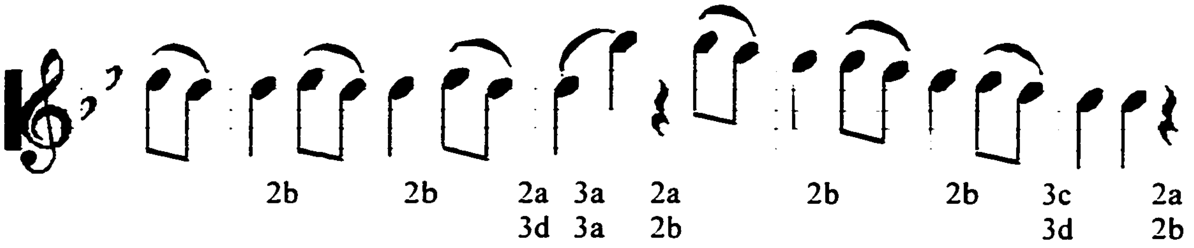


Figure 2.2: More complex applications of Lerdahl and Jackendoff's Group Preference Rules 2 and 3 to demonstrate the interaction of rules (top) and the application to actual music (bottom). The interactions of rules can lead to mutual reinforcement (a; leading to a stronger perception of a boundary) or conflict (b, c and d; leading to a weaker perception of a boundary). Note that the conflict exists only by virtue of the application of Group Preference Rule 1. The real application is the opening of Mozart's G Minor Symphony. Both examples are adapted from Lerdahl and Jackendoff, 1983, pp. 47-47. The rules are listed by number below each figure.

Note that, as is shown in the figure, the rules define the ideal case. As the divergence from the ideal increases, the "grouping intuitions are . . . much less secure" (Lerdahl & Jackendoff, 1983, p. 46). It should be noted that this conflict is really a manifestation of Rule 1: that is, groups should not consist of a single note event. A further example of the reinforcing or conflicting nature of these basic rules is shown in Figure 2.2. Lerdahl and Jackendoff's application of the rules to the opening of Mozart's G Minor Symphony, is also presented in Figure 2.2. Note the rules occasionally reinforce, often conflict (using the previous consideration) and more importantly, note that Rule 2 is applied much more often than Rule 3.

Before leaving this section, I would like to return to consideration of GPRs 4, 5 and 6. Rule 4 seems straightforward in its predictions (but see Chapter 4 for detailed analysis), and when comparing listeners' parsings with predictions its effects should be obvious. That is, in places where the Rules 2 and 3 are stronger, there should be a higher probability for a boundary at the next level. This rule also implies that when two or more aspects of GPRs 2 and 3 coincide, there should be a higher probability for a boundary at the next level (remember that listeners are probably parsing at some intermediate level). Rules 5 and 6 are more ambiguous. It is possible to imagine situations in which boundaries are created solely on the basis of these rules (i.e., places where Rules 2 and 3 do not apply, but Rules 5 and 6 would). At this point, Symmetry and Parallelism, despite their prevalence in music, are not well enough defined for conversion into an algorithm (however, see Chapter 4 of this work). For example, in the absence of Rules 2 and 3, should Symmetry be invoked for spans as small as four notes, six notes or perhaps only for spans as long as twelve notes? Because of the lack of definition, Symmetry, at this point, should be constrained to the creation of the second level of the hierarchy, and as such, to the arbitration of digressions between Rules 2 and 3 (i.e., if Rule 2 applies at one note, and Rule 3 applies at the adjacent note -- violating GPR 1 -- but Rule 5 supports Rule 2, then a boundary would be placed at Rule 2). The same, to a more serious degree, can be said of Rule 6, Parallelism. What is to be considered the basis of parallelism? It could be pattern of note durations, articulations, attack points, pitch heights or even tonal function. Furthermore, what is to be considered the appropriate span of notes over which parallelism works? Four notes seems too short (i.e., two groups of 2), but is 20 notes too long (two groups of 10, which might be beyond short term memory span)? As with Symmetry, it seems, at this point, prudent to constrain parallelism to the creation of the second level of the hierarchy, and as such, to the arbitration of digressions between Rules 2 and 3. As before, it is important to point out that Lerdahl and Jackendoff (1983, p. 52-53) are aware of the ambiguity in their formulation at this point of development. They believe that their lack of specificity is reflective of a more basic problem in human psychology (perception/cognition): That is, how do people recognize similarity? As they have implied, there is a second problem with respect to music: That is, how does one model the ambiguity that listeners hear, and accept, as a part of music? Before getting too embroiled in a discussion of parallelism or symmetry within music, it seems prudent to determine whether or not the lower level rules are a reasonably accurate reflection of human music perception and cognition.

Other Tests of Group Preference Rules 2 and 3

Given the influence of Lerdahl and Jackendoff (1983), it is not surprising that there has been other research on the application of their rules. In this review, the focus will be on tests of the low-level grouping principles though many other aspects of the theory have been tested (cf., prolongation reduction: Krumhansl, 1996).

In the first of two experiments, Deliege (1987) tested the parsing of 32 short sequences (3 to 16 notes) extracted from the western tonal repertoire (Bach to Stravinsky) presented auditorially. There were six groups of subjects, musicians (first, second and third level harmony students at Belgian Royal Conservatories) and non-musicians (no formal training), who heard the extracts in isolation, or after providing the musical context of the extract or after providing a musical context that was similar to, but not the true context of the extract (i.e., extracts from the same work, or from the corpus to which the work belonged). Subjects indicated their parsings by placing a line between a series of dots presented visually: the number of dots matched the number of sounds in the upper voice of the extract. Stimulus selection was designed to permit the examination of all the

Group Preference Rules. Two additional rules were defined for study: boundaries due to timbre changes (Rule 3e: Timbre Change), and large register changes (Rule 3f: Large Register Change) were considered separate from small register changes (Rule 3a: Register Change). The basic findings were that both musicians and non-musicians parsed the sequences in accordance with the rules but musicians did so significantly more often (musicians averaging 77.2% and non-musicians averaging 50.8%). Within the musical training group, level of musical training did not matter (though the range was quite limited). Only musicians were affected by context (producing slightly different patterns of responding). Verbal reports of the subjects explaining their reasons for parsing corresponded to the rules for both musicians (87.5%) and non-musicians (69.9%). As for the rules themselves, both musicians and non-musicians used Rule 2b (Attack-Point: 99% and 99%), Rule 3b (Dynamics: 81% and 75%), and Rule 3e (Timbre: 80% and 70%) consistently. Musicians tended to make much more use of the remaining rules than did non-musicians: Rule 2a (Slur/Rest: 80% and 30%), Rule 3a (Register: 75% and 48%), Rule 3c (Articulation: 56% and 13%), Rule 3d (Length: 70% and 30%) and Rule 3f (80% and 46%). Deliege noted that subjects rarely parsed in accordance with Rule 3d (Length): Rather than parse between notes n_2 and n_3 (of the sequence of four notes n_1, n_2, n_3 and n_4, where n_1 and n_2 are of equal lengths and n_3 and n_4 are of equal lengths but n_1 and n_3 are not of equal lengths) as predicted by the rule, subjects parsed between n_3 and n_4 (delayed segmentation). However, it should be notated that such delayed segmentation is actually in accordance with a one-sided version of Rule 2b (Attack-Point). This interpretation is supported by the observation that the other similarity rules (Rules 3a, 3b, 3c, 3e and 3f) do not suffer this delayed segmentation. Deliege used a slightly different classification scheme that grouped Rules 2a, 2b, 3c and 3d together (length or temporal changes), but the other rules in this set did not exhibit such delayed segmentation. While it seems likely that Rule 3d is not particularly important for parsing, it is also likely that the parsing Deliege observed was due to some other rule and not indicative of the need to recast Rule 3d to parse after the first different note. Deliege also noted that many subjects cited a change in the direction of the melodic contour as a basis for parsing.

Some of these findings were incorporated into the present algorithm for the tests of parsing (see the next section for details). For example, the strength of both Rules 2b and 3b was a factor in the decision to control dynamics at the level stimulus (i.e., since both were strong, to avoid confusion, changes in dynamics were eliminated). Secondly, the lack of a distinction between Rules 3a and the new Rule 3f (Large Register Changes) was the basis for the type of quantification behind Rule 3a (see Figure 2.5): The quantification does not grant ever increasing values to larger register changes. The lack of strong effect for Rule 2a (Slur/Rest) was a factor in the decision to eliminate slurs from the current stimuli. Deliege also noted that, of the rules concerning length, some pertain to aspects indicated in the score (Rule 2b: Attack-Point) while others pertain to aspects of performance (Rule 2a: the slur of Slur/Rest, Rule 3c: Articulation). She found that musicians were sensitive to both types while non-musicians were only sensitive to the first type. This was also a part of the decision to eliminate Rules 2a (Slur/Rest) and 3c (Articulation) from current studies.

From Deliege's results, there is an additional consideration for any experimental tests of the rules. Essentially, training did seem to have some effects, but the effects were not dramatic, particularly when one considers the range of training. That is, some rules are used by all regardless of training, but the use of other rules is enhanced by training

In the second experiment of the same study, Deliege (1987) pitted the various rules against each other to determine which would be more important using specially designed nine-note sequences. In each case, one rule predicted a boundary before the third note (or the fourth, fifth or sixth notes) and a second rule predicted a boundary after the third note (or the fourth, fifth or sixth notes). This is, in effect, a test of Rule 1 (avoid groups with a single event) but stimuli were controlled so that there could be no consideration for higher order rules (i.e., Symmetry or Parallelism). This seems odd because Lerdahl and Jackendoff clearly imply that context will be the determinate of which rule is more important: Hence no rule should be intrinsically stronger than any other. Be that as it may, the test considered all the previous rules, but dropped the new Rule 3f (Large Register Changes) and introduced a new rule, which I call Rule 4a

(Changes in Melodic Contour: a form of good continuation). It is temporarily labeled as such to highlight that its basis is distinct from the proximity and similarity principles that underlie Rules 2 and 3 (though arguably, Rule 3a: Register is a form of good continuation). It is not surprising, given the more constrained stimuli, that responses were in accordance with the rules much more often for both musicians (97.0%) and non-musicians (92.9%). In this case, agreement implied the use of either rule that could be applied. In the analysis, for the musicians, the rankings were Rule 3e (Timbre: 89%), 3b (Dynamics: 66%), 2a (Slur/Rest: 56%), 3a (Register: 48%), 3c (Articulation: 45%), 2b (Attack-Point: 41%), 3d (Length: 33%) and 4a (Melody: 10%). For non-musicians, the ranking was somewhat different: Rule 3e (Timbre: 77%), 3a (Register: 75%), 2b (Attack-Point: 59%), 3b (Dynamics: 46%), 2a (Slur/Rest: 42%), 3d (Length: 27%), 3c (Articulation: 25%) and 4a (Melody: 20%). Deliege also noted that subjects tended to parse in accordance with the key-defined tonal function of notes: That is, subjects often parsed after the dominant. Secondly, Deliege noticed a marked preference for parsing towards the centre of the nine-note sequence, particularly for musicians. This is in accordance with Lerdahl and Jackendoff's (1983) Group Preference Rule 5 (Symmetry).

Two points can be gleaned from this experiment. Firstly, a change in the direction of melodic contour is not likely an important parsing mechanism (remember that such change is distinct from Rule 3a -- register). Secondly, musicians and non-musicians might parse differently if rules conflict (which might be the case in a "real" melody). Given the observation of parsing on the dominant, one might consider degree of training as an indicator of degree of instantiation of the notions of tonality. That is, training enhances the ability to distinguish such tonal centers.

Although generally well designed and executed, there are some minor concerns with respect to this study. Firstly, the method of applying the rules (choosing or designing the sequences that exemplified the rules) was never clearly stated. Hence, the degree to which the rules actually apply cannot be known. That is, the rules were not quantified so it is unclear that the different rules were given equal footing for the tests. Secondly, in the first experiment, the extracts seem to have been fairly complex (only some examples were presented). Hence, there may have been applications of multiple rules within a single extract (reinforcement and competition) that were not noticed by the author. The previous discussion of delayed segmentation highlights the possibility of such a situation. In the second experiment, the sequences were contrived and totally lacking in context, which is a large divorce from music and the goals of Lerdahl and Jackendoff (1983). Both these experiment highlight a need for quantification and algorithmic application of the rules (one could not miss an application of a rule). Thirdly, the parsing was not done on-line. That is, subjects heard the sequences (in the first experiment, subjects could hear the sequence more than once and the number of repetitions was monitored) and then parsed. Hence, parsing could have been based, retrospectively, on global aspects of the stimulus, rather than on features inherent within the stream of sound signals. Fourthly, in the first experiment, given that the segments were extracted from longer pieces, it is not clear that metrical accents were controlled. Although it is not clear, numerous comments within the published paper, seem to imply that performance variations were retained so it is likely that subjects might have been parsing on the basis of metrical structure. All of these concerns were considered in the design of the experiments in this work.

A similar study was conducted by Peretz (Experiment 1, 1989) using extracts from twelve French folk tunes, although Peretz only looked at Group Preference Rules 3a (Register), 3d (Length) and 6 (Parallelism). In the first experiment, subjects heard an extract (less than 14 notes) and then indicated the boundary on a line of dots in the manner of Deliege (1987). Again musicians and non-musicians reliably parsed music into components that could be meaningfully related to intuitions of phrase boundaries and the Group Preference Rules of Lerdahl and Jackendoff. Musicians' parsings generally coincided more with the rules than did non-musicians: Musicians performed at an average of 90.1% concordance with the rules (Rule 3a: 65%, Rule 3d: 100% and Rule 6: 98%) while non-musicians performed at an average concordance 75.0% with the rules (Rule 3a: 54%, Rule 3d: 83% and Rule 6: 75%). The results clearly mimic those of Deliege (1987).

The experiment shares the same concerns as those of Deliege (1987). In the absence of a quantified estimate of the rules, one cannot be sure that no application has been missed. In fact, in one example of parallelism, the note preceding the boundary for parallelism would imply a boundary based on Rule 2b (Attack-Point). As in Deliege, in Experiment 1, boundary positions were collected after hearing the entire piece (subjects could request a second presentation of each stimulus), and as such, might have been based on retrospective considerations of global structure. In Experiments 2 and 3, Peretz (1989) changed the procedure to avoid this. In Experiment 2, subjects heard the extract and then a three-tone probe. Subjects were required to indicate whether or not the probe had been a part of the extract. Some probes crossed grouping boundaries (determined theoretically) while other probes did not cross group boundaries. If the probe did not cross a boundary, then the comparison task involves retrieving a single unit and then comparing that unit to the probe. If the probe did cross a boundary, then the comparison task involves retrieving two units, splicing those units and then comparing them to the probe. It was assumed that the latter operation is more difficult. It was found that the latter operation generally lowered performance (assessed by errors and reaction time), thereby confirming the hypothesis, though the proof is somewhat indirect (i.e., one has to assume that the listener actually segmented at the expected location). There was no substantial difference between musicians and non-musicians. There were some performance differences between the type of boundaries, and most importantly, boundaries based on Rules 3a and 3d were not the cause of the probe effect, though a speed-accuracy trade-off clouded the results.[19] In the third experiment subjects received the probe before hearing the extract. As such, the dependent measure was the ability to spot the probe in the melody. The results were essentially the same, though Peretz concluded that the monitoring task was more reliable than the recognition task.

Palmer and Krumhansl (1987a) tested the theoretical predictions of Lerdahl and Jackendoff (1983) using the first four measures of Bach's Fugue XX in A minor (from the *Well Tempered Clavier*, Book 1), although they did not separate tests of metrical structure from those of grouping structure. Palmer and Krumhansl (1987a) asked subjects

---

[19] Peretz and Babai (1989, 1992) have found a strong Rule 3a boundary effect on probe recognition in an experiment that explored the chunking of melodies.

to rate the goodness or completeness of a short musical segment (from Bach's Fugue XX in A minor). The segments either ended on a phrase boundary or at some point within a phrase (phrase analysis by Lerdahl, based on Lerdahl & Jackendoff, 1983). Lerdahl's analysis was converted into numerical ratings that expressed the degree to which each note should provide a sense of phrasing. That is, each note was tagged for the role it played in the time-span hierarchy (or the level it belonged to in the hierarchy). Higher ratings of completion were produced by segments that ended on a phrase boundary (highest levels in the hierarchy), but generally, the sense of completion correlated with the ratings based on the (high-level) time-span reduction. Because of the focus of this work on the grouping structure, it is interesting that a quantification of the basic metrical pattern did not correlate with the sense of completion. This induces one to conclude that it was the grouping structure (the focus of this work) that determined the overall rating (these conclusions were supported in a regression part-analysis). It is interesting to note that Palmer and Krumhansl thought it important to consider the tonality separately: They found that tonality (using Krumhansl & Schmuckler's quantification of the tonal hierarchy; Krumhansl, 1990) was significantly correlated with the perception of completion for the pitch pattern and almost significant for the overall melody pattern. Tonality was not correlated with metrical stress. In fact, metrical stress was not associated with completion ratings for either pitch pattern or time pattern (or the overall melody pattern). Palmer and Krumhansl (1987b) produced very similar results using a harmonically more complex piece (Mozart's A Major Piano Sonata, K 331). In this case, musical training was explicitly analyzed, but no major differences were noted (the theory is supposed to apply to the expert listener, but all had some musical training: 2.5 to 14 years: no training effects were mentioned).

Generally, the data from Deliege (1987), Peretz (1989a) and Palmer and Krumhansl (1987a), support the theory of Lerdahl and Jackendoff (1983). However, there is some ambiguity about the detailed application of the rules. The present work quantifies those rules in a hope that such quantification might begin to disentangle the ambiguity. Secondly, there is some question about the roles of both tonality (or diatonicism) and training. The current study and further studies will have to carefully consider both.

Quantification of Group Preference Rules 2 and 3

The goal of this section is a quantification of the two low-level rules of Lerdahl and Jackendoff (1983). The intent is to create a means of embedding the rules in an algorithm so that the rules may be applied to a melody without bias or influence (i.e., the automatic top-down processing that a human might use when deciding how to apply these rules).

Unfortunately, the rules as defined by Lerdahl and Jackendoff (1983) are not in a form that is directly amenable to conversion to an algorithm. However, Lerdahl and Jackendoff do acknowledge difficulties that arise when the rules conflict (1983, p. 47, 54) and they do acknowledge the potential need for more quantitative approach to rule strength (1983, p. 47, 54). They feel that the current level of understanding is not sufficient for such a delineation, so that the premature application of numerical recipes would not assist in the development of their theory.

Hence, fully cognizant of the Lerdahl and Jackendoff's (1983) trepidations, this section will describe a quantification of the low-level rules (Rule 2 and 3). This quantification can then be related to the actual performance of subjects and in turn, provide additional feedback on the relative utility of each of the rules, thereby bootstrapping the theory. In this development, the intent is to depart as little as possible from the rules as they are stated. Although Lerdahl and Jackendoff (1983) do not believe that quantification can be achieved at this point, they do provide a number of insights that might serve as validity checks in any subsequent analysis. They imply that judgements of boundaries should depend on the degree to which the rule is met (i.e., longer time disparities between attack points should produce stronger boundaries in Rule 2b). At another level, they imply that the different rules may be used to different degrees (perhaps idiom specific variations) so that boundaries would be the product of the intrinsic strength of the rule and the conditions for the rule manifest in the stream of auditory events. For example, they imply that Rule 3 is generally weaker than Rule 2 (1983, p. 47). They further imply that, although the application of all the rules is implicitly two-sided within their definitions (i.e., boundaries are defined with reference to notes on both sides of the boundary; see Table 2.1), one-sided applications might result in weaker perceptions of boundaries (1983, p. 46). That is, as stated, all the rules consider the possibility of a boundary between the two middle notes of a four-note span. However, Lerdahl and Jackendoff imply that some version of their rules might yield boundary predictions between the second and third notes of a three-note span (these would be weaker than the corresponding two-sided version).

Of these considerations, only the relation between the degree of rule adherence (i.e., the degree to which the note events in the sequence of music instantiate the rule) and boundary strength is developed herein. The reason is that this relation is an unambiguous extension of the model. On its own, it is sufficient to allow for the quantification of the rules. Considerations of the relative intrinsic strengths of the different rules are not considered: Simply stated, it is better to let listeners demonstrate which are more important (with due consideration for the opinions of experts as a check on validity). Considerations of one-sided versus two-side rules are not developed because Lerdahl and Jackendoff (1983) did not consider those in sufficient detail to be modeled: any extensions would be unnecessary and possibly discordant within their system.

To that end, the extensions of the rules of Lerdahl and Jackendoff (1983) were limited to an assessment of strength of the boundary implied by the application of the rule. Although this may seem sufficient for the design of experimental tests, there are still many unanswered questions before such an implementation can be made. Firstly, the major question is "what is the primitive of each rule?". In this context, primitive is a statement of the underlying construct. For example, Rules 2a, 2b and 3d are clearly about time. Rule 3a is about pitch, but is it log-frequency pitch (i.e., physical pitch) or is it diatonic pitch? Rule 3b is about intensity, and Rule 3c is about different styles of expression. The second, less critical, question is "How should the rule be scaled?" A failure to capture the correct primitive would mean that the theory was not being tested. A failure to find the right scaling would imply that either important relationships may be lost or that unimportant relationships may be emphasized.

Table 2.2 The Primitives assumed for Group Preference Rules 2 and 3 of Lerdahl and Jackendoff.

Rule 2	Proximity	
	a. (Slur/Rest)	time, scaled linearly, measured in SI units (milliseconds), but discussed, more simply, in terms of note durations
	b. (Attack-Point)	time, scaled linearly, measured in SI units (milliseconds), but discussed, more simply, in terms of note durations
Rule 3	Change	
	a. (Register)	difference in pitch height, scaled linearly, using note number (MIDI format, with middle c as note number 60) as a basis
	b. (Dynamics)	absolute loudness on a logarithmic scale.
	c. (Articulation)	categorical scaling of a change in type
	d. (Length)	time, scaled linearly, measured in SI units (milliseconds)

Notes: SI is Systeme International, the standards for measures in science.
MIDI is Musical Instrument Digital Interface, an international standard for the electronic encoding of musical signals

Basically, to summarize briefly (see Table 2.2), one primitive was assumed to be time, measured on a linear scale. Numerous studies in psychoaccoustics have demonstrated a linear scaling of time intervals for durations in the range of musical notes (Allen, 1979; Schiffman, 1982). Non-linearities only manifest at very short or very long times. Hence, a quarter note was assumed to be twice as long as an eighth note, and one half the length of the half note. Musical notation (in place of actual millisecond timing) will be used wherever possible in the presentation and discussion of results because it is more straightforward. Actual millisecond timings will only be cited when necessary. The important point is that durations were scaled linearly[20]. The second primitive was assumed to be pitch height. This is a much more ambivalent decision because several studies have demonstrated that the perception of notes in music is not simply a function of pitch height (cf. Krumhansl & Shepard, 1979; Shepard, 1982; Chapter 1, this work). The distances or the perception of distances between notes are not simply a function of the differences in frequencies (Krumhansl, 1979). Be that as it may, the rules cited by Lerdahl and Jackendoff strongly imply absolute pitch heights (logarithmically scaled frequencies: equal tempered tuning) by virtue of their lack of reference (except in higher rules) to notions of tonality. It is true that later works by these authors (Jackendoff, 1992; Lerdahl, 1992) do implicate tonality, but these later works do not provide any way to instantiate tonality. Since the construct is so complex, it was felt that the simplest primitive would be the best. To simplify, the notion of pitch height was recast in terms of note number using the standard MIDI format for the scaling notes (i.e., semitone differences in pitch are translated to whole number differences, with middle C as note number 60, and the C an octave above middle C is note number 72). Effectively, each semitone step was treated as equivalent to listeners, which is in fact, what logarithmic frequency perception does. Scales for the other changes (articulation and dynamics) were also defined (see Table 2.2), but for the remainder of this work, these were not utilized since changes on these dimensions were controlled at the level of the stimulus. This allowed the present work to focus on a smaller set of rules.

Both time and pitch height used a linear scaling, even though it seems obvious that many aspects of boundary formation should be based on non-linear scaling. For example, boundaries based on the presence of a rest should be more obvious the longer the rest, but there is a limit. Once one approaches the level of the semibreve rest (whole rest), longer rests might not result in a greater perception of a boundary. In the other

---

[20] Note that the results of Lantz and Cuddy (1996, 1998) might imply a non-linear scaling of the perception of the relative importance of duration of musical tones, but their stimuli covered a much broader range of durations than is typically found in melodies like those used in the present work. In addition, their work was not available when the algorithms were developed.

direction, a rest shorter than a quaver (eighth rest) or possibly a semiquaver (sixteenth rest) are not likely to result in a boundary. The actual function is more likely to resemble that of a logarithmic, sigmoidal, or hyperbolic tangent, but over the interesting range (i.e., values used in simple melodies), the approximation of linearity seems reasonable. In general, the use of linear scaling is considered a first order approximation. That is, all scales are linear over some range, and in fact, the logarithmic, sigmoidal or hyperbolic tangent functions are correlated with the linear scale in the restricted ranges for note durations and pitch jumps found in simple melodies (relative to complex orchestral pieces; non-linear scaling might be essential if one wishes to deal with orchestral music). Figures 2.3 to 2.5 demonstrate, in musical notation, the meaning of these scalings for each of the rules. Several examples are provided, along with the algorithm and a magnitude estimate of the value of the rule within the examples.

Before discussing the rules in greater detail, it is important to state a few considerations that apply to the algorithm as a whole. Firstly, in all cases the implementation of the rules were kept as simple as possible, without incurring the loss of too much information. Obviously, all rules, given their definitions, can be represented by binary conditions (i.e., on or off). For some rules the move to a more quantitative scaling seems straightforward and even necessary (attack-point, pitch height), while for other rules, such a quantification might seem unwarranted given current knowledge (i.e., articulation). Much of the decision of how to implement is based on the complexities of the algorithm necessary to implement the particular rule. If a fully quantitative algorithm would be very complex (in my view as the programmer), then it is likely that a simpler binary version should be retained. If the quantification is simple, then it should be implemented. This tacitly assumes that the complexity of the algorithm reflects the complexity of the human processing system. Secondly, in all cases a tie was treated as a single note with a length equal to the total of the notes tied. Thirdly, inter-note durations (the interstimulus interval) were not included. Although variations in such intervals are a part of the expression of music, the first implementation of this algorithm would require that they be a constant. The main reason for this decision is that it is much too difficult to attempt to address the expressive variations of such intervals within an algorithm that is driven only by the score of the music[21]. Given that the interstimulus intervals are to be treated as constant, presupposing all researchers could agree on an appropriate interstimulus duration, inclusion of interstimulus times would only serve to unnecessarily complicate issues at this point. That is, although it would seem that some of the rules to be cited below would be affected by the lack of consideration of interstimulus times (e.g., Slur/Rest), the effect would be small when considering all rules. This tacitly assumes that interstimulus intervals are smaller than the smallest perceivable rest (likely a 16th, or possibly a 32nd, rest: this may be idiom or style dependent) and as such are of very short duration relative to the "usual" notes of a melody.

Finally, for later simplicity, the application of the rules is normalized to a range from 0.0 to 1.0. This allows the rules to be plotted using a common scale, and more, importantly, this allows for an intuitive comparison across rules. Remember, however, that the scaling is somewhat arbitrary. That is, while the scale maximum for each rule should represent a strong boundary and the scale minimum for each rule should represent no boundary at all, direct comparisons at the intermediate points cannot be made: for example, a magnitude of 0.5 may represent a 50% probability of a boundary for one rule, but only a 20% probability for another rule.

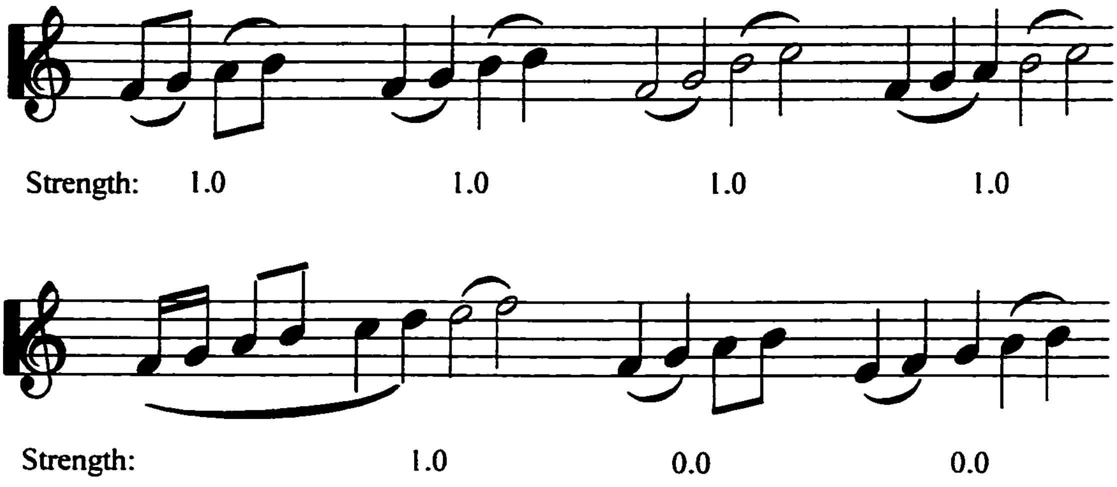
GPR 2a (Slur/Rest) actually represents two operations, dissimilar enough that one wonders why they are included in the same rule (however, both examine relative pauses between notes: the perception of offset to onset changes). The difference is even more obvious when one tries to convert each to an algorithm. The slur represents the blurring of the distinctions between notes, while the rest is the absence of sound (see Figure 2.1). The slur is effectively the categorical analogue of a frequency glide and as such the rule seems to address the issue that there is no interstimulus distance between notes under a

---

[21] At this point in time, algorithms cannot make subjective expressive decisions; algorithms can be programmed to model the subjective decisions that their programmer would have made, but this is not the same as making subjective decisions.

slur, followed by some interstimulus distance from the end of the slur (the end of the last note of the slur) to the next note (the onset of the next note). In the examples cited by Lerdahl and Jackendoff (1983), the slur rule seems to require that there be two sets of slurred notes, with no intervening isolated notes (i.e., one slur ends and the next begins: see Figure 2.1 for an example). As such, the slur rule seems to be a binary representation. The rule either applies or does not. This is how it has been implemented (see Figure 2.3) with a 0.0 indicating no application of the rule and a 1.0 indicating an appropriate application of the rule. For the future, should the slur rule show some promise, it might be useful to treat the durations of each slur (before and after the boundary) as a measure of the boundary strength. It can be assumed that two eight-note slurs would create a stronger sense of boundary than two two-note slurs, but the implementation of this would be problematic because it involves spans of notes that are not adequately described by Lerdahl and Jackendoff. That is, though it is obvious that such long spans under a slur would certainly constitute a group, this reflects more the application of the higher level rules (symmetry, parallelism or time-span reduction), than it does the application of the slur rule per se.

The rest aspect of GPR 2a clearly deals with the rest in comparison with the usual interstimulus interval between adjacent notes. The idea expressed is that the absence of sound for a duration that is in the normal range of notes is cause for a boundary. The absence of sound is likely important: In studies of sentence comprehension, subjects have been found to be generally oblivious to the replacement of a syllable (or three) by a nonsense noise (cough, buzz or tone), but highly sensitive to the removal (replacement with silence) of a single syllable (Warren, 1970, Warren & Warren, 1970; Warren & Obusek, 1971; Obusek & Warren, 1973 as cited in Anderson, 1990, and/or Goldstein, 1984, and/or Reynolds & Flagg, 1983). That boundary is placed on the rest. It is so obvious that the duration of the rest should reflect the strength of the boundary, that it must be quantified. However, there are two ways to approach this. One could consider the absolute magnitude of the rest as a measure of the boundary strength.

Rule 2a: Slur

Rule 2a: Rest

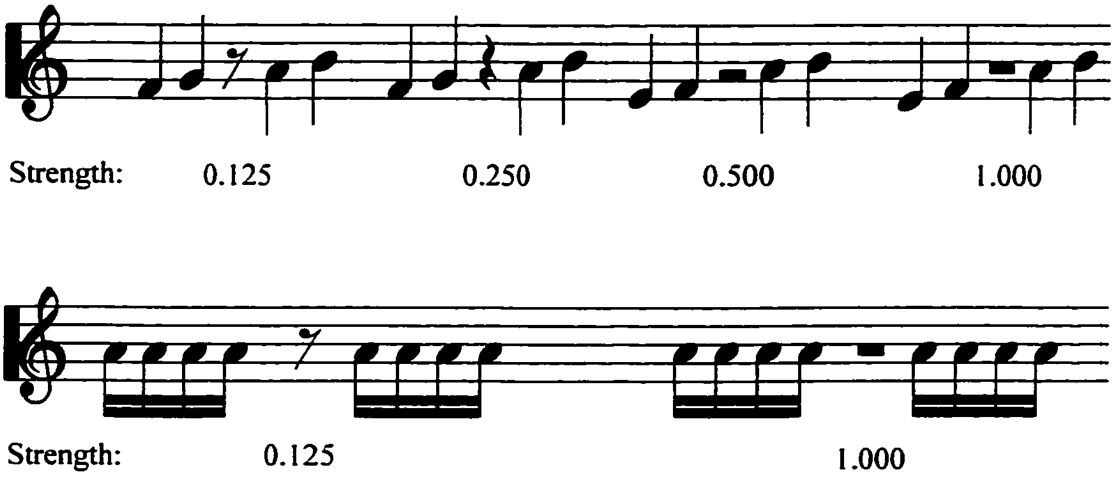


Figure 2.3: Quantification of the Group Preference Rule 2a of Lerdahl and Jackendoff. (1983). The strength notated under each example is the relative measure of the instantiation of a boundary at that point. Note that the slur version is represented as a binary condition, while the rest version is the relative magnitude of the rest regardless of context.

Hence, a whole rest would receive a value of 1.0, a half rest would receive a value of 0.5, a quarter rest would receive a value of 0.25 and so on. This has simplicity and cross-music comparisons to recommend it, but it does fail to address relativistic concerns. For example, a semibreve rest (whole) in the midst of a stream of quarter notes would seem much more important than a semibreve rest (whole) in a stream of whole notes. Absolute magnitudes do not capture these kinds of relationships. Which is closer to the goals of Lerdahl and Jackendoff (1983)? The answer is not obvious. The attack-point variation of GPR 2 (i.e., Rule 2b) would imply that the relativistic coding is more important than absolute coding, but the wording of Rule 2a seems to compare the rest size (plus an associated interstimulus interval, if any be added in front of a rest), in an absolute sense, to the usual interstimulus interval between notes. If we assume a constant, or an average interstimulus interval, then the absolute size of the rest seems closer to ideals of Lerdahl and Jackendoff. Although my first implementation of this rule acted in a manner like Rule 2b (see below), I now use an absolute implementation (see Figure 2.3) in which the semibreve rest (whole rest) implies a boundary with strength 1.00, a mimim (half rest) implies a boundary with strength 0.50, a crochet (quarter rest) implies a boundary with a strength 0.25, et cetera. This has the beneficial effect of creating a clear distinction between Rules 2a (Rest) and 2b (Attack-Point). However, this creates a slight problem for scaling because breve rests (double whole) will exceed and multibar rests might exceed the length of the semibreve rest (defined as boundary implication 1.0). One can either allow that the scale will, on rare occasions, exceed a value of 1.0, or try to encode the longest ever possible rest as a value of 1.0, and scale accordingly. Since rests longer than a semibreve are rare, particularly in a single melody line, the former method has been chosen. Note that this is also one occasion where a non-linear scaling (logarithmic or sigmoidal) might be more appropriate in the long run: A knee (point of inflection) at or near the semibreve rest would effectively mean that any rest longer than the semibreve would be equivalent in its ability to induce a boundary.

GPR 2b deals with the differences in attack points of adjacent notes that are due to the variations in the lengths of notes (this is distinct from Rule 3d: length changes, that will be discussed later) or due to the presence of rests. Rule 2a can be conceived as an offset to onset rule, while Rule 2b can be conceived as an onset to onset rule. This rule does not include differences that are due to rests (see Rule 2a above) because it is assumed that there is some sound throughout the duration (see Figure 2.1 and Table 2.1). The rule only applies if there is a note with a relatively longer duration inserted between two notes of shorter durations; the boundary is placed after the longer duration. The rule, as defined, implies that longer discrepancies should be associated with stronger boundaries, but that leads to the question of how to quantify those durational difference. The simplest implementation, for a span of four notes (n_1, n_2, n_3, n_4), compares the interval between the attack points of notes n_2 and n_3 with the average of the associated intervals between n_1 and n_2 and n_3 and n_4. Since the interstimulus interval is not considered herein, this reduces to the length of n_2 compared to the average of n_1 and n_3. That is, the interval between the attack points is the length of the notes. To normalize to the range of 0.0 to 1.0, the ratio was taken:

boundary = $1.0 - (n_1+n_3) / (2 * n_2)$ where the n's are lengths
$n_2 > n_1$ and $n_2 > n_3$

Because n_2 must be longer than n_1 and n_3 the rule produces a boundary value of 0.0 in the limit as n_1 and n_3 approach the length of n_2. When n_1 and n_3 are much shorter than n_2, the rule produces a value that approaches 1.0. If either n_1 or n_3 are longer than n2, the rule does not apply. Note that this quantification removes consideration of the units (e.g., milliseconds) used to measure note duration. Also note that the rule really only involves a three note span (n_4 plays no role).

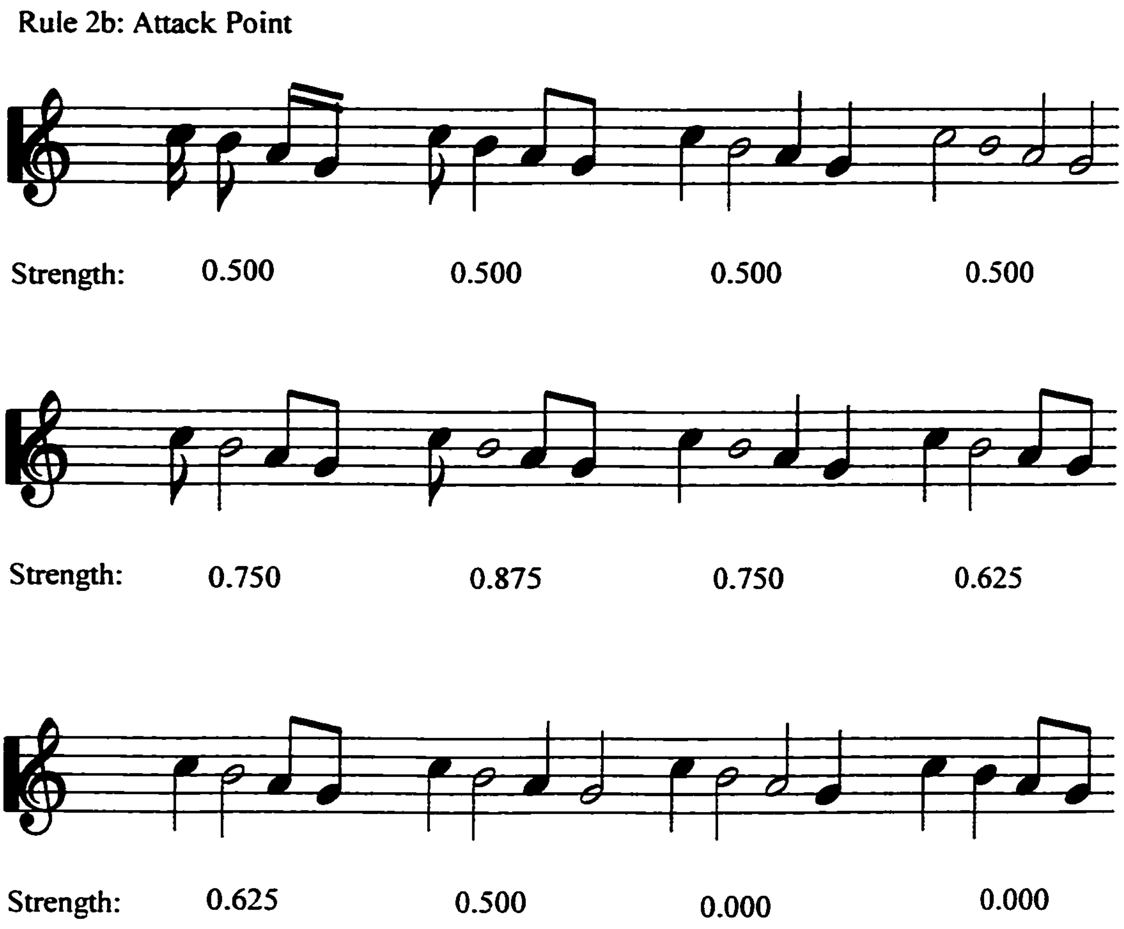


Figure 2.4: Quantification of the Group Preference Rule 2b of Lerdahl and Jackendoff (1983). Note that the rule produces a range of values between 0.0 and 1.0, and that it is symmetric in its application. The rule assumes that the second note is longer than both its flankers, and returns a 0.0 if it is not.

For those concerned with the previous dismissal of the interstimulus interval, it should be noted that as the interstimulus interval approaches the magnitude of the note durations, the scaling becomes a non-linear representation of that which is intended. However, this is only a problem for very short notes. For example, if we assume a whole note surrounded by 32nd notes, with a duration of 320 ms for the whole note, 10 ms for the 32nd notes and a 10 ms interstimulus interval, the boundary value is 0.94 when the interstimulus interval is considered (the true expression of the rule) and 0.97 when the interstimulus interval is ignored. For the case of a quarter note surrounded by 32nd notes, the values are 0.78 and 0.88 respectively. In the more common case of a whole note surrounded by quarter notes, the values are 0.73 and 0.75. The distinctions are not large because their interstimulus interval would be counted twice in the numerator and twice in the denominator. Common examples of this are presented in Figure 2.4. To accommodate changes in attack-point that are due to rests, a simple modification was added. If there is a rest within the span of the four notes, the value of the rest is added to the note preceding the rest before the computations are started. Hence, Rule 2b is confounded with the rest aspect of Rule 2a (a separation is not possible).

GPR 3 has four delineated conditions, but for reasons mentioned earlier, only two (Rules 3a: Register change and Rule 3d: Length change) will be discussed here. Rule 3a (Register change) occurs when the first two notes of a four note sequence are relatively close in pitch and the last two notes of that same four note sequence are relatively close in pitch, but the first pair is not close, in pitch, to the second pair. The boundary is placed between the first and second pair. The quantification of Rule 3a, as with Rule 2b above, is a somewhat ambiguous situation. (In this discussion, one must remember that for reasons already cited, register change does not include considerations of diatonicism or tonality.)

Firstly, register change is a potentially unbounded change: That is, there is no theoretical limit to the magnitude of pitch changes, though consideration of what is normal will certainly lower the ceiling. Secondly, the change is, again, expressed as more change between two adjacent notes than is shown in the flanking pairs of notes. Hence, register is a relative change. To implement this, the amount of jump between notes n_2 and

n_3 (see Table 2.1 or Figure 2.1) was compared to the average of the associated jumps between n_1 and n_2 and n_3 and n_4. To normalize to the range of 0.0 to 1.0, the ratio was taken:

boundary = 1.0 - (n_1-n_2 + n_3-n_4) / (2 * n_2-n_3) where the n's are pitch heights in MIDI notation

Because the transition between n_2 and n_3 must be greater than that between n_1 and n_2 as well as n_3 and n_4, the rule produces a boundary value of 0.0 in the limit of no difference in the size of the transitions and a value of 1.0 when the pitch heights of n_1 equals n_2 and n_2 equals n_4 (see Figure 2.5 for further examples).

As implemented, this rule has the possibly undesirable effect of producing a value of 1.0 whenever n_1 equals n_2 and n_3 equals n_4, regardless of the magnitude of the transition between n_2 and n_3. That is, a sequence like c-c-d-d would produce a value of 1.0 (as would b-b-c-c or b-b-f-f). On the other hand, as with Rule 2b above, this implementation has the advantage of removing consideration of the units used to measure pitch height (as long as we consider linear transforms of the usual logarithmic interpretation of pitch perception). Since Lerdahl and Jackendoff did not provide any refinement, this implementation seems to represent the best balance that can be obtained at this time. Again, it should be noted that a non-linear scaling might be the more appropriate quantification for Rule 2b in the long run. In the same vein, it must be remembered that the scaling uses whole numbers to represent each semitone difference. Hence, a difference of an octave is treated as subjectively equal to 12 times the difference of a minor second. Over the limited ranges found in simple music, this seems reasonable, but it might not suffice for more complex works.

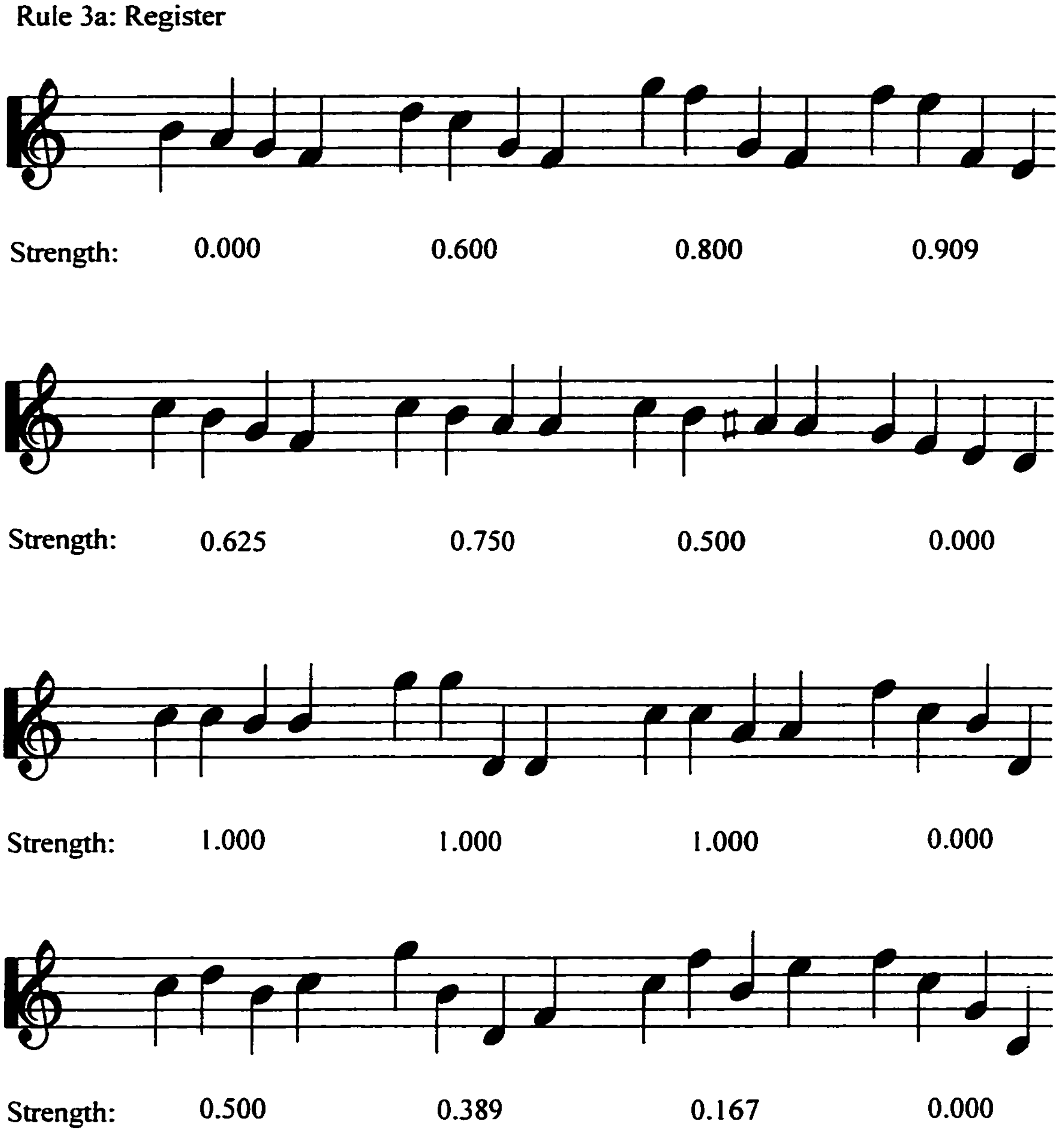


Figure 2.5: Quantification of the Group Preference Rule 3a of Lerdahl and Jackendoff (1983). Note that the rule produces a range of values between 0.0 and 1.0, and that it is symmetric in its application. Note that the change between the first and second notes must be less than the change between the second and third notes and that the change between the third and fourth notes must be less than the change between the second and third notes for the rule to apply.

GPR 3d (Length changes) is simpler to implement because it is more constrained in its application (in many ways, this rule seems like a special case of Rule 2b). As stated, the rule says that there should be a boundary whenever two notes having equal durations are followed by two other notes having equal durations that are different from the first two. The boundary is placed between the two pairs of notes. The only question to address is how to quantify this rule. It seems obvious that the amount of difference in the duration is a useful first step. Hence, the rule was quantified as:

boundary = $1.0 - (n_1 / n_3)$ if $n_3 > n_1$ where the n's are lengths (note that $n_1 = n_2$ and $n_3 = n_4$)

or

boundary = $1.0 - (n_3 / n_1)$ if $n_1 > n_3$ where the n's are lengths (note that $n_1 = n_2$ and $n_3 = n_4$)

This coding results in a 0.0 when the two pairs have the same length, and a value that approaches 1.0 when the lengths are very different (e.g., for whole notes compared with 32nd notes, the value is 0.97). Figure 2.6 presents some examples.

Having defined the basic algorithm to implement each rule, it was only necessary to write the programs that could implement them. Programs were written in-house using Borland C/C++ Version 3.0 to run under DOS[22]. A dedicated file structure (ASCII) was designed to code musical material (i.e., melodies); this file is similar to early file types (e.g., Creative Labs Sound Files; the MIDI file format was considered too complex for this work) for coding musical material. This file type contained the structures to address the usual notations of music (i.e., tied notes, slurs), but was limited to a single melody line.

---

[22] Although C was the language of choice, there does not seem to be any reason precluding the analysis of simple melodies within a program like Quattro or Excel. The format of the ASCII melody file is designed for such a purpose. This should enable an expansion of such analyses. Initially it was planned to include these algorithms within this document, but the amount of space required was prohibitive.

Rule 3d: Length

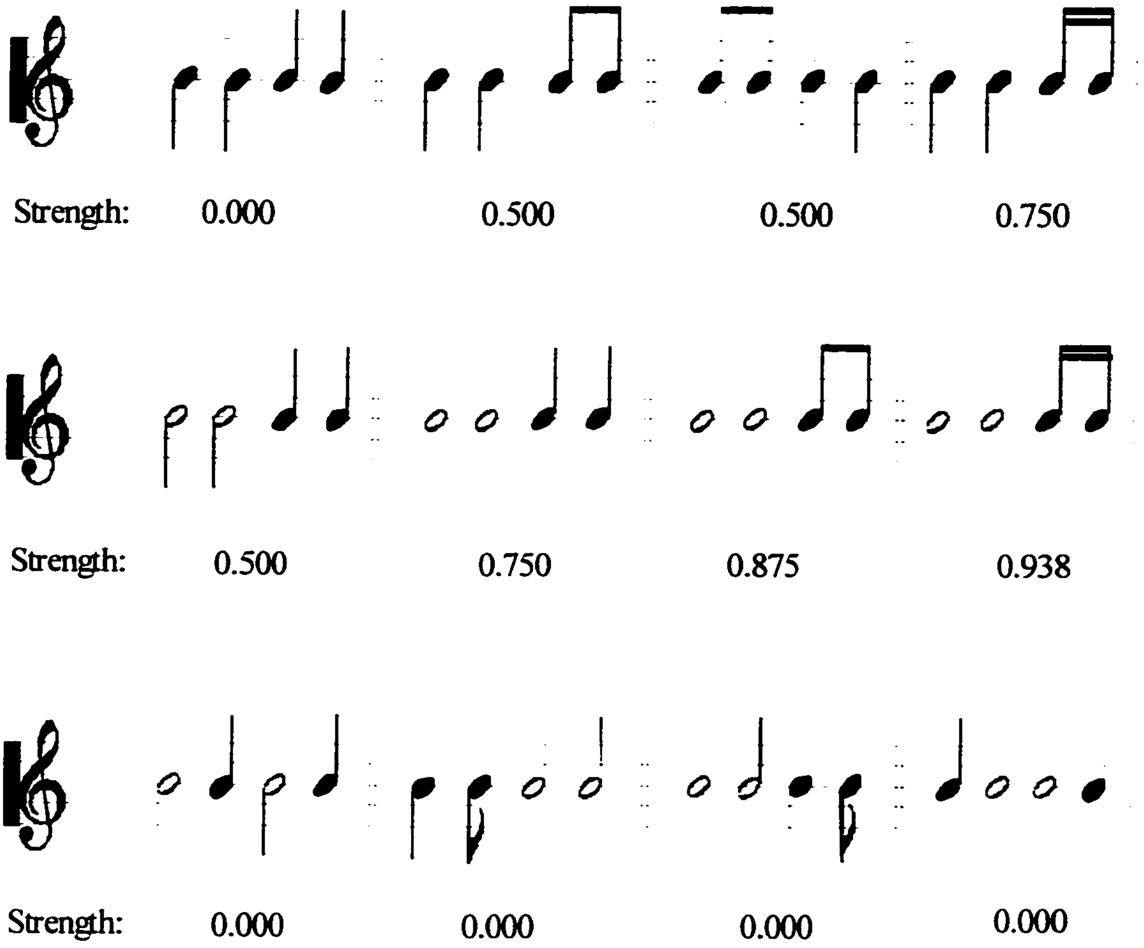


Figure 2.6: Quantification of the Group Preference Rule 3d of Lerdahl and Jackendoff (1983). Note that the rule produces a range of values between 0.0 and 1.0, and that it is symmetric in its application. Note that the rule does not apply unless the notes within each of the first pair and the second pair have the same durations

Application of Group Preference Rules 2 and 3

This section represents an initial attempt to apply the quantified versions of GPRs 2 and 3 to music. As such, this was not intended as a rigid test of those rules, but rather as a means of gaining some insights and intuitions concerning the applicability and utility of those rules. Several simple melodies were analyzed (were parsed algorithmically) and these analyses were visually inspected to see if they "made sense". In addition, the analyses were compared to the grammatical structure of the lyrics of the song. It was this type of analysis that help with various decisions, cited in the previous section, pertaining to the quantification of the rules (e.g., the method used to represent the rest aspect of Rule 2a). Simple melodies were analyzed even though this seems to run counter to the desires of Lerdahl and Jackendoff (1983). They argue that many folk melody and nursery rhymes exhibit a:

> "stereotypical" structure . . . [with] a maximal reinforcement of grouping preference rules. And here lies the danger for research. [sic] . . . the stereotypical structures are totally unrevealing, since they represent the confluence of a great number of interacting factors whose individual effects therefore cannot be identified. It is essential to begin with more sophisticated examples in order to arrive at any notion of what is going on.
>
> Lerdahl and Jackendoff, 1983, p. 67

Lerdahl and Jackendoff are correct in their concerns for the redundancy inherent in simple music. However, it is difficult to accept the argument that the confluence exhibited in simple melodies is not separable while the confluence that must be inherent is more complex melodies is separable. In their discussion of the dilemma facing music theory as a science, Brown and Demspter (1989, p. 98) raise a similar point:

> . . . on the one hand, music theory aspires to an objective understanding of musical phenomena and eschews extravagant and loose talk. On the other hand, theorists want to contemplate the so-called "masterpieces" of music, those consummate acts of "genius," by probing their infinitely rich detail and organization. These ambitions can work at cross purposes, the one encouraging and the other deploring the search for musical generalities and regularities.
>
> 1989, p. 98

People are not born with an understanding of complex classical pieces. This runs counter to numerous studies demonstrating the developmental course of musical understanding (cf., Krumhansl & Keil, 1982, but see also Krumhansl & Jusczyk, 1990). Understanding of more complex pieces must be predicated on the understanding of simpler pieces: Such is the basis for most methods of instruction in musical performance, composition and even musical appreciation (cf., Daniels & Wagner, 1975; Dalhousie Arts and Social Sciences Undergraduate Calender. 1993). Even the history of music has demonstrated a trend to more complex pieces (while retaining the simpler).

The approach taken herein is to take advantage of the stereotypical structures of simple melodies (hence boundary formation should be obvious and unambiguous) while minimizing the confluence by the examination of a limited set of rules. The material used had to be complex enough to be heard as music (i.e., not as a string of notes, or as a sequence of unrelated sounds), but simple enough to provide useful experimental data when presented to subjects (the actual experimental techniques are the topic of the next chapter). Melodies were chosen using two basic criteria. Melodies had to belong to the corpus of western-tonal music, since this is the domain of the theory: Even though the authors claim that their principles are, in some sense, universal, their theory is directed at western-tonal music. Admittedly, "western-tonal" is something of a judgement call, but the intent was mainly to eliminate those pieces that clearly did not belong within the idiom (e.g., music based on other scales [e.g., whole-tone, pentatonic, microtonal] or music based or played on non-standard instruments [e.g., industrial machines, environmental sounds]). Secondly, melodies had to be reducible to a single melody line. In this regard, melodies arranged for the formative years of piano instruction were most useful.

To maintain musicality while minimizing the number of applicable rules, several aspects of grouping that are under the control of the performer were eliminated from the analysis since they can also be eliminated from the presentation of the stimulus (when that stimulus is presented to subjects for parsing). This means that parsing based on dynamics and articulation (including slurs and subtle timing differences) were not considered. Even though there is some loss, melodies can be played at a constant intensity with precise, invariant timing of notes (relative durations and interstimulus intervals) and still retain the character of music. Metrical structure was also not analyzed since metrical accents can be minimized in the performance of a piece (i.e., the theory implies that grouping and metrical structures can be separated). Beats (implicated by intensity changes or subtle changes in the onset times, or by subtle increases in duration) were not provided in the presentation of the melodies. This left, essentially, four rules to examine: Rule 2a (Rest), 2b (Attack-Point), 3a (Register Changes) and 3d (Length Changes). The role of these four rules in a number of pieces was examined.

The algorithmic application of the rules is first compared to Lerdahl and Jackendoff's (1983) analysis of the opening of Mozart's G Minor Symphony (see Figure 2.7). In this comparison, it must be remembered that Lerdahl and Jackendoff place boundaries between units, while this algorithm indicates boundaries on the first note of corresponding unit. In Figure 2.7, the notation of Lerdahl and Jackendoff has been shifted to align with the algorithm (cf., Figure 2.2).

Generally the program finds the same boundaries as indicated by Lerdahl and Jackendoff, but with a few differences. Firstly, the program did not analyze for slurs, so the program did not indicate the first 2a boundary (between Note Events 8 and 9). The program located all the 2b boundaries. The program found two unmarked 3a (Register change) boundaries. After checking, it was determined that both these boundaries fit within the definition of Lerdahl and Jackendoff. This highlights the distinction between a rigorous algorithm and a subjective interpretation. The staff notation is not ideal for the detection of subtle pitch differences because some steps on the staff represent semitone differences (i.e., b to c, d to e) while others represent whole tone steps (i.e., from f to g to a). The algorithm did not find the 3a (register change) boundary that Lerdahl and Jackendoff imposed (Note Event 10). This is not an error, but rather an interpretational ambiguity. The rule states that there should be four notes (not four notes with a rest in the middle) with a most dramatic change between the second and third notes. Their analysis placed a boundary within a four-note span that contained a rest. A rest is a note event having, in principle, no defined pitch (i.e., a rest is qualitatively different from a note).

Lerdahl and Jackendoff, in their other rules (i.e., GPRs 2a and 2b), imply that the absence of sound (i.e. the rest) is an important phenomenal accent, and yet, in this example, treat the absence of sound as irrelevant. I chose to try to maintain consistency with the overall design of their theory, such that, in the algorithm, the rule does not apply when there is a rest within the span. Also, the inclusion of rests within Rule 3a would further confound the analysis of the rules since rests would then be involved in Rules 2a, 2b and 3a. The algorithm did not analyze for Rules 3b or 3c and the algorithm found both the instantiations of Rule 3d.

In general, the algorithm produces the same analysis as Lerdahl and Jackendoff (1983), with two exceptions. The algorithm also treats rests within Rule 3a in a manner that seems more consistent with the whole of the theory. The algorithm also quantifies the rules and as such is more sensitive to subtleties that may slip past the human eye. This is not so much a critique of Lerdahl and Jackendoff as a comment on human perception. Deliege (1987, Fig. 14) also seems to have applied rules somewhat idiosyncratically. A boundary is labeled Rule 3d (Length Change), when a stronger interpretation might be Rule 3? (Timbre). Later (Fig. 17) boundaries are attributed to pairs of identical sounds when it might be more appropriate to label them as Rule 3a (Register), as they would be by the algorithm developed herein. In the examples of Peretz (1989) there are several places where the algorithm would detect an application of Rule 3a (Register) or 3d (Length) or even Rule 2b (Attack-Point).

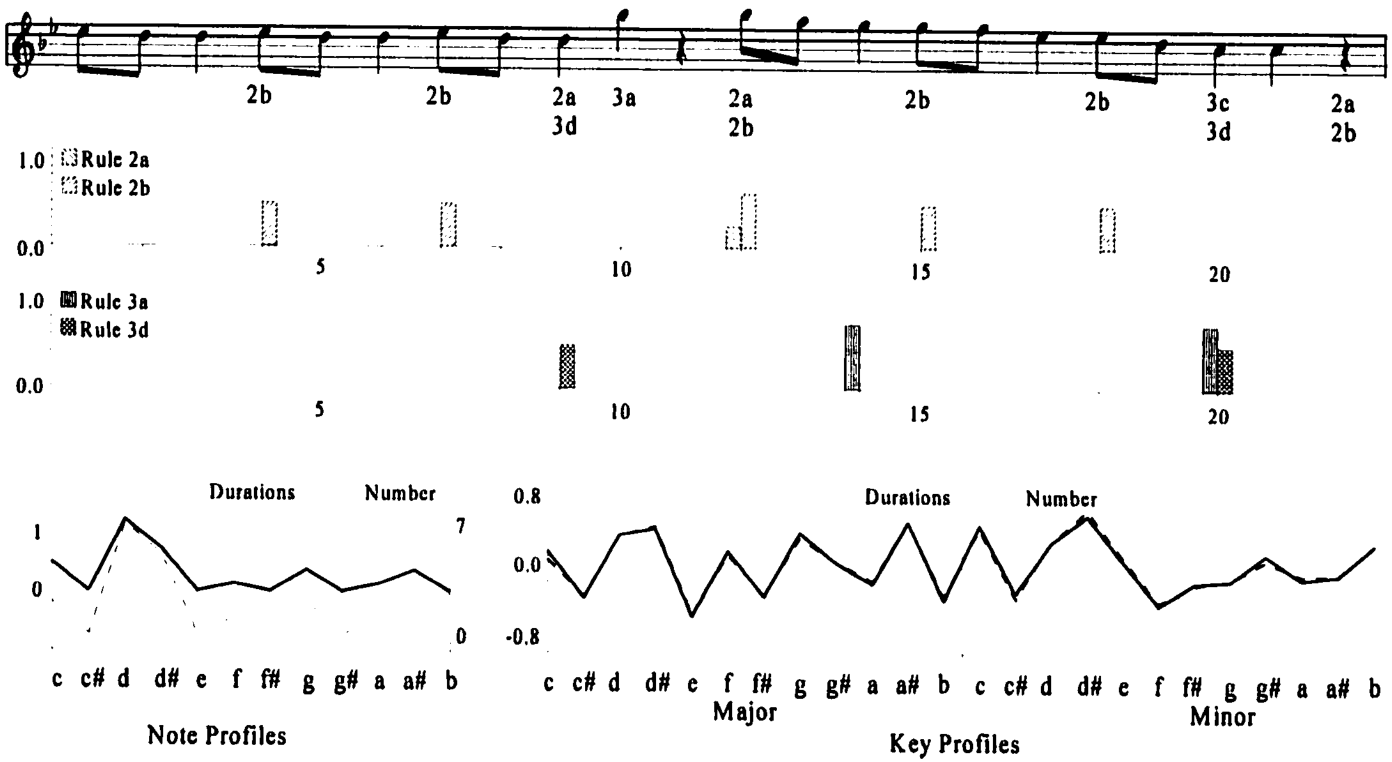


Figure 2.7: The analysis of the opening of Mozart's "G Minor Symphony", including Rules 2a (Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the melody Rules 2a (Slur), 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided. Note that the algorithm finds 3a boundaries where none had been placed by Lerdahl and Jackendoff (1983). Note that Lerdahl and Jackendoff's boundaries are between notes, while the algorithm places them on the first note of the subsequent unit.

In addition, the algorithm allows one to examine other aspects of parsing, aspects that are only mentioned in passing by Lerdahl and Jackendoff. Firstly, one can examine the degree of "confluence" in the application of the rules to the melody. The note-by-note correlations of the boundaries detected by the rules is such a measure. High confluences would be high correlations, while low confluences would be low correlations. For this melody, the correlations are presented in Table 2.3.

Table 2.3: The intercorrelation between the application of the low-level Group Preference Rules, in the opening of Mozart's G Minor Symphony.

	Rule 2a	Rule 2b	Rule 3a	Rule 3d
Rule 2a (Rest/Slur)	1.000	0.499	-0.069	-0.069
Rule 2b (Attack-Point)		1.0000	-0.171	-0.171
Rule 3a (Register Changes)			1.000	0.450
Rule 3d (Length Changes)				1.000

Note that the correlations between Rules 2a and 2b and between Rules 3a and 3d approach r=0.5. (these can be seen in Figure 2.7). As a assessment of confluence, the figure is more informative. Note that the only occurrence of a boundary due to Rule 2a coincides with one of those indicated by 2b. This is the source of the correlation. The same relationship is true for Rules 3a and 3d. However, Rules 2a and 2b do not overlap with Rules 3a and 3d. For further insights, and as a check on the criteria of adherence to western-tonal idiom, the note profile and the key profiles are also provided. There are two versions of the note profile: The first is based on the summed durations of all notes, and the second is based on the total number of occurrences of each note. These two profiles often align: That is, greater total durations are associated with greater total number of occurrences. In this melody, the correlation is actually r=0.99 (r^2=0.97). From the profiles, one can use the Krumhansl (1990; see Frankland & Cohen, 1996) key-finding algorithm to obtain the relative measures of key strength or association. On the basis of the duration profile the best key is e♭-minor (*r*=0.59), which is next to the notated key of b♭-minor on the circle of fifths. The second best key is b♭-major (*r*=0.51), producing a q-factor of 0.09 (the difference between the squared correlations for the two best keys). For the number profile, the same keys are found with correlations of *r*=0.65 and *r*=0.51 respectively and q factor of 0.16. The implication is that key is defined, though it is not particularly strong (one must remember that this is an extract from a larger work. As such it may not contain all the information that the original did). Unfortunately, Lerdahl & Jackendoff did not present many examples with simple melodies, so it was necessary to draw upon other sources for insights.

Ultimately, the rules will be compared to the parsings of subjects, but before that was attempted, it was considered useful to have some idea of which rules and which melodies provide the best (or worst) opportunities for experimentation, with "best" being left as a vague concept. Although it might be possible to refer the melodies to a panel of experts, disparities between the rules and the panel might be more reflective of differences in opinions than inadequacies of the rules. A more appropriate approach was to compare the parsing of the melodies based on the rules to the parsing of the lyrics associated with each melody. This tacitly assumes that there are lyrics to parse. It can be shown that the lyrics and the phrasing of melodies coincide (cf., Christ, DeLone & Winold, 1975 pp., 48-81; Serafine, Crowder & Repp, 1986; see also *The New Groves Dictionary of Music*, 1980). The analysis is also a useful way to model the pattern that might be produced by subjects' parsing of melodies. That is, we would expect the parsing of melodies by subjects to resemble the parsing of melodies based on the grammar of the vocal line, and we can use an examination of the grammatical coding to highlight how such an analysis can be used. This analysis was never intended as a rigid test of the theory. It was considered only as a fruitful (and resource efficient) means to verify the implementation of the algorithm and the broad principles of the theory. It is useful because it can be applied quickly to a large number of works -- empirical tests with subjects will have to be much more selective of material and this analysis helps in that regard.

The analysis of the grammar of the lyrics of melodies is difficult because such lines do not follow the rigid rules of grammar, but rather, must be treated as poetry (not as prose). Notwithstanding such concerns, the lyrics of the melodies were parsed, using the grammatical syntax of the individual who arranged each piece, within a simple six level scheme:

single words -- level 1
two or three words -- level 2
phrase -- level 3
dependent clause -- level 4
independent clause -- level 5
sentence -- level 6

Though admittedly crude, the scheme was found to be adequate to the task[23] because the goal was to simply provide a firmer base for intuitions for the direction of further work. That is, this was not considered an end-point, nor in fact, was it ever thought that this type of analysis could be refined sufficiently to be an end-point. The melody and the words of a melody likely correlate in a large amount of music, but music is not constrained to this and composers often depart considerably from this.

In the following melodies, the same analysis as was used with the example from Lerdahl and Jackendoff (1983) is applied. That is, each melody is presented with its analysis for Rules 2a, 2b, 3a and 3d. The associated grammatical parsing of the lyrics is also presented. The note profiles and the key profiles are provided as an estimate of the degree of tonality exhibited by a piece (cf., Frankland & Cohen, 1996). Finally, the correlations between the rules, including the grammatical coding are provided. The list of melodies, their sources, their notated key and their best key (cf., Krumhansl, 1990) and q-factor determined from the Krumhansl key-finding algorithm (cf., Krumhansl, 1990, Frankland & Cohen, 1996), the correlation between their duration- and number-note

---

[23] Initially, two (either present or absent) and three level schemes were attempted, but these did not seem to provide enough range to help.

profiles and their associated figure number are presented in Table 2.4.

Table 2.4: The background details for the melodies analyzed.

Name	Figure No	No of Notes	Notated Key	Best Key	Key Strength	Q-Factor	Profile r^2
The Mulberry Bush[1]	2.8	36	C-Major	C-Major C-Major	0.691 0.694	0.281 0.271	0.979
Tom, Tom, the Piper's Son[1]	2.9	30	C-Major	C-Major C-Major	0.795 0.773	0.484 0.323	0.927
Three Blind Mice[1]	2.10	48	C-Major	G-Major G-Major	0.900 0.853	0.482 0.318	0.793
Mary Had a Little Lamb[1]	2.11	53	C-Major	G-Major G-Major	0.256 0.211	0.001 0.018	0.980
Away in a Manger[2]	2.12	44	G-Major	D-Major D-Major	0.745 0.788	0.073 0.092	0.967
Silent Night[2]	2.13	46	C-Major	C-Major G-Major	0.776 0.656	0.242 0.091	0.785
Jingle Bells[2]	2.14	103	C-Major	C-Major C-Major	0.441 0.394	0.192 0.144	0.970
Wish You a Merry Christmas[2]	2.15	54	C-Major	G-Major G-Major	0.683 0.752	0.046 0.100	0.967

Notes: Key strength is the square of the correlation returned by the key-finding algorithm.

In all cases, the associated key values based on duration-note profile is reported first and then the value based on the number-note profile.

When the best key disagreed with the key signature, the second best key was invariably the key signature, except for "Three Blind Mice" in which case, it was D-Major.

[1] From *Nursery Melodies at the Piano* arranged by J. Bastien.

[2] From *Favorite Melodies the World Over* arranged by J. S. Bastien.

Note that all melodies, with the exception of "Mary Had a Little Lamb" had well defined tonal centers (defined by the Krumhansl key-finding algorithm: the square of the correlation with key profiles), and high q-factors (the difference between the best and second best keys). The results were the same regardless of the use of note durations or number of note occurrences (the duration profile and number profile were always highly correlated). For each of these melodies, the correlations between the rules and the grammatical coding of the lyrics are presented in Table 2.5

Table 2.5: The correlations between the rules and the grammatical coding of the lyrics of the melody.

Name	Rule 2a	Rule 2b	Rule 3a	Rule 3d	Misses
The Mulberry Bush	---	0.724	0.036	-0.056	0/4
Tom, Tom, the Piper's Son	---	0.897	0.062	0.067	1/6
Three Blind Mice	---	0.622	0.147	---	1/7
Mary Had a Little Lamb	---	0.880	0.424	---	1/8
Away in a Manger	---	0.464	-0.109	---	0/4
Silent Night	---	0.332	0.135	0.556	0/7
Jiggle Bells	0.623	0.806	-0.134	-0.085	0/14
Wish You a Merry Christmas	---	0.602	-0.135	-0.151	2/4

Notes: The dashes indicate that the rule did not apply in this melody.
Misses is the number of places in which there was a grammatical pause in the lyrics that did not correspond to any boundary based on the rules: This is represented as the number out of the total number of times that there was a grammatical pause in the lyrics (it is called a miss because it implies that the theoretical predictions missed a boundary location).

From this table, one can see that all eight melodies cited used Rules 2b (Attack-point) and 3a (Register Change), five used Rule 3d (Length Changes) and only one used 2a (Slur/Rest; slurs were not analyzed but rests were). The findings are consistent with the statements of Lerdahl and Jackendoff (1983, p. 47); the low emphasis on Rule 2a can be explained by the decision to not examine slurs in this work. Table 2.5 implies that Rule 2b demonstrates the most promise: Rule 2b is generally correlated with the grammatical parsing of the lyrics of the melody, while other rules are not.

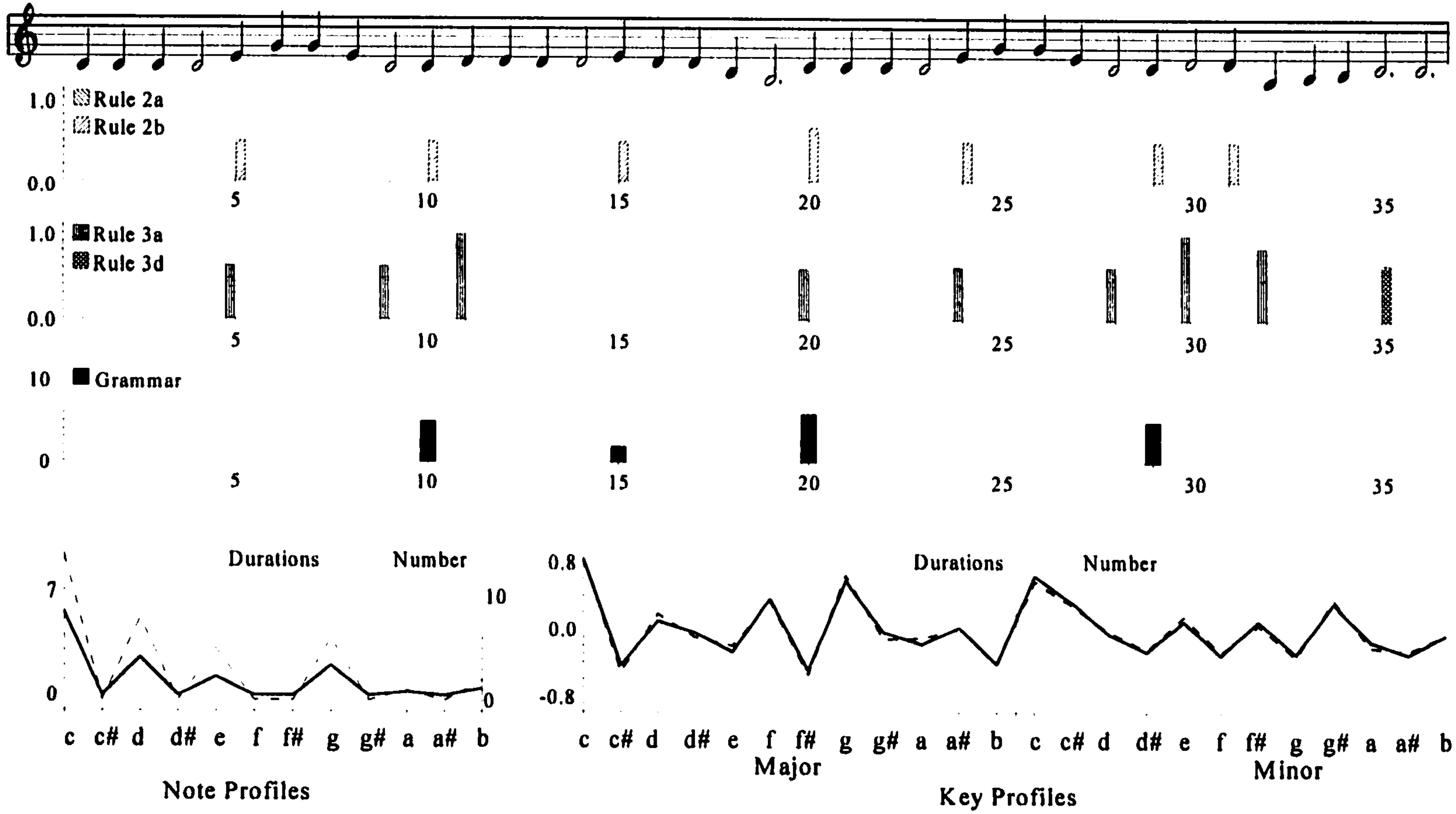


Figure 2.8: The analysis of "The Mulberry Bush", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

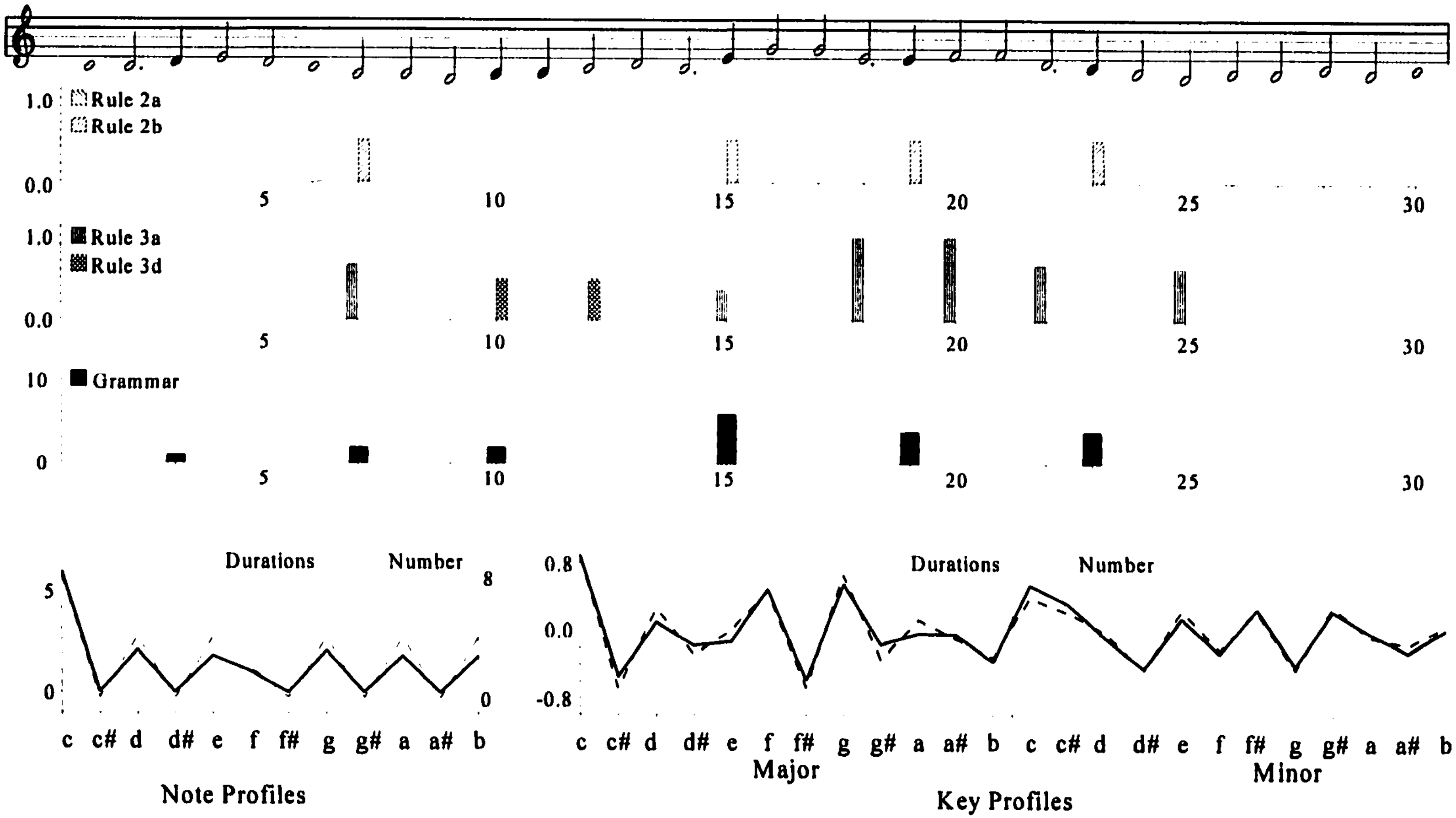


Figure 2.9: The analysis of "Tom, Tom, The Piper's Son" including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

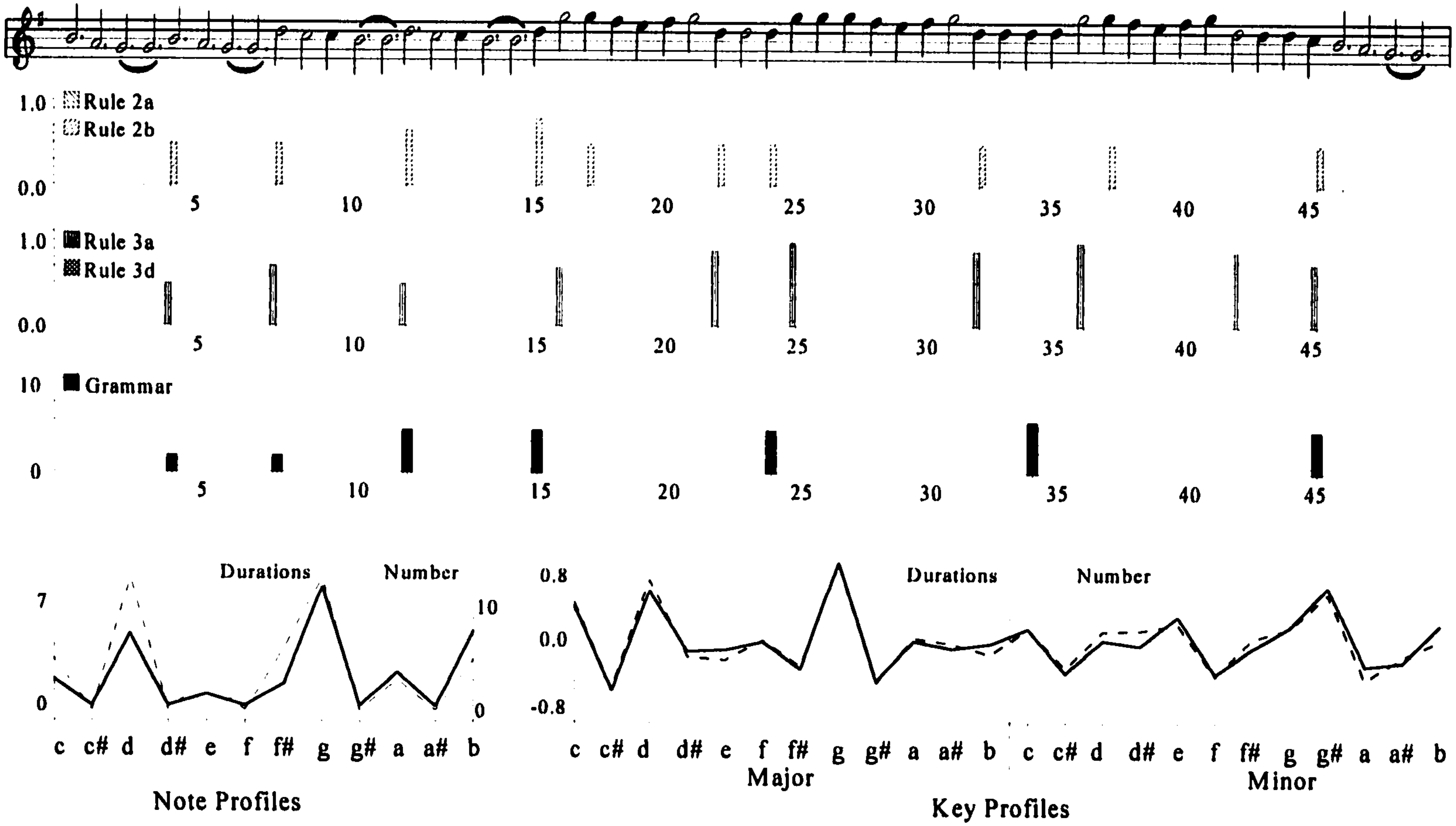


Figure 2.10: The analysis of "Three Blind Mice", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

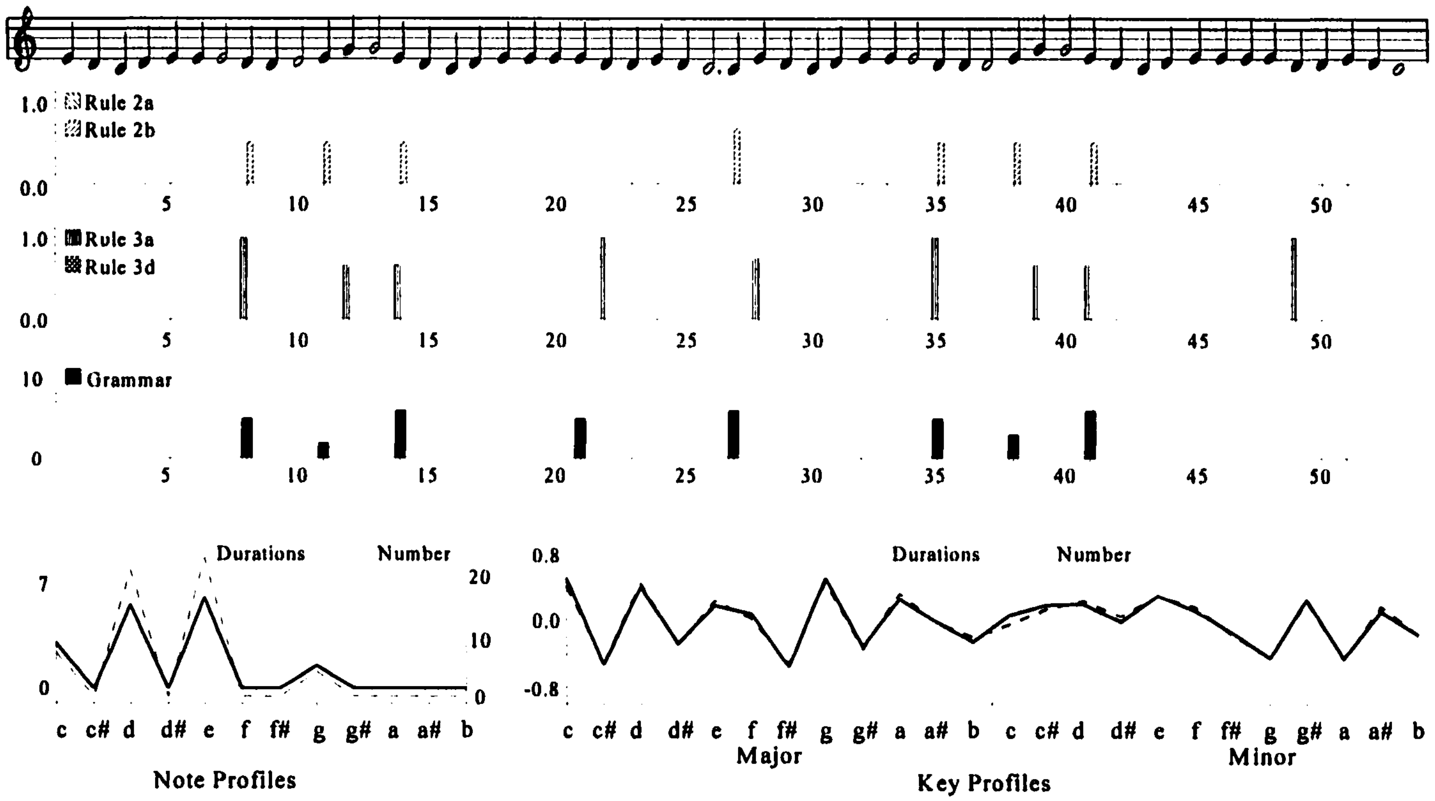


Figure 2.11: The analysis of "Mary Had a Little Lamb", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

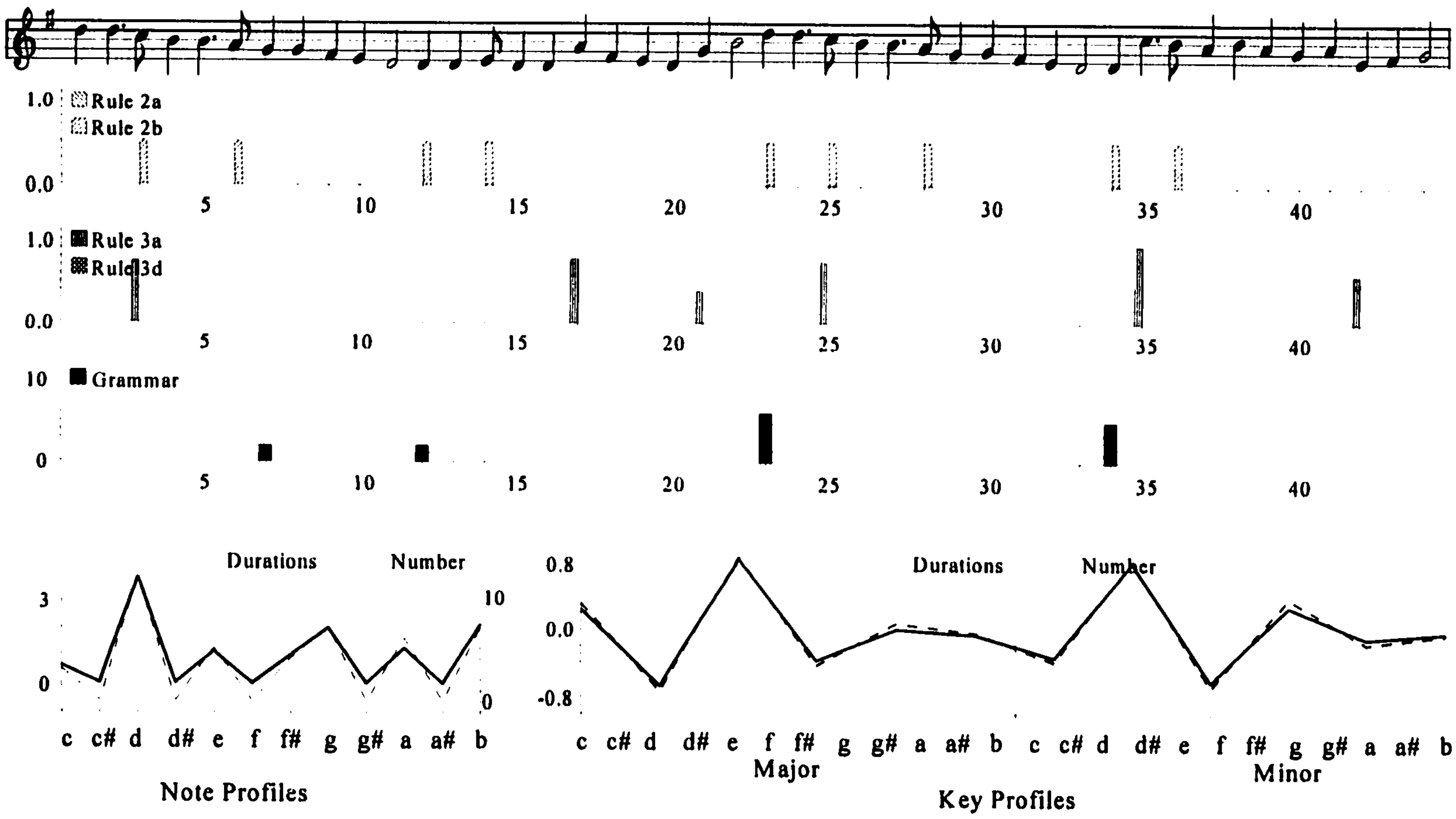


Figure 2.12: The analysis of "Away in a Manger", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

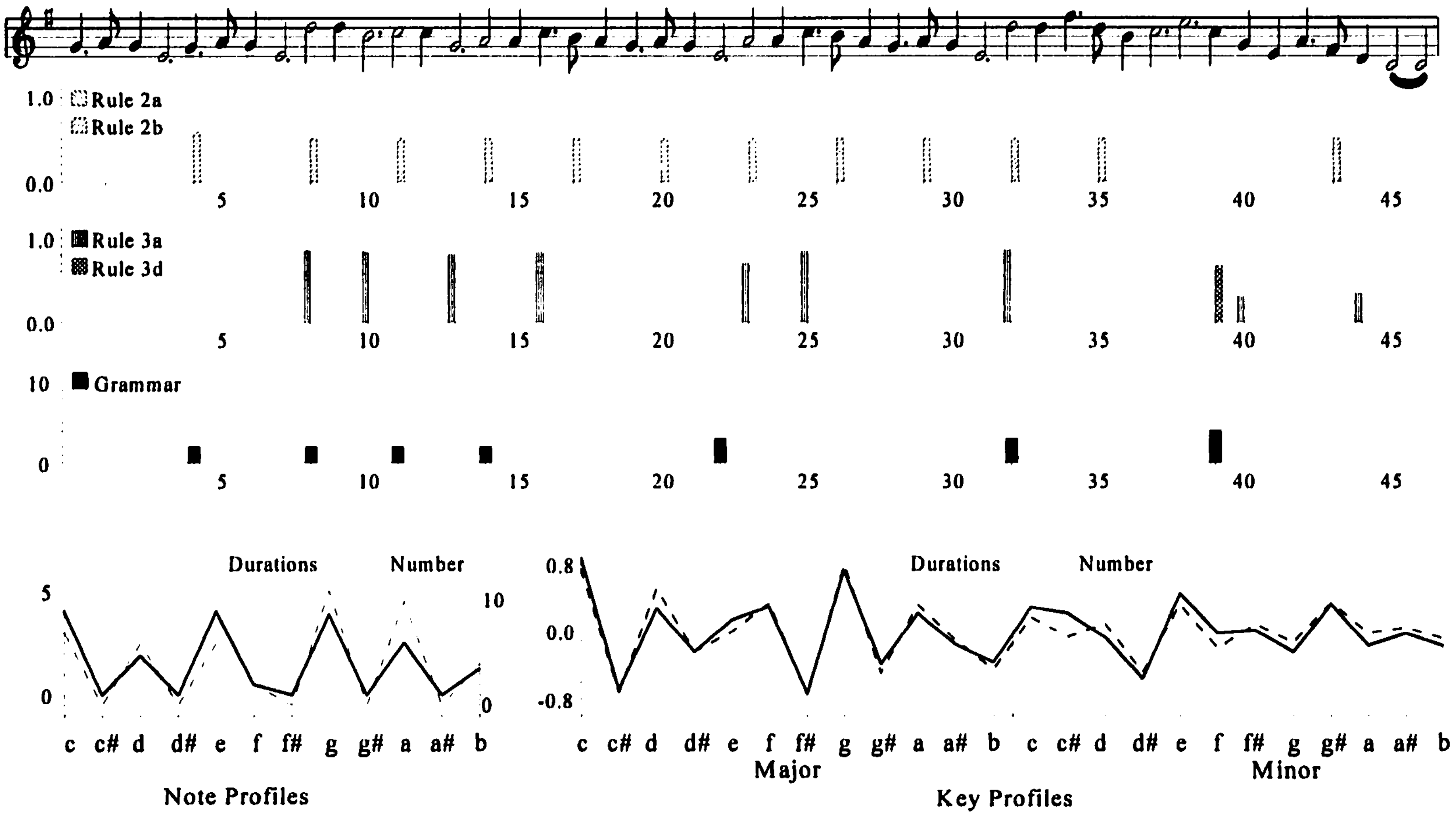


Figure 2.13: The analysis of "Silent Night", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

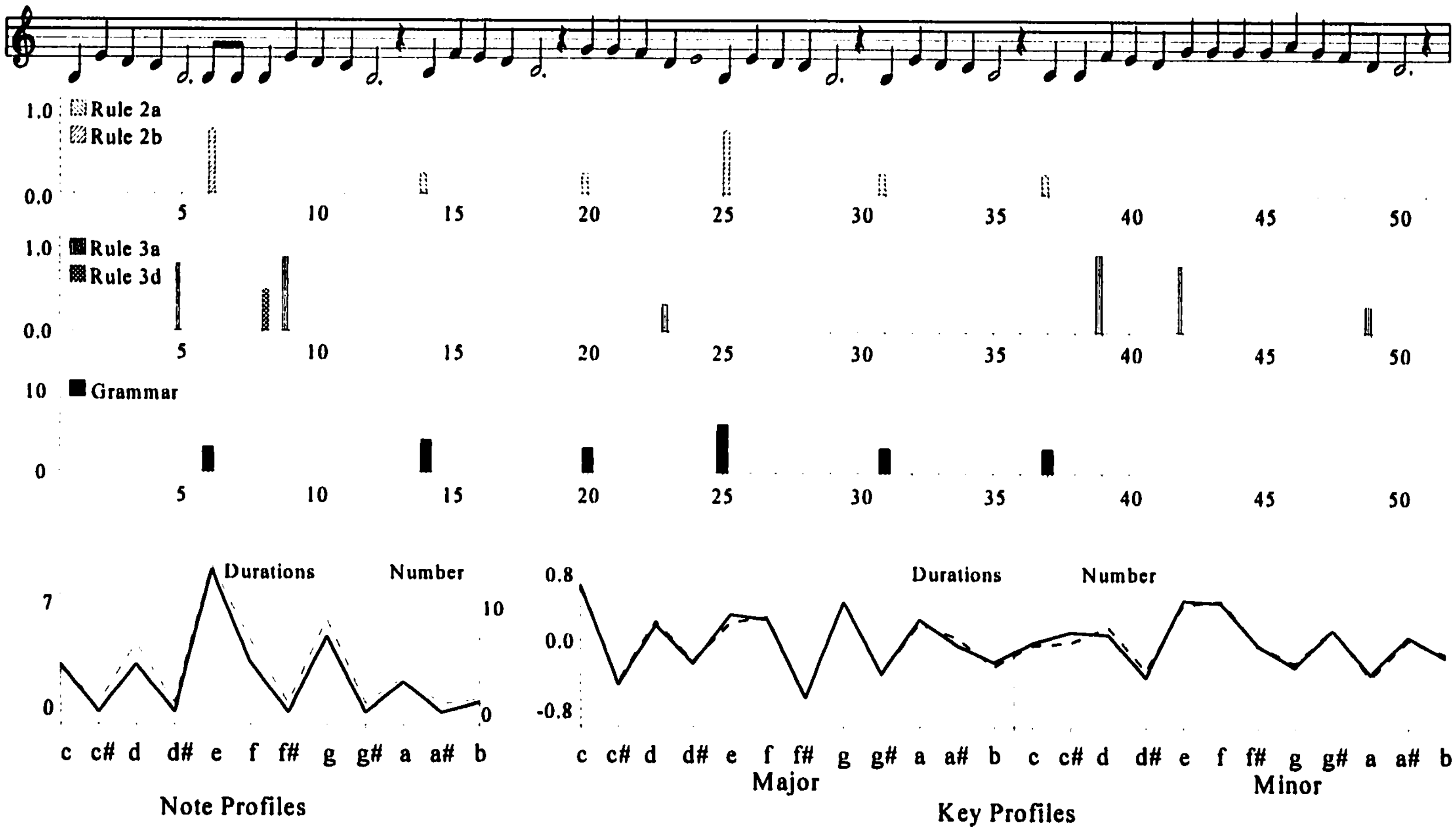


Figure 2.14a: The first half of the analysis of "Jingle Bells", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

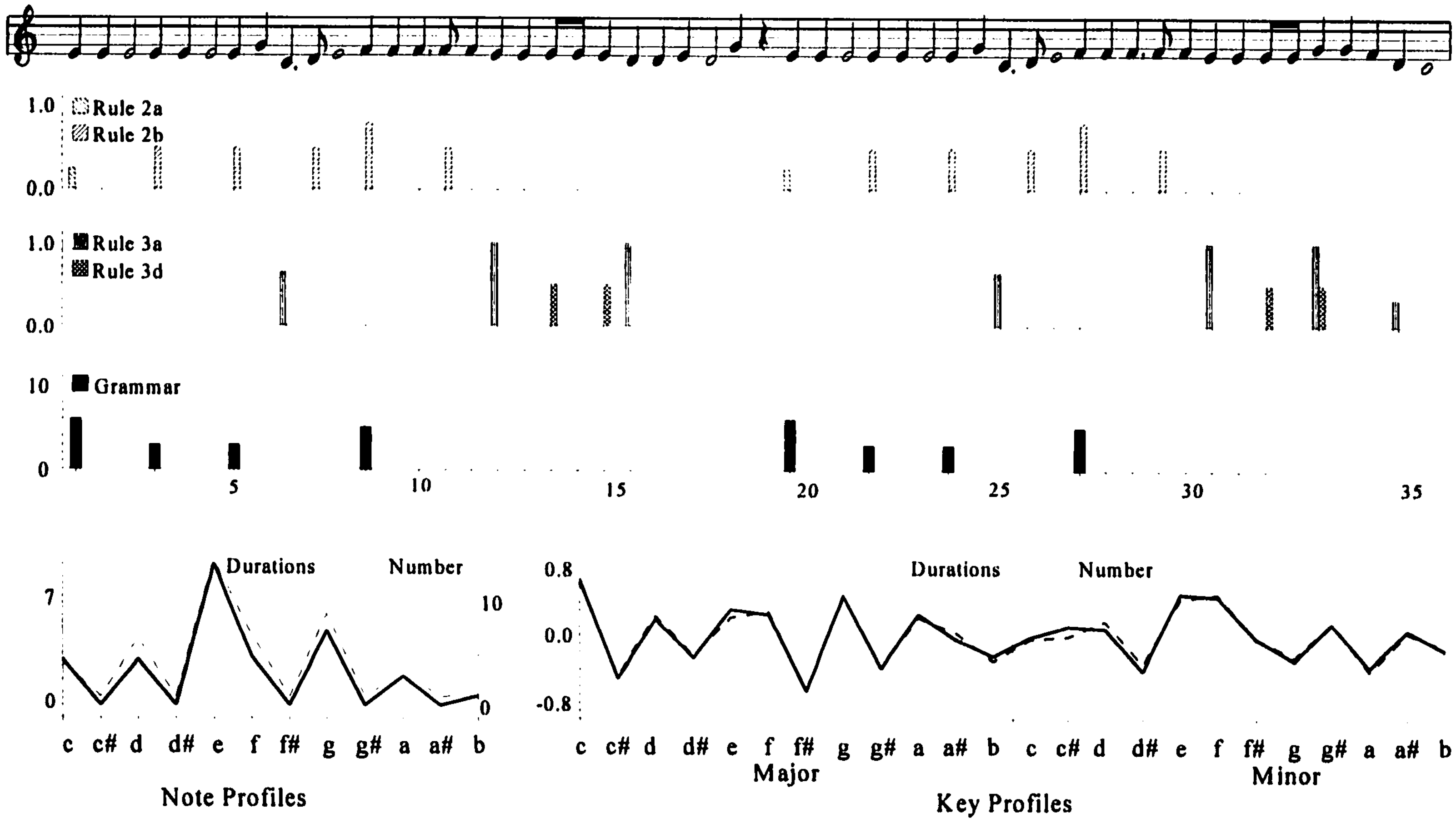


Figure 2.14b: The second half of the analysis of "Jingle Bells", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance

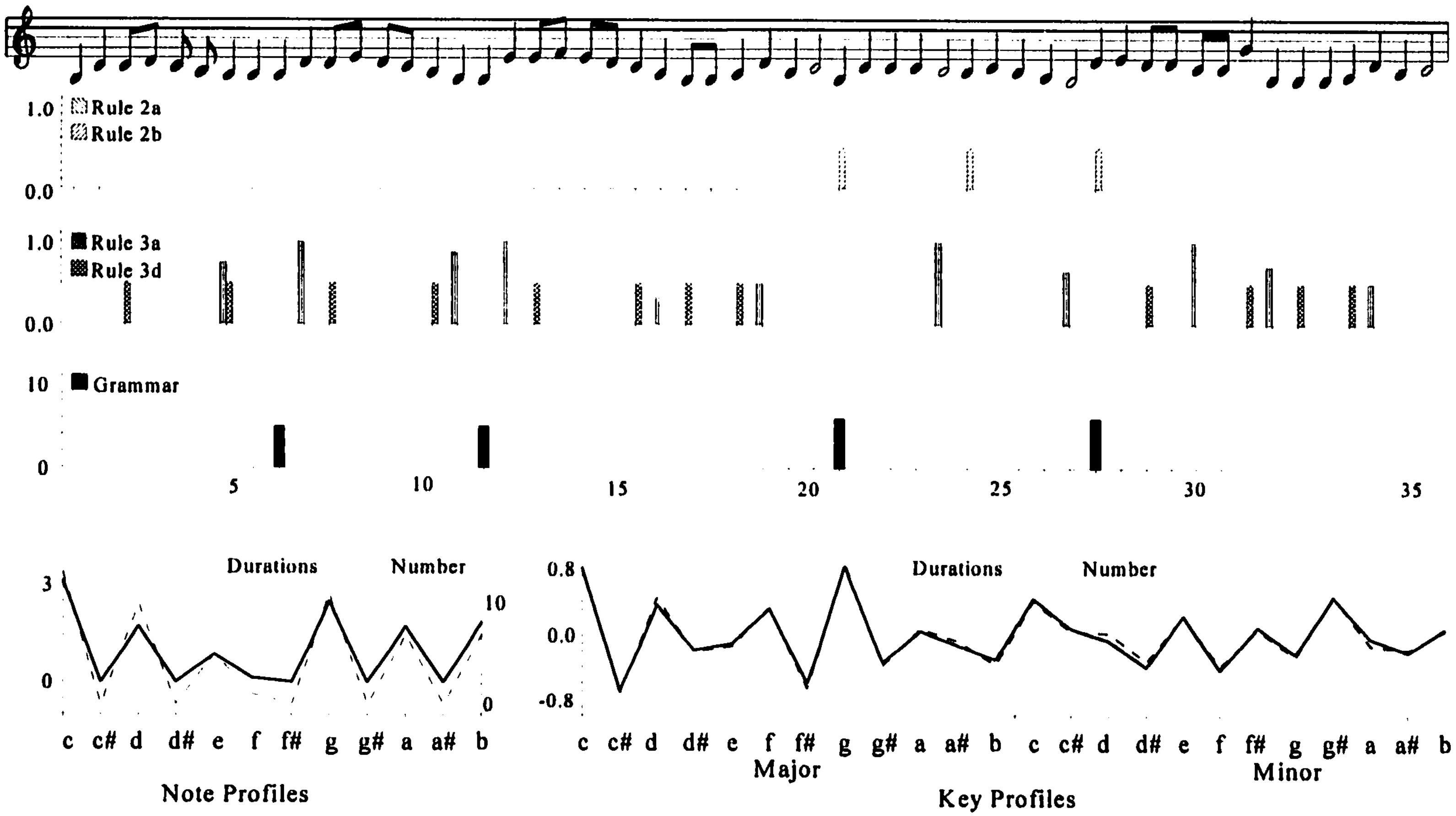


Figure 2.15: The analysis of "We Wish You a Merry Christmas", including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

"Mary Had a Little Lamb" is the only melody to "use" Rule 3a as a marker for grammatical pauses. "Silent Night" is the only melody to "use" Rule 3d as a marker for grammatical pauses. In both cases the word "use" is in quotations because one cannot expect the grammatical coding and the rule-based boundaries to necessarily coincide, although one would expect that they generally coincide. Also presented in Table 2.5 is the number of times there was a grammatical pause without any boundary based on any of the rules (i.e., a miss). Although, this will be more important when grammatical parsing of the lyrics is replaced by the actual parsing of subjects, it can be seen that for three of the eight melodies, there was one miss and one of the eight melodies had two misses. This is not a positive sign for the theory.

Pretend for a moment that the grammatical parsing is actually the parsings indicated by subjects based on listening to the melodies, with all the aforementioned controls (precise control of the timing of notes: no phenomenal or metrical accents; no changes in dynamics, no changes in articulation). These misses would imply that Lerdahl and Jackendoff have missed at least one basic rule for parsing. One might try to argue that these misses were the instantiation of rules not tested. This is not a valid argument because the rules not tested do not exist in the melody (i.e., slurs do not exist because notes do not run together, dynamic changes are controlled, articulation is controlled). One might then try to argue that these misses represent parsing on the basis of memory for the usual presentation of the melody which does contain those other rules. This is possible, and to control it, subjects must be asked to parse unfamiliar melodies[24]. To be fair, Lerdahl and Jackendoff end their presentation of these low-level group preference rules with, "One might add further cases to deal with such things as changes in timbre or instrumentation" (1983, p. 46), but these, too, could be controlled at the level of the stimulus without affect the perception of the melody as music.

---

[24] One might still argue that misses in unfamiliar melodies is based upon analogical parsing of similar material, but each step away from the original intent (parsing based on auditory events), each post hoc rationalization, makes the theory as a whole less tenable.

Table 2.6 presents a measure of construct that Lerdahl and Jackendoff described as confluence. Note that none of the correlations is particularly high, implying that there is not much confluence. The interesting question will be whether or not those few occasions (where the rules do indicate boundaries on the same note) produce more empirical boundaries.

Table 2.6: The correlations between the Group Preference Rules 2a, 2b, 3a and 3d.

Name	Rule 2a Rule 2b	Rule 2a Rule 3a	Rule 2a Rule 3d	Rule 2b Rule 3a	Rule 2b Rule 3d	Rule 3a Rule 3d
The Mulberry Bush	---	---	---	0.174	-0.082	-0.088
Tom, Tom, the Piper's Son	---	---	---	0.149	-0.105	-0.127
Three Blind Mice	---	---	---	0.369	---	---
Mary Had a Little Lamb	---	---	---	0.385	---	---
Away in a Manger	---	---	---	0.143	---	---
Silent Night	---	---	---	0.120	-0.088	-0.070
Jingle Bells	0.521	-0.089	-0.562	-0.155	-0.098	0.090
Wish You a Merry Christmas	---	---	---	-0.116	-0.130	-0.152

Notes: Dashes indicate that the rule did not apply (i.e., there were no places in the melody at which the rule applied).

The final melody presented (Figure 2.16) is an extract from instrumental music. This melody exhibits the aforementioned properties: A defined tonal centre and correlations between the rules. This melody will be used in subsequent experiments.

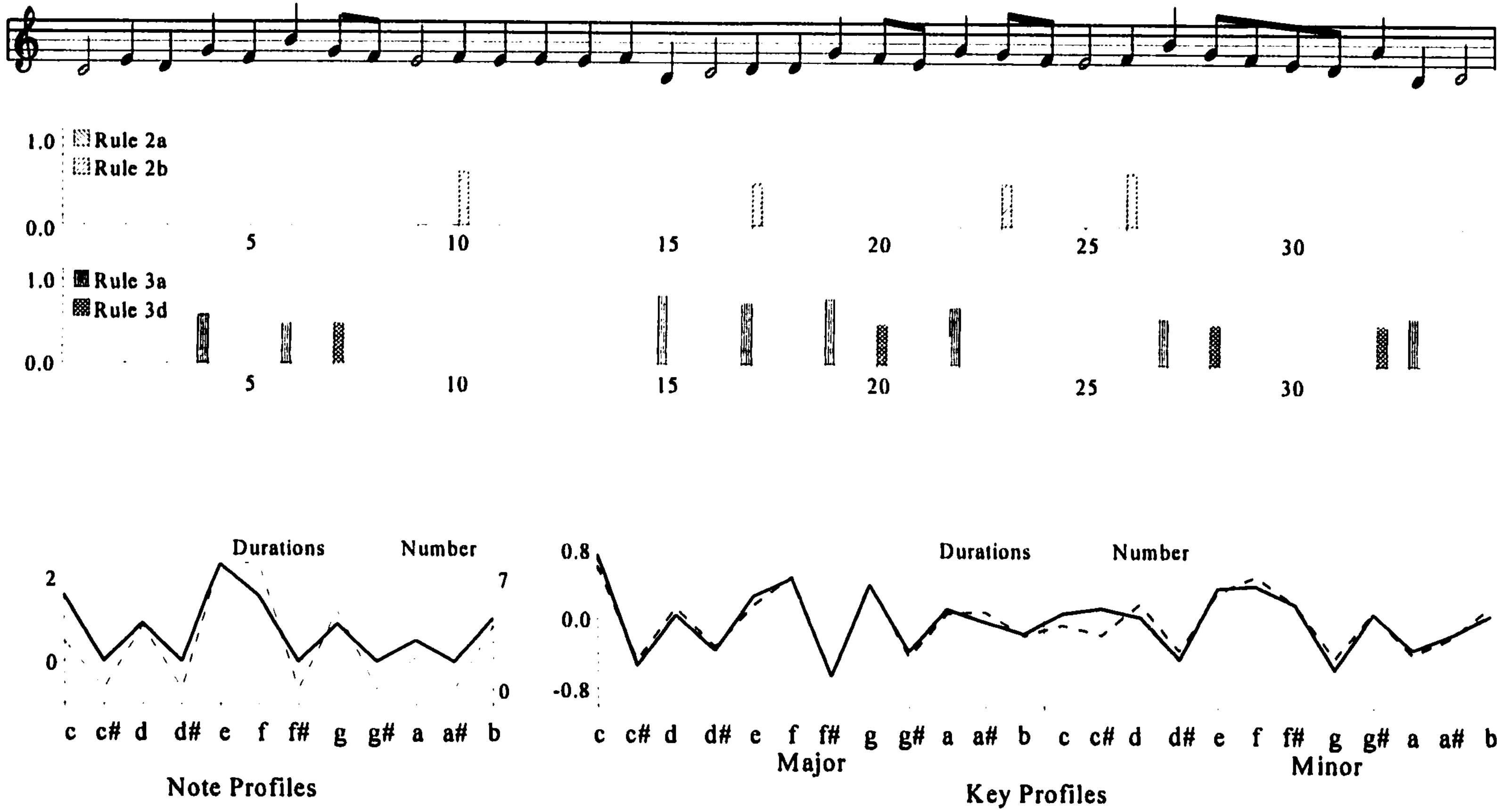


Figure 2.16: The analysis of the extract from "Softly Now the Light of Day" (G. Doane), including Rules 2a (Slur/Rest), 2b (Attack-Point), 3a (Register) and 3d (Length), as well as, the grammatical coding of the song. Rules 3b (Dynamics) and 3c (Articulation) are not included because these are aspects that are determined by performance. The note profiles by total duration (equivalent to the number of whole notes: left axis) and total number of occurrences (irrespective of individual durations: right axis) and the associated key profiles (see Krumhansl, 1990) for the entire song are also provided.

General Comments, Concerns and Critiques

Although Lerdahl and Jackendoff (1983) can offer some powerful insights into the parsing of music, there are several important concerns. Some of these deal with global aspects of the theory, while others deal with the local aspects of the theory.

The first concern is one of the overall perspective. Even though they claim their theory is a theory of perception (1983, pp. 1-11), Lerdahl and Jackendoff clearly take the perspective of a composer (i.e., their notion of the "experienced listener" [1983, p. 7] -- not necessarily an individual with training) who has the musical score to examine. The observation has been made by others. For example, Cook cites Rosner's review of Lerdahl and Jackendoff's text:

> the intuitions of a sophisticated listener *who has the score of the piece!* The present theory (and others that claim perceptual validity) generally rest on the tacitly accepted but little noticed use of scores . . . Lerdahl and Jackendoff are not alone in assuming that complex hierarchical trees in *visual* form are isomorphic to the auditory results of listening, which necessarily occurs across time and involves memory. The assumption seems very dubious psychologically, when carried to higher levels.
>
> Rosner (1984), cited in Cook (1989), p. 119

Cook's concern is with the larger structures of time- and prolongation-reduction. However, even though the present work has focussed on the low-level parsing that does not depend on memory to the same extent, it is not possible to ignore completely the broader issue of perspective. The problem is that Lerdahl and Jackendoff may have inadvertently "created" low-level rules and "ignored" other possible low-level rules in order to make their global structure match their beliefs about the global structure.

If Lerdahl and Jackendoff (1983) have generated one (or more) useless rule, then such a rule will generate false positives. False positives occur when a low-level rule predicts a boundary that does not match with empirical observations. The interpretation of false positives must be taken with great care since all theories can reasonably be expected to generate false positives. It is only possible to eliminate a rule (from the set of all rules) if that rule generates nothing but "absolute" false positives (empirical boundaries are never located where the rule predicts). Anything less than this absolute requires interpretation. A rule may generate a lot of false positives -- a rule may only predict a single boundary within a piece -- but it could still be an important rule. That is, it is possible for a rule to be used rarely, but be important (thermal nuclear explosions are hopefully rare, but extremely important to human existence). This situation is more important if the particular rule in question is the only rule that can predict a particular boundary. In this situation, it is an important rule. On the other hand, if a rule generates a lot of false positives, and other rules generate predictions at the same empirical boundaries, then the rule is likely unimportant.

On the other hand, if Lerdahl and Jackendoff (1983) have missed an important rule, then their theory as a whole will generate a miss (an empirical boundary where none has been predicted). This is a more difficult case to interpret since the implication is that the cause of boundary formation (i.e., the selection of particular locations from the set of possible locations) is not known. In this case, the theory will have to be adjusted, assuming that it has been properly implemented and assuming that subjects have indicated the locations of boundaries.

The second concern focuses on the basic dimensions that Lerdahl and Jackendoff (1983) have identified. They use changes in time, pitch (and timbre) and intensity. Time pattern and intensity present few concerns, but the specific use of pitch and not the diatonic set or higher order rules of tonality is equivocal. The problem is that pitch is not tonality, and yet most would grant that tonality (or diatonicism) is a critical dimension of pitch within western-tonal music. All of Lerdahl and Jackendoff's examples seem to revolve around diatonicism or tonality and not pitch; in previous discussion it was demonstrated that Lerdahl and Jackendoff missed an application of their own Rule 3a (Register change), and this miss was likely due to their thinking in terms of diatonicism rather than pitch. The difficulty with diatonicism (and tonality) is how to cast it in theory or within an algorithm. That is, if issues concerned with pitch (e.g., Rule 3a: Register Change) were to be converted into issues concerned with diatonicism, how would one define changes. If the steps along the diatonic scale are perceived as equivalent (cf., Shepard, 1982), then how can the steps to the non-diatonic pitches be represented.

Furthermore, to use a diatonic scale for Rule 2b (Register Change), one would have to "know" which diatonic scale was being used. Hence, one would need to "know" the key that the piece was perceived in (or at least, likely to be perceived in). One would need to know that subjects actually used an internal representation of key when listening to the piece. The point of this discussion is not to attempt to clarify or extend Lerdahl and Jackendoff's (1983) discussion of the pitch dimension. Rather, it is to say that Lerdahl and Jackendoff seem to implicitly use a diatonic representation of melody rather than a purely pitch height representation of melody. However, the role of diatonicism in the use of the low-level rules is not clear. It will therefore be necessary to find some empirical method to monitor the role of diatonicism in the analysis of parsing, even if only to demonstrate that the simple logarithmic representation of pitch is sufficient. One must always remember that diatonicism and the logarithmic representation of pitch are monotonically related.

The main point is that in the current work it was felt that it would be prudent to consider the potential effect of tonality on the parsing of melody. This consideration was secondary to the main goals of the empirical assessment of boundaries, and the relationship between those empirical boundaries and theoretical predictions of those boundaries.

CHAPTER 3

Experiment 1: Introduction

This first experiment was designed to explore several aspects of boundary formation within a melody. This section explores the methodological options and justifies the choice of an appropriate methodology. That is, theoretical discussion of the parsing of music has been presented in Chapters 1 and 2: The goal herein is to decide which experimental task can best address such issues. First and foremost such a task must be able to empirically determine boundary locations in melodies. These empirically determined boundaries could then be compared with the theoretically predicted boundaries given the low level parsing rules of Lerdahl and Jackendoff (1983) for the determination of boundary locations (see Chapter 2, this work). In addition to these main analyses, there were several supplementary analyses designed to explore possible relationships between boundary formation, tonality and musical background, as well as several checks on the integrity of other analyses.

The focus of this introduction is on the rationale for the choice of experimental task. Hence, studies are cited to the extent that they bear on this problem. To address the main purpose, several techniques were considered. Ideally, one would like to attach a probe to those neurons that are responsible for the detection of boundaries. Although not yet possible, this ideal does serve to inform. Presumably, the closer to the ideal, the better. Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) and Event Related Potential (ERP) directly image the brain (in some sense). Of these, only ERP has the necessary temporal resolution for studies of boundary location in music. ERPs have been used in a number of studies related to musical stimuli (Besson and Macar, 1987; Klein, Coles & Donchin , 1984; Konovalov & Otmakhova, 1984 [EEGs] Osborne &Gale, 1992; Paulus, 1976 [EEGs]). However, although promising, the technique is yet not useful for the assessment of boundaries. There are simply too many unknowns: Do boundaries exist? Where are the boundaries if they exist? Do boundaries require the updating of working memory (associated with the P300) or require semantic processing (associated with the N400)? Until more behavioural assessments have determined some of the important parameters for the study of boundaries within music, the use of ERPs will have to be postponed.

Several behavioural tests have been used for the assessment of boundaries. Gregory (1978) and Sloboda and Gregory (1980) used a click detection task to demonstrate that subjects' responses tend to migrate to the regions where phrase boundaries should occur. Gregory (1978) presented subjects with six-note sequences monophonically at either 200 ms or 400 ms per note (with an equal duration pause). One group of subjects was instructed to treat the sequence as two three-note subsequences and a second group of subjects was instructed to treat the sequences as three two-note subsequences. That is, subjects imposed a grouping on the subsequences. Clicks, centred on the presentation of the second, third, fourth or fifth note or on the pauses between those notes, were presented to the other ear. Of relevance to boundary location, the perception of clicks tended to migrate to the boundary between subsequences regardless of the group to which subjects were assigned. The design suffered from the short length of the sequences (6 notes) and the fact that subjects were instructed to parse the sequences in a particular way (this may not be the "natural" parsing for these sequences). Also the interpretation of the click reaction time data requires that one know, in advance, the location of boundaries. Consider what the results would have been if Gregory had not instructed subjects to parse as either two- or three-note subsequences. Given that both parsings were easily accomplished, one might expect that, if not instructed, some subjects might "naturally" use two-note subsequences and some might "naturally" use three-note subsequences. The resulting mixed distribution would not support the theory.

Sloboda and Gregory (1980) used a similar format, but with longer sequences designed to assess the reality of the cadence as a marker of the phrase boundary. Two phrase stimuli were constructed that contained no phrase markers, structural phrase markers (tonal/harmonic considerations), acoustic phrase markers (lengthened notes) or both structural and acoustics phrase markers. The results were similar in that the clicks tended to migrate to the theoretically notated phrase boundary. Phrases marked by either structural or acoustic or both structural and acoustic markers affected performance, though it seemed that structural information was necessary for subjects to predict upcoming phrase endings. The design, however, suffered from the artificiality of having only one (very obvious) phrase boundary that was designed theoretically, and from the fact that one must know where this phrase boundary is before one can interpret click reaction time. It may have been that there were individual differences in reaction times for the neither condition (neither structural or acoustic markers) that were washed out in the averaging process. That is, each subject may have parsed the sequence at a different location, and these parsings may have affected reaction times in the predicted manner.

Such studies belong to the family of dual task experiments that explore aspects of music (cf., Bereft & Parfait, 1993; Scoffer, 1985). In such tasks, subjects listen to music while performing a second task, and changes in performance on the second task are assumed to be related to the cognitive load imposed by the primary task, listening to music. Although attractive, such dual task procedures suffer from several problems within the goals of this thesis. Firstly, assuming that the assumptions are all valid, then changes in performance on a secondary task are related to cognitive load. Cognitive load may imply the location of a boundary and the associated creation of a memory unit, but cognitive load may exist for a number of other reasons that do not entail boundary formation. That is, different points of a melody may be more difficult to process, requiring more resources, thereby affecting reaction times on the secondary task. As such, the task cannot be used directly to assess boundary location.

If one cannot use the task to locate boundaries, one might use the task to check the location of boundaries (as in Gregory, 1978), under the assumption that the placement of a boundary and the creation of an associated memory unit increases cognitive load. Unfortunately, this tact requires some a priori knowledge of boundary location in addition to the assumption that unit closure imposes a cognitive load. Such knowledge is available theoretically, but not empirically. The use of this task also requires the additional assumption that all other factors can be controlled (e.g., cognitive load caused by musical complexity is not associated with unit closure) well enough to represent only random sampling error. If one were to rely on the theoretical knowledge, and if the theory were correct, then the results would be clear. In the event that the theory were only partially correct, or that other factors could not be controlled, the results would be very ambiguous. Such dual tasks also suffer from a second more serious concern: Namely, they are a dual task. One can never be certain (in the absence of independent verification) that the second task is not interfering with the normal operation of the first task.

Many researchers have attempted to assess boundary location by reference to other behavioural measures that must be associated with boundaries. For example, Palmer and Krumhansl (1987a, 1987b) and Boltz (1989, 1991a) asked subjects to rate the sense of closure associated with phrase endings (or melody endings). As a check on other aspects of their stimuli and conclusions, Krumhansl and Jusczyk (1990) and Jusczyk and Krumhansl (1993) asked subjects to decide which of two versions was more natural sounding: the one with breaks imposed at (theoretically defined) phrase boundaries or the one with breaks imposed between phrase boundaries. While each task has some virtue, all suffer from the same general problem of needing to know where boundaries might be located a priori. Furthermore, it is questionable that such general impressions would be useful at the level of structures below the phrase.

Another approach has been demonstrated by Deliege (1987, Experiment 1)[25]. Basically, musicians and non-musicians were required to parse short (3-16 notes) instrumental or orchestral extracts from classical music (baroque, classical, romantic and early twentieth century). Subjects listened to the extract and then indicated their parsing within a line of undifferentiated dots. A similar study was conducted by Peretz (1989, Experiment 1) using extracts from twelve French folk tunes. Both studies were intended to test aspects of Lerdahl and Jackendoff's (1983) theory of music parsing. As noted previously, Deliege[26] examined Rule 2a (Slur/Rest), 2b (Attack-point), 3a (Register), 3b (Dynamics), 3c (Articulation) and 3d (Length): She also added Rule 3e (Timbre) and split

---

[25] These studies were presented previously.

[26] Deliege (1987) relabelled the rules, but the correspondence is one to one with the original rules.

Rule 3a into two components retaining Rule 3a for pitch changes less than an octave while calling pitch changes in excess of an octave Rule 3f (3a and 3f were later recombined). Peretz looked at Group Preference Rules 3a (Register), 3d (Length) and 6 (Parallelism). The procedure was successful: In both studies musicians and non-musicians reliably parsed music into components that could be meaningfully related to intuitions of phrase boundaries and the Group Preference Rules of Lerdahl and Jackendoff (1983; see Chapter 2 of the current work, for more detailed analyses of the results). Both Deliege and Peretz concluded that training was not necessary for the invocation of the rules, but it did sharpen the ability to use the rules. Both Deliege (Experiment 2) and Peretz (Experiments 2 and 3) conducted additional experiments to test aspects of boundary formation.

Although useful, this procedure is limited to short sequences of notes (cf., Deliege, 1987, p. 335): A subject would be likely to lose serial position in a longer work. To extend the basic paradigm to longer sequences, one might try to introduce place markers into the line of dots, but this, in turn, might induce some parsing biases. The procedure also suffers from another more subtle processing bias: Subjects listen to the entire extract and then indicate their parsing. Such a task, then, is really a retrospective analysis of the parsing of the sequence. That is, the parsing indicated on the dots may represent only those boundaries (units) that could be recalled. To avoid memory problems, subjects in Deliege's study could hear the sequence as often as desired (the number of repetitions was recorded) whereas subjects in Peretz's study were permitted one repetition of the sequence, but these create or enhance yet another problem: The parsing likely represents global structure or top down processing (particularly for musicians) because the units would be delineated after hearing the entire sequence (cf., Peretz, 1989a). Finally, there is yet another subtle bias in both studies. In Deliege's study (Experiment 1), and in Peretz's study, sequences were selected to contain phrase units delineated on the basis of the rules of Lerdahl and Jackendoff (1983). In Deliege's study (Experiment 2), sequences were designed to pit the various rules against each other. As such, verification of the existence of boundaries and verification of the theory was likely because the design and stimulus selection constrained and encouraged subjects to use boundaries and to place boundaries at particular locations. That is, designs that select stimuli to meet certain criteria may have poor external validity (see Cooper & Richardson, 1986, and Loftus, 1996). In fairness, neither experiment was attempting to demonstrate the reality of boundaries or the utility of the rules in a global sense, but rather, each was trying to demonstrate that the rules could be used: The distinction is between what "could be" a basis for boundary formation and what "would be" the basis for boundary formation. However, the present experiment, and more generally, this thesis, was aims at the "would be" aspect of boundary use and boundary location.

One can avoid the problem of external validity by selecting musical material at random from a defined corpus (i.e., western-tonal music). One could use longer sequences by asking listeners to indicate the parsing of music on a score (or a simplified representation of the score). Although this alterative method does avoid some problems, it also creates others. Aside from the obvious limitation to musically trained or experienced individuals, the task suffers from a more subtle concern. Having the score in front of the "listener" may lead to parsing on the basis of the visual spacial pattern rather than the auditory temporal pattern (cf. Cook, 1989, p. 119). While it is arguable that for a highly trained musician, "seeing the music" is "hearing the music", this is not an established fact and, even if true, it is still a concern that seeing the music may help to induce a particular cognitive representation, or to select a particular cognitive representation from among alternatives.

One could possibly avoid the retrospective problem and use longer sequences by having listeners indicate their parsings while listening to music (the ideal of on-line processes; cf., Peretz, 1989a). In fact, this was the task chosen for the present thesis, though the rationale was actually borrowed from studies in social psychology. Newtson (1973, 1976; Hanson, & Hirst, 1989; Massad, Hubbard & Newtson, 1979; Newtson & Engquist, 1976; Newtson, Engquist & Bois, 1977; Newtson, Hairfield, Bloomingdale & Cutino, 1987) attempted to study the parsing of the stream of behaviour by having subjects indicate, by the use of a button press, when one meaningful event ended and a different one began. Subjects observed individuals (on film) engaging in various activities (without dialogue) and were simply instructed to break the stream into its meaningful units. In his first experiment, Newtson (1973) asked two groups of subjects to parse the same film of a person building a molecular model. The only difference between the film presentations for the two groups was that in the middle of one film, a short sequence of random action was inserted (the actor took off one shoe, put it on the table, and rolled up the other pant leg). Before the segment, both groups parsed in a similar manner. After the segment the group having seen the insert used many more units to parse the behaviour ("parsing behaviour" is Newtson's term). In Newtson's view, it is the parsing of the behaviour stream that regulates the amount of information available. Hence, the observers in the "strange insert group" tried to increase the amount of information absorbed (in order to make sense of the behaviour) by the use of smaller units. After viewing the film, subjects in both groups rated the personality of the actor. Subjects in the strange insert case were more diverse in their impressions of the actor, which would be predicted if these subjects had absorbed and processed different pieces of information (relative to each other) and generally more information (relative to the no-insert group). In the second experiment, two groups watched a film of a person filling in a questionnaire, followed by the same personality assessment of the actor. One group of subjects was asked to create smallest meaningful units of behaviour and the second was asked to create the largest meaningful units of behaviour. Small unit subjects produced more boundaries, small unit subjects were more diverse in their personality assessments of the actor and small unit subjects were more confident about their assessments (which is consistent with the absorption of more information). However, both groups were internally consistent in the location of their boundaries. Newtson also observed that the small units created by the small unit group seemed to be subsets of the larger units created by the large unit group (however, see MacCallum, 1986, for a critique), implying a hierarchically nested perception of behaviour.

For present purposes, it seems that the on-line parsing aspect of the task might be useful. The results were internally consistent and seemed reasonable from a theoretical perspective. In addition, the task was simple for subjects to perform, requiring no special skills.

Newtson and Engquist (1976) investigated the reliability of the technique by having subjects parse the same film (an actor exploring and randomly rearranging objects in a cluttered room) in a five-week test-retest procedure with a small unit group and a large unit group. Test-retest reliability was high for both groups. Newtson, Engquist and Bois (1977) repeated the five-week, test-retest analysis with eight films and three groups of subjects, one coding at the small unit level, a second coding at the natural unit level and a third coding at the large unit level. The results replicated the previous results and provided further support for the notion that the behaviour stream was parsed in a hierarchical fashion. Since boundary locations seemed to be stable, Newtson and Engquist next determined that material at the boundary was more important by presenting subjects with films in which material had been deleted at either the boundary (Newtson & Engquist used the term "breakpoint") or between the boundaries. Subjects were asked to press a key whenever they detected a missing element in the sequence and they did so more often when the element was at a boundary. Furthermore, the duration of the missing element mattered when the deletion occurred at a boundary but not when the deletion occurred between boundaries. In the second experiment, subjects were presented with slides extracted from either the boundary positions or from between boundary positions. Subjects could accurately reconstruct the sequence when provided with the boundary slides (comparisons were made with respect to a control group that had seen the original film), but not when presented with the material from between boundaries. In the final experiment, subjects were presented with six short films, and then given a recognition test for slides taken from the films. There were two groups of subjects: The first group simply watched the film while the second group performed the usual parsing of the film at the natural level. In both groups, slides taken from boundary positions (defined by the second group) were better recognized than slides taken from between boundaries. There was no difference in performance between the two groups implying that the act of parsing was not the cause of better memory, or more generally, that the act of parsing did not interfere with the natural process of observation.

In summary, in a number of variations on the basic task, subjects were capable of parsing a visual sequence of events creating units that were "as small as possible", or "natural or normal", or "as large as possible". For investigations of boundary formation in melody music, the ability to obtain different on-line parsings that reflect the natural (phrase?) or smallest (motive?) could be useful. The act of parsing (key-pressing) did not interfere with the normal perception of events. Subjects in the Newtson and coworkers series of studies found the procedure simple to learn (hence can be used with musically untrained subjects) and subjects were capable of producing consistent parsing for sequences as short as six seconds and as long as seven minutes (a reasonable range for musical stimuli). In the film context, consistency implied within-subject reproducibility across repetitions and between-subject similarity (both of which one would expect with music). Newtson and coworkers also noted that the number of units was correlated with depth of understanding of the visual sequence by several measures. Further studies (e.g., Newtson, Hairfield, Bloomingdale & Cutino, 1987) implied that the boundaries corresponded to points of meaningful or distinctive change between states and not to points representing the attainment of a meaningful or distinctive state. Both of these notions are consistent with the idea of phrase boundaries in music. Lastly, results were consistent with the notion that parsing affected long-term memory (i.e., parsing reflected subsequent processing).

The Basic Experimental Design

This film-parsing task was adapted for music: One can view this adaptation as an extension of the work of Deliege (1987) and Peretz (1989) or as a new application of the work of Newtson and coworkers (1973, 1976). Basically, subjects listened to melody and then indicated with a button press when one unit ended and a new unit began. Subjects indicated the "boundaries" or "breakpoints" between units, thereby informing us of the parsing of music and the delineation of the components of music. The task has the advantage of being a direct measure of boundary location. Based on the work of Newtson and colleagues, one would predict that subjects should have no difficulty learning the task, that subjects should show high within-subject reproducibilities and high between-subject similarities. Furthermore, one would expect that the boundaries would correspond to points of distinctive change, which is, in fact, what the low level Group Preference Rules of Lerdahl and Jackendoff (1983) are designed to capture. The task has other advantages. The stimulus input aspect of the task is completely confined to the auditory domain, so parsing based on other modalities cannot interfere. The task is also conducted on-line and so avoids problem of retrospective parsings. Finally, the task does not require any expertise with music. As such, the task has face validity.

Issues of Task Validity

However, despite the face validity[27], there are some concerns. Does the task have content and construct validity? Content and construct validity are much more difficult to assess than face validity. These types of validity are, essentially, judgements (cf., Kerlinger, 1986, pp. 418-423) and as such, can only be made by individuals experienced in the field, ideally, experienced in both theory and experimental methods. These require an examination of the concepts using a bootstrapping process within the web or net of knowledge that is related to the concept (Helmstadter, 1964, pp.134-144, based on Cronbach & Meehl, 1955). The task does appear to have content and construct validity within the series of experiments developed by Newtson and coworkers.

Regardless of the validity in the domain of social psychology, in moving this task to the new domain of music cognition, it is necessary to reconsider validity. The task could be said to have criterion validity if the task results in boundaries that agree with those determined by other tests. Unfortunately, there are no other empirical tests (to my

---

[27] Properly, the ideas behind face, content, construct and criterion validity require a book-length exposition. However, to be brief, in this work face validity is uncritical appearance that the task measures the associated concepts. Content validity is the question of sufficiency: Does the task capture all or most of the concepts behind the task? Construct validity is an assessment of the theoretical considerations behind the task: Why does the task produces the results it does? Is task performance a result of the concepts defined? Criterion validity is concerned with whether or not the particular task produces results compatible with other tasks that address the same concepts.

knowledge, beyond those cited previously; see also Chapters 1 and 2) and, in part, the purpose of this experiment is the empirical verification of theoretical notions of boundary location. In other words, although a theory is being tested (only one of many), we cannot assume that this theory is true (in some sense), and therefore we cannot use criterion validity as a marker for construct validity. At best, all that can be said at this point is that if theory and the results of this experiment coincide, then the task and theory together have, at least, the beginnings of construct validity.

To attack the question from another direction, if subjects are asked to perform this task with music, then what are the likely determinants of performance? The task does have some elements of a dual task contained within it: Subjects must listen and indicate their parsings at the same time, which is similar to the problem of thinking and articulating that thinking simultaneously. However, unlike the other cited dual tasks, both of these are directed at the same stimulus and its processing. If we assume that subjects are capable and do attempt to follow instructions (as is implied by Newtson and coworkers) , then the question is really one of determining what other factors might influence key-presses (possibly, unconsciously) other than boundaries and the associated units of music. What might influence a subject to "feel" a boundary in the music where none exists? Subjects may respond to a musical interjection (i.e., the musical equivalent of a verbal "oh!", "wow!", "surprise!", "ouch", "err...", "eh...", or "ah..."), if such a device exists. In this case, one might not wish to call the key press a boundary because, as in verbal discourse, the material before and after the interjection forms a single unit. Possibly, one might wish to label the single key press as two boundaries so as to delineate the interjection from its surroundings. This example is something of a red herring. It is intended to demonstrate that such problems are not as serious as they might seem: The key press of the subject does delineate units in some sense, but the nature of the delineation is ambiguous. Detailed concerns of the actual parsing of units can wait until some data are collected (the process of construct validation is one of bootstrapping between theory and experimentation).

Krumhansl[28] (1996) has used a similar task that provides some evidence for the validity of the design. However, the evidence is indirect because her work was concerned with analysis on a much larger scale (phrases and sections). In a multiphase design, she asked subjects to parse a complex piece of Mozart's Piano Sonata in E$^{\flat}$ Major (K. 282, First Movement) to indicate the endings of sections (by button press), the amount of tension (by mouse motion) and the beginnings of new musical ideas (by button press). The boundaries between sections and new musical ideas that were indicated by subjects aligned with the empirical measures of tension, as well as, with many intuitions of theoretically appropriate constructs (see Lerdahl's [1996] analysis of the piece). However, the analysis conducted by Krumhansl was at a fairly global level which negated the necessity of fine temporal distinctions. She smoothed subject responses by averaging over a two-beat window. As such, the results and analysis might not apply when smaller units are requested. In addition, the relation between the responses and the analysis was largely qualitative in nature. Although the analysis was directed at the more global constructs of music, for the present work it is relevant to note that the responses of subjects were highly consistent for all measures, despite the complexity of the music (Lerdahl & Jackendoff might argue that this is a consequence of the confluence of multiple factors leading to the same interpretation). The consistency is more impressive given the wide range of training of the subjects.

For the question of validity, particularly with a low level analysis, the most serious concern is that subjects might indicate boundaries using principles like those described by Lerdahl and Jackendoff's (1983) Group Preference Rules, but those breaks mean nothing for the unitization of music. That is, the boundaries indicated by subjects could be an artifact of task demands (subjects are asked to press a key when...). More subtly, there may be reliable breaks in the music, but those breaks are not boundaries in the sense of creating word-like units of music. This is particularly important for parsing below the phrase, where the literature has been most ambiguous. To address this concern

[28] Note that the studies of this thesis were designed and conducted prior to the availability of Krumhansl (1996).

it is necessary to step back to examine what these boundaries really mean.

If boundaries exist, then boundaries delineate sequential units of music. If music is processed as a sequence of units, then subsequent processing should reflect those units. Hence, to demonstrate the reality of boundaries, a second task is necessary to insure that the key presses do mean boundaries and the creation of associated units. In the work of Newtson and coworkers, the reality of the unitization of the visual sequence was correlated with several other indices of social perception: reconstruction of the original narrative from segments (near boundary information was more important), recognition of deleted segments (near boundary deletions were more obvious) and recognition memory for segments (near boundary segments were more recognizable). Of these three, the third seems the most suitable for experimentation within a musical context. Some recognition tasks involving boundaries in music have been completed by Dowling (1973), Peretz (1989, Peretz & Babai, 1989) and Tan, Aiello and Bever (1981).

Peretz (1989; see previous citations), using extracts from 12 French folk tunes, tested aspects of Lerdahl and Jackendoff's (1983) Group Preference Rules by presenting both musically trained and untrained subjects with short melodies to parse. After the melodies had been parsed by one group of subjects (see the previous discussion of Peretz's study), a second group of subjects heard the melodies, followed by (after a two-second pause) a three-note probe. Subjects were required to decide if the probe had been in the sequence. Subjects were presented with 144 trials in which 96 contained true probes (had been in the melody) and 44 contained false probes (had not been in the melody). True probes represented one of four conditions: just before a boundary (a within-unit probe), on or crossing a boundary (between-unit probes) and just after a boundary (a within-unit probe). False probes consisted of novel sequences matched with the true probes on pitch range, key, contour and duration. Although the accuracy of musicians was slightly higher than that of non-musicians, no other differences between the two groups were noted. Also, accuracy was equivalent for true and false probes (only true probes were subsequently analysed). Probes that did not cross the boundary were better recognized than probes that did cross the boundary, but this effect interacted with the type of boundary presented (as defined by Lerdahl and Jackendoff's Group Preference Rules). For boundaries based on Group Preference Rule 3a (Register Change) and 3d (Length Change), within-unit probes were not recognized as well as between unit probes. For boundaries based on Group Preference Rule 6 (Parallelism) the reverse was true and very pronounced. Peretz also measured reaction times for the true probes (for the false probes, subjects could make a decision before the completion of the probe and hence such trials could not be counted), and obtained similar results but also evidence of a speed accuracy tradeoff, particularly for boundaries based on Rules 3a and 3d. Peretz (1989, Experiment 3) repeated the experiment with a third group of subjects, but reversed the order of the probe and sequence. In this case, boundaries based on Rules 3d and 6 both demonstrated more accurate responses with the within-unit probes. These results, at first glance, seem to contradict those of Newtson and Engquist (1976, Experiment 2): In their work, material near boundaries was better remembered than material far from boundaries. One explanation is that all of Peretz' probes were drawn from the area near boundaries, and as such, Peretz examined the subtle differences between variations of a single condition of Newtson and coworkers.

The results of Peretz parallel the results of Dowling (1973) in a similar task. Dowling presented subjects with sequences of notes that were designed to contain four phrases of five notes each, although the atonal construction of sequences constrained the interpretation of the sequences as music. Phrases were delineated by both rhythmic properties and by a return to middle C. After hearing the sequence, and a two second pause, subject heard a five-note test sequence (probe) and decided, using a four-point scale, whether or not that probe was contained within the preceding sequence. False probes were designed to change the overall contour, which implies that they were different from the original sequence. Two main groups of tests were constructed: probes that crossed phrase boundaries (corresponding to the between-unit probe of Peretz) and probes that did not cross phrase boundaries (corresponding to the within-unit probe of Peretz). The basic findings indicated that probes that crossed boundaries were not recognized as well as probes that did not cross boundaries, and the effect was magnified for slower presentations (slow: 3 tones/s, fast: 6 tones/s). There was also a positional effect with respect to the phrase number: later phrases were generally remembered better (a recency effect).

Tan et al. (1981) also produced similar results. Subjects were given a recognition test for a two-tone probe taken from one of three locations in a two phrase melody (16 to 27 notes): just before the phrase boundary, crossing the phrase boundary, and just after the phrase boundary. False probes did not match any portion of the sequence, though they were taken from the same key. The phrase boundary was defined on the basis of music-theoretic consideration. In 11 melodies, the phrase boundary was marked by a full cadence (from the dominant) to the tonic and in the remaining 13 melodies, the phrase boundary was marked by a semicadence (motion to the dominant). Performance at 0.44 was lowest for probes that crossed the boundary, 0.58 for probes before the boundary, and 0.62 for probes after the boundary. However, there were several interactions between type of cadence, training and probe. Musicians had a greater range of effects (in fact, the simple effects within the non-musicians were not significant), and it was only for musicians presented with the full cadence that the probes extracted after the boundary exceed the performance of probes extracted before the boundary.

Details of Design

Based on the previous discussion, it seemed like a two phase experiment would be capable of addressing the purposes of this work. The first phase assessed boundary locations using a key-press task (hereafter: *boundary location*). The second phase assessed the impact of such boundaries on subsequent processing (hereafter: *boundary efficacy*).

Subjects heard a melody and indicated the position of boundaries by a simple key press. Subjects were requested to make their units as small as is meaningfully possible. For each subject, this resulted in a boundary profile -- a point-to-point map of the locations of boundaries -- for the melody. Each individual boundary profile is a binary representation of the location of boundaries. To maximize external validity (and ecological validity), the stimuli were melodies from existing songs or extracts of such melodies, which is consistent with Peretz (1989), Deliege (1987, Experiment 1) and Krumhansl (1996). In contrast to Dowling (1973), Tan et al. (1981), and Deliege (1987, Experiment 2), stimuli were not created on the bases of theoretical notions of parsing. The point was to determine where boundaries exist in real music and to determine whether or not the theoretical notions of parsing (Lerdahl and Jackendoff, 1983) can predict the empirical data.

The second phase assessed boundary efficacy; subjects were tested with short extracts (Peretz, 1989, used the term probes, but the term extracts will be used herein to avoid confusion with the tonality task) from the melody and asked to indicate whether or not each extract was actually in the melody. As in studies by Dowling (1973), Peretz (1989) and Tan et al. (1981), subjects were tested for extracts that did not span boundaries and for extracts that did span boundaries. However, unlike all previous studies, the definition of a boundary was determined for each subject individually. That is, the definition of a boundary was unique for each subject. Hence, the distinction between what spanned or did not span a boundary depended on where that subject had actually placed a boundary. This change has two important consequences. Firstly, by virtue of this design, the extracts presented to subjects actually represented their perceived boundaries, and therefore their musical units. The various extract conditions did not reflect music-theoretic notions of boundary location (cf., Dowling, 1973; Tan et al., 1981). The various extract conditions did not reflect some average boundary based on a group of subjects (cf., Peretz, 1989: in her study, different groups of subjects supplied the parsings and responses to test sequences; the same is true for all the cited work in social psychology). Secondly, this made the actual execution of the design much more complex because subjects had to first provide the boundary location, then extracts had to be designed for that subject and finally, subjects were tested for their memory of those extracts.

Boundary Location

For the boundary location task, there are several considerations. Because the task is relatively new to music (albeit, an extension of existing work) it is critical that the responses of subjects be analysed in detail, both to insure the internal and external validity (essentially, reliability both within- and between-subjects) and to provide some measure of confidence in the constructs that underlie the design (i.e., the unitization of the melody). To this end, subjects were given fairly extensive practice so that the manual aspect of the task would not interfere with the processing of information. Subjects were also given repeated trials with the same stimulus (i.e., subjects parsed the same melody on 3 [or 6] occasions). This enabled an assessment of the within-subject reliability.

It is important to realize that high within-subject reliabilities are not required, nor necessarily expected. If subjects fail to demonstrate reliability across repetitions, repeated trials might provide a measure of the development of boundaries over repetition (i.e., the change in parsing that develops with familiarity). To highlight this point, the term "repetition consistency" will be used in place of the term "reliability", because there is no a priori requirement that subjects should be reliable (though previously cited work implied high reliabilities). The changes from repetition-to-repetition can be an error of reliability if we assume subjects are supposed to produce the same pattern of unitization. However, such change could also be the natural learning of the unitization of a melody. Because the online aspect of boundary location within the domain of music is relatively new, it is important that the repetition consistency be assessed by a number of measures to minimize possible errors or omissions of analysis and subsequent interpretation. This results in a rather large set of data and analyses, so large that only the luxury of a thesis permits a full exposition.

Once repetition consistency has been dealt with, it is possible to examine between-subject reliability. Again, a preferred term would be between-subject similarity since there is no reason to assume that all subjects should produce the same unitization pattern. Indeed, Lerdahl and Jackendoff's (1983) rules are called preference rules because they allow for a multitude of interpretations of the same piece, while constraining the total number of interpretations. Even though there may be differences, it is expected that groups of subjects will demonstrate similar boundary patterns: The shared language of music suggests a commonality of parsing (cf., Krumhansl, 1996; Peretz, 1989). As noted previously (see Chapter 2 of the present work), differences between groups may be due to issue of diatonicism or tonality. Differences between groups may also be related to training (amount and type) or musical tastes. Effects of training have been inconsistent within the literature (e.g., Krumhansl found none, Tan et al., 1981, found some). Again, the use of a number of measures of similarity insured that as many as possible of the nuances of the material had been captured.

Once within-subject consistency and between-subject similarity had been assessed, it was possible to relate the boundary profiles to the Group Preference rules of Lerdahl and Jackendoff (1983). To this end, it was important that the musical stimuli were not too complex, while retaining ecological validity (cf., Chapter 2, this work). This limitation was realized by selections that could be represented by a single, simple melody line (as presented in collections of musical works), with a well defined sense of key (e.g., modulations between keys were avoided, twelve-tone music was avoided). Of course, the definition of simple is somewhat subjective, but the guide for this was pieces arranged to provide instruction at the novice level (up to and including Royal Conservatory of Music, Grade 3 for piano: Hereafter RCM will be used for Royal Conservatory of Music).

The rules of Lerdahl and Jackendoff (1983) were quantized on a continuum (cf., Chapter 2, this work). The continuum for each rule was designed to reflect the strength or probability of boundary placement based on that rule. In the comparison of each theoretical rule with the empirical data, more power can be gained by comparing groups of like-minded subjects (those who produce the same boundary profiles) with the particular theoretical predictions. By adding several binary profiles from different subjects, one can get an estimate of the probability of a boundary at each location in the population. That is, any location at which a large number of subjects placed a boundary could be considered a high probability location, and any location at which only a few (or no) subjects placed a boundary would be a low probability location. In fact, if desired, each location can be treated as a binomial distribution with a given probability of success (a boundary) estimated by the empirically observed proportion (it is important for subsequent analyses, to remember that, in the limit of high n, and/or $p=q=0.50$, a binomial distribution becomes a normal distribution). These probabilities can be compared to the theoretical predictions using correlational techniques (which rely on normal distributions at each possible value for the variables involved).

It is admitted that this analysis is not optimal. For example, this analysis does not consider the conditional probability for a boundary at location k+1 given a boundary at the previous location, k. One might predict fewer boundaries just after the placement of a boundary, and in fact, this is represented by the Group Preference Rule 1 of Lerdahl and Jackendoff (1983). However, given the paucity of information to act as a guide, it is considered reasonable to ignore this for now. That is, since Rule 1 has not been experimentally verified, it should not be used as a limiting factor for any analyses at this point. A second point is that the boundaries placed by subjects represent boundaries based on any and all rules that might apply in music (regardless of the ability of Lerdahl and Jackendoff to capture those rules). In the comparison of the empirical boundary profiles and the theoretical boundary profiles, it was not expected that any one rule would capture all the empirically determined boundary positions. Conversely, it would be a remarkable simplification of music if one rule did capture all the empirical boundary locations (however, recall in Chapter 2 [this work] that it seemed that Rule 2b, Attack-Point, was the most prevalent for boundary locating). As was mentioned previously (cf., Chapter 2), it is important to watch for misses (an empirical boundary in the absence of a theoretical boundary) and to devote less attention to false alarms (a theoretical boundary in the absence of an empirical one).

One could continue to considering a multitude of alternative problems and analyses to deal with those problems, but such discussions and interpretations are best considered after the data have been seen. In this study, analyses were limited to the basic set necessary to clarify the relationships between the empirical and theoretical data. More complete analyses must be postponed until some of the more fundamental issues (e.g., Do any rules work at all? Which rules work? Is there a reduced set of rules that is both necessary and sufficient?) have been dealt with.

Methodological Issues: In addition to the previous concerns, there are a number of issues created by the methodological difficulties of an on-line assessment. Firstly, inevitably there is a delay between the actual decision and the associated manual response. This delay is likely exceed 100 ms (reaction times for simple detection task; Emmerich, Fantini, & Ellermeier [1984] found times in the range of 250 ms) but not likely to exceed 1000 ms (a reasonable maximum reaction time for a choice task). These values are within the range of the times of individual notes, so subjects may indicate boundaries with a delay of one or two notes. Hence, boundary profiles of individual subjects must be aligned to the most plausible intuitive and/or theoretical interpretation of boundary positions. Of course, as subjects gain familiarity with the piece, it is possible that their reaction times will decrease (anticipation of boundaries), and this will have to be considered as well. Given the vagueness of the timing of the responses of subjects, it was thought that there would be no point in timing boundary location with millisecond precision. Instead, boundary locations were tied to the currently sounding note event (a rest can be considered to be a note event) at the time of the key-press. It was then only necessary to examine the nearest neighbours of that note event in order to determine the "real" boundary location. That is, given expected subject delays and the expected durations of notes, the true boundary location should be no more than a note event or two (likely before) from the actual key-press. Krumhansl (1996) also noted such vagaries in response timing. Her solution was to smooth the responses of subject by averaging over a two-beat interval (she did not indicate the actual algorithm used to smooth). Such a procedure is unacceptable here because the desire is to analyse responses near the temporal resolution of individual notes (i.e., below the two-beat resolution). Krumhansl also noted that it was necessary, on occasion, to shift various averaged response profiles relative to each other (by up to 5 beats!) in order to discern the relationships between various measures at various times.

Boundary Efficacy

The boundary efficacy task was designed as a check on the reality of boundaries. Following Dowling (1973), Peretz (1989) and Tan, et al. (1981), subjects were presented with a short extract extracted from the melody. Subjects then decided if the extract was in the melody or not. At the most basic level, extracts were designed to either cross a boundary or not cross a boundary. However, this assessment was made on-line, based on the actual boundaries indicated by subjects. That is, once a subject had parsed a melody, extracts were created based on that parsing. Each extract was four notes in length. This length was chosen on the basis of Lerdahl and Jackendoff's Group Preference Rule 1 which states that units as small as one or two notes should be avoided, while units of three or four notes are acceptable. It is also a commonly cited length for the motive. It is also close to the values used by Peretz (three-note extract sequences) and Dowling (five-note test sequences). Each subject was presented with 16 extracts.

To address all possible delays for the subjective parsings, extracts were extracted such that the boundary (actually the key-press indicating a boundary) falls on the first, second, third or fourth note of the extract. As will be elucidated more fully later, these extracts should represent one or two units, depending on the meaning of the key-press with respect to boundaries. This, of course, adds to the complexity of the analysis, but it cannot be avoided. Additional extracts were extracted from immediately before the boundary (actually, immediately before the key-press) and from immediately after the boundary (actually, immediately after the key-press). An additional filler condition consisted of extracts extracted far from any boundaries.

Based on Dowling (1973), Peretz (1989), and Tan, et al. (1981), one would predict lower performance for those extracts that straddle a boundary (see the subsequent analyses for more details about specific predictions). Based on Newtson and Engquist (1976, Experiment 2), one would predict lower performance for extracts far from boundaries. For each melody, 16 extracts were created: one half were converted to *foils* (false extracts) by changing one note by one or two semitones (up or down). This decision was motivated by the desire to match foils to true extracts (hereafter, true extracts will be called *literal* extracts; the word true has too many connotations and uses) and the desire to see what types of change (position of change, size of change, tonality of change: see later analyses) most impaired performance. There was some trepidation that such a small change might not be detectable. For example, untrained subjects in Cuddy and Cohen (1976) performed at about 60% when detecting such a change in a single note of a major triad, though musically trained individuals reached 90% on average. In Peretz's study, errors ranged from about 8 to 36% for true extracts (i.e., literal extracts; Peretz only analysed literal extracts). In the study of Tan, et al., the performance of musicians ranged from 34 to 72% on literal extracts and from 27 to 46% on false extracts. For non-musicians the corresponding numbers were 45 to 65% and 43 to 45%. The size of the change were mitigated by the desire to challenge musically trained subjects (small changes), while being attainable by untrained subjects. Because extracts were generated on-line individually for each subject, the analysis was, by necessity, complicated.

It should be noted that Peretz (1989) advocated the use of a monitoring task (Experiment 3) in place of a recognition task (Experiment 2). That is, she observed better results when subjects were presented with the extract (called a probe) before the melody, and then asked to detect the extract when it occurred within the melody. Although it might be more useful, this limits the experiment to the presentation of only one probe per melody. Hence, much less data could be collected.

<u>Methodological Issues:</u> The limit of 16 extracts was chosen as a balance between the desire for many extracts and total testing time. The comparison of the extracts with the melody should reflect the effects of long term memory. The melody and associated extracts are too long to be retained verbatim in short term memory. However, given the mandatory similarity for the melody and its extracts, one can easily expect proactive and retroactive interference if too many melodies and/or extracts are used. That is, if too many extracts were used, the subject might confuse (compare) the current extract with other extracts (which are similar to the melody) and not to the melody per se. If the delay between the end of the melody and the presentation of the extract is too long, the subject might compare the extract to other melodies within the same experiment. In addition, one

might predict some serial position effects due to primacy when learning of the melody. These would manifest as primacy for extracts extracted from the beginning of the melody. To avoid such effects, the number of extracts was limited, there was no delay between the presentation of the melody and the presentation of the extracts, the delay between the individual extracts was fixed (i.e., a time-out for failure to respond) and test extracts were only presented after the last repetition of the melody (in the boundary location phase). Experimental evidence suggest such concerns are necessary. Dowling (1973), Peretz (1989), and Tan, et al. (1981) all showed substantial decrements in performance for delays of only two seconds on a sequences of only 14 (Peretz), 20 (Dowling) or 16-27 (Tan, et al.) notes. In addition, in these studies, subjects were only presented with a single test sequence on any particular trial. As a further control, the serial position of the point of extract extraction from the melody was treated as a covariate in analyses. In addition, the order of presentation of the extracts was also used as a covariate in analyses.

The Role of Tonality in the Parsing of Melody

In addition, because diatonicism and tonality are important, if not central, to music, an attempt was made to understand the role of tonality within the parsing and memory of music. However, as has been noted in a number of studies (Frankland & Cohen, 1990; Krumhansl & Shepard, 1979; see also Krumhansl, 1990) the representation of tonality may vary considerably between subjects. Hence, a secondary task was added to assess tonality. The secondary task is a variation on the probe-tone task of Krumhansl (see Frankland & Cohen, 1990). Basically, subjects were presented with a key-defining context (an ascending or descending scale in the major mode), followed by a probe tone (each of the thirteen chromatic tones). Subject rated the fit of the probe to the context. From this, the response profile for the thirteen probe tones was constructed. Subjects having an internalized representation of the role of various notes within a given key tend to produce a what has been called a triadic profile. In this profile, tonic triad tones (tonic, mediant, and dominant) are given the highest ratings, other diatonic tones (second, fourth, sixth and seventh) are given intermediate ratings, and non-diatonic tones are given the lowest ratings. The perfect fourth is often close to the mediant in rating and the seventh (the leading tone) is often close to the non-diatonic notes in rating. Subjects lacking this internal representation tend to produce what has been labelled as a recency or proximity profile, in which high ratings are given to those notes heard most recently.

It has also been observed that these types of profiles are associated with the amount of musical training. In this experiment, tonality profiles were used as the bases for group delineation (cf., Frankland & Cohen, 1990; Krumhansl & Keil, 1982; Krumhansl & Shepard, 1979). This allows subject selection to include individuals without formal training (or formal assessment of expertise by RCM examination) and subjects who may be deeply involved with music but not properly tapped by measures associated with training. That is, the use of tonality profiles as an objective assessment of musical knowledge (or a least the rudimentary beginning of such knowledge) helps to extend the study to all people (all consumers of music). The approach allows one to assess the musical knowledge of the average person, even the ataractic, not just the musical knowledge of the expert, but without ignoring issues associated with musical expertise. The relationship between the type of profile and training is fairly basic: The probe-tone task seems to be capable of separating those with training from those without, but it does not delineate well between levels of training. For the purposes of this experiment, the desired goal is the separation of those who have and use an internalized representation of diatonicism (or tonality) from those who may not and the link between diatonicism (or tonality) and boundary location.

Degree of tonality might be important in parsing because diatonicism defines the potential set of notes in the melody and tonality defines the relative importance of notes within a key (cf. Chapters 1 and 2 of the present work). Hence, for any note in a melody, there may be a relationship between the role of a note within a key and aspects of parsing. Individuals sensitive to the role of various notes might parse in a manner that is different from those who are not sensitive to such distinctions. To use a familiar example, consider the standard structure of the musical phrase as a motion away from the tonic, to the dominant (or mediant and subdominant) and then back to the tonic. An individual sensitive to tonal structure might parse on long notes that are tonally important (see

Deliege, 1987, p. 352-353) or on changes in harmony (see Deliege, 1987, p. 343). Musically trained subjects might ignore pitch changes of an octave, or large pitch changes between notes that are not tonally important. An individual not sensitive to tonality might parse on any long note or any large pitch change. Newtson and coworkers (1973, 1976; Newtson & Engquist, 1976; Newtson, Engquist & Bois, 1977) argued that boundaries represent points of change, not points of relative stability. What constitutes change and stability depends on the classification scheme that applies to the subcomponents of the stimulus. For example, when in the key of C-major, an individual sensitive to the tonal hierarchy, may not consider the diatonic non-triad tones, A_4 and F_4 different, while the tonic triad tone E_4 would be very different from the non-diatonic tone $D^{\#}_4$. The reverse may be true for someone who does not utilize the tonal hierarchy and instead basis assessments of similarity on pitch proximity. The role of tonality is also implicated with Lerdahl and Jackendoff's (1983) theory, but its impact is only considered at a much higher level of analysis (a level that is beyond the scope of this work).

As was stated, tonality profiles have been shown to be correlated with training (Frankland & Cohen, 1990). A multitude of studies in a multitude of domains have demonstrated that experts use larger chunks and have better retention for domain related material (cf., Haberlandt, 1994, or any text on cognitive psychology). Hence, in the boundary location task, one might predict that musically trained individual would use fewer boundaries (larger chunks). In the boundary efficacy task, one might predict that musically trained individuals would have higher retention overall. Because the correlation between training and the evidence of the tonal hierarchy is not perfect, measures of training and more generally, measures of involvement with music, were assessed separately. Several analyses examined the relationship between tonality profiles, training and other measures of involvement with music.

Notes on the Communication and Interpretation of Results

The following experiment may be conceived as four interrelated, but largely separable experiments. The first aspect explores empirical boundary location in melodies. The second aspect compares those locations with predictions of the theory of Lerdahl and

Jackendoff (1983). The third aspect determines whether or not those locations are meaningful indices of processing. The fourth aspect delineates subjects on the basis of their internal representation of tonality and relates this to musical training and other measures of involvement in music. All the different aspects are assumed, a priori, to be related, but the empirical data will drive the final conclusions. Because there is a large quantity of data and because the analyses are necessarily complex, details of data reduction, rationalization for the choice of analyses, analyses and initial interpretations of those analyses are presented together as a unit in the section titled Annotated Results (in place of the more usual dichotomy between Results and Discussion). The intent is to make it easier for the reader to follow. A general summary is presented in the Discussion.

A critically important point is that the overall design is a mixture of experimental and quasi-experimental variables. The determination of boundaries is quasi-experimental since the stimulus (i.e., the melody) was not designed to contain specific predictors of boundaries. That is, the stimulus contained many features of music, some of which might be related to boundary formation. Hence, all analyses and statistics pertaining to boundary position must be interpreted within the context of a quasi-experimental (correlational) design. This is particularly important for the interpretation of significance and effect sizes. Conversely, the analysis of repetitions of the same melody is an experimental design. That is, the only thing that changes from repetition to repetition is time (the stimulus is constant). Hence, all analyses and statistics related to repetitions (significance and effect sizes) must be interpreted in that context. The analysis of relationship between the rules of Lerdahl and Jackendoff (1983) and the empirically determined boundaries is best considered quasi-experimental, because Lerdahl and Jackendoff only considered aspects of music that they felt important to perception/cognition (i.e., they may have failed to control or consider some effects). The analysis of test sequences is also a quasi-experimental variable because, although test sequences were based on boundary locations, there are probably many uncontrolled (or unassessed) aspects of melodic complexity that will affect retention of test sequences (just as there are many aspects of music that affect boundary location that Lerdahl & Jackendoff may have missed). Finally, the analysis of the internal representation of tonality is an experimental variable *to the extent that the results are not overgeneralized too much* (i.e., subjects can be classified by their responses to probe tones preceded by a key-defining context). The generalizability of the results is a source of disagreement between Krumhansl and others, including Butler and coworkers (Brown, Butler & Jones, 1994; Butler, 1989, 1990; see the discussion in *Music Perception*, 1994). However, Krumhansl (1990) has provided an impressive amount of support to indicate that the task used herein can be generalized to the notion of an internal representation of tonality. Generally, for the analysis of experimental factors, high effect sizes should be expected. For the analysis of quasi-experimental factors, lower effect sizes can be expected because there may be many uncontrolled factors each of which accounts for some of the variance.

Before delving into the data in detail, it is useful to make a few general comments pertaining to style. Firstly, the presentation of raw data (means, standard deviations, medians and modes) is extensive. Other researchers should have access to the raw data so that they may make their own inferences, especially given the number and breadth of theories in music that could make use of such raw data (only citing analyses tends to make the data inaccessible to any interpretation other than those of the original design). All error values cited (i.e., "±") are standard deviations, not standard errors (standard errors reflect, in part, the sample size). Some, such as Lotfus (1996) feel that means and standard errors (effectively standard deviations) are more important than other types of analyses lamenting the fact that space constraints in published work often lead to the elimination of the presentation of raw data. In work lacking such constraints, such as this thesis, one should avail oneself of the rare luxury of a detailed presentation.

Analyses are presented in detail so that individual readers can make their own inferences of statistical, and more critically, theoretical, significance. The discussion of non-significant results is more thorough than is usual, but in a work like this, non-significance is important[29]. For example in a test of Lerdahl and Jackendoff's (1983) theory, the "non-significance" of Rule X is critical, particularly if Rule Y is "significant". This would imply that Rule Y is predictive and Rule X is not predictive. Speaking more generally, Lerdahl and Jackendoff have generated several rules. All, or at least some, should be "significantly" related to behaviour if the theory is valid. If none of the rules were to be "significant", then the theory should not be used for further research. "Significance" matters only within a background of "non-significance".

When considering the dismissal of a theory on the basis of non-significant results, one must insure that the lack of significance does not arise from a problem of design. That is, the issue is not significance and the reporting of significance: The real issue is external validity. Given the theory, methods and execution, are the non-significant results indicative of a lack of an effect (where one might reasonably be expected)? The issue of significance is related to the previous notion that some aspects of the design can be considered quasi-experimental (i.e., the analysis of boundary locations and the analysis of test sequences) while other aspects of the design can be considered experimental (i.e., analysis of repetitions). Reporting and interpretation of significance must reflect these differences. For example, as will be noted, many of the analyses I will present are "highly significant", but irrelevant because effect sizes are small and the variable under consideration is experimentally manipulated. Conversely, some effects are not significant,

---

[29] Often it is non-significance when significance was expected that is interesting (and it is not a power problem, cf., Cohen, 1992). One usually designs experiments with conditions that can be expected to be different. Many texts devote substantial space to means of creating significance (e.g., Hayes, 1995; Kerlinger, 1986). In fact, experiments are heavily biased to find significant results (see Cooper & Richardson, 1986). According to Loftus, the typical design "revolves around the testing of a null hypothesis that could not really be true to begin with" (1996, p. 162). With respect to hypothesis testing, the situation has gotten so bad that the APA has recently considered a ban on the publication of hypothesis testing (1996 APA Convention, cited in Shrout, 1997) to parallel such bans in other areas (e.g., American Journal of Public Health, cited in Shrout, 1997; see commentaries by Abelson, 1997; Estes, 1997; Harris, 1997; Hunter, 1997; Lofus, 1996; Scarr, 1997). The problem is not hypothesis testing per se; the problem is one of external/ecological (or construct) validity.

but interesting because effect sizes are high. Note that the important variables are quasi-experimental.

Method

Subjects

Sixty-one subjects (41 females and 20 males) were recruited from the university community, primarily from the Departments of Psychology and Music. Subjects had a mean age of 23.61±7.48 years (range: 16 - 52) with a skew of 2.28±0.31[30]. The clustering of ages in the low twenties is typical for a university based population. Of the males, one was ambidextrous, two were left handed and 17 were right handed: Of the females, three were left handed and 38 were right handed. Because the basis for the assignment of subjects to groups would be a cluster analysis, no rigid selection criteria were imposed, although a wide range of training was desired. Subjects either volunteered, participated for class credit or financial compensation for the one hour experiment.

Several measures of involvement with music were obtained using a self-report questionnaire (administered as a short structured interview at the end of the experimental session)[31]. For each instrument formally studied, subjects reported the date lessons commenced, date finished and RCM Grade (or equivalent) as well as the hours per week of lessons and the hours per week practising. For these same instruments, subjects also reported the hours per week playing for those periods when they were not having formal instructions. If the intensity of instruction, practising or playing varied, several notations were made for the same instrument. In addition to lessons studied, subjects reported comparable information for instruments that they had "picked up" without formal instruction (a common phenomenon for musicians): date commenced, date finished and hours per week of playing. The distinction between formal lessons and playing was intended to capture the distinction between getting feedback and assessment and the lack

---

[30] The error is the standard error of skew.

[31] This questionnaire has been refined through its use in a number of previous studies.

of feedback and assessment. This distinction was kept in mind during the gathering of information. With respect to the above, classroom instruction was considered training (since individuals could get individual attention), formal choirs and chorus were considered training, but public performances were treated as playing. For each subject, for each instrument studied ("studied" is used as an umbrella term for instructions plus practising), within each category, the variables shown in Table 3.1 were retained.

Table 3.1: Training, Practicing and Playing variables retained for analysis.

Training	TotalInst	total hours of instruction (hours)
	MaxCatInst	maximum total hours of instruction for any single category (hours)
	RecInst	the recency of instruction (years before experiment)
	MaxSinInst	highest intensity of instruction on a single instrument (hours/week)
Practicing	TotalPrac	total hours of practice (hours)
	MaxCatPrac	maximum total hours of practice for any single category (hours)
	MaxSinPrac	highest intensity of practice on a single instrument (hours per week)
Grade	Grade	highest grade achieved (on any single instrument)
Playing	TotalPlay	total hours of playing (hours)
	MaxCatPlay	maximum total hours of playing within any single category (hours)
	RecPlay	the recency of playing (years before experiment)
	MaxSinPlay	highest intensity of playing on a single instrument (hours/week)

For the variables concerned with playing, measures were taken exclusive of the periods in which instruction was received (in lessons or practicing). As such, playing time could only occur during periods when the subject was not taking lessons on that instrument. It would be difficult to delineate playing from practicing. Hence when taking instruction, all additional time spent on the instrument was considered practice. For each instrument played but never trained, similar measures were taken (see Table 3.1). When both training and playing are discussed, the terms "involved with" and "studied" will be used.

For instruments studied, one can see that it is total hours that is retained (TotalInst, TotalPrac and TotalPlay). This is designed to allow for the equating of subjects across different study schedules (e.g., 1 hour/week for 2 years versus 2 hours/week for 1 year). For comparison purposes, one can divide the total hours by n, where n is multiple of 52, in order to obtains years of studying at n hours per week. Maximum total hours within a single category indicates the degree of focus of a subject on a particular style of instrument (MaxCatInst, MaxCatPrac and MaxCatPlay). One would expect higher performance (and associated grades) for those who had focussed more time on a single instrument category (e.g., violin, and viola) because time would be available for (devoted to) abstract issues of theory rather than mechanical aspects of execution. Conversely, individuals who had spread their time over a lot of instrument categories (violins and piano) might be associated with lower maximum achievement because proportionally more time would be consumed in the acquisition and refinement of the mechanic skills associated with each instrument category. Of course, this tacitly assumes that the instruments within a category are similar enough for the transfer of mechanical skills: This is undoubtedly true for closely related instruments (e.g., violin and viola) but is less likely to be true for more distant relatives (violin and bass) and is likely untrue for some members of a category (e.g., both violin and fretless guitars are classified as fretless string: no one played both). Maximum intensity on a single instrument (MaxSinInst, MaxSinPrac and MaxSinPlay), regardless of category, was also retained for similar reasons similar to maximum intensity within a category.

Note that it is possible within this scheme for a subject to have studied but never played an instrument: This simply means that once instructions were finished, the subject never returned to the instrument. In addition to the previous variables, the number of instruments trained and the number of instruments studied were also retained.

To compress these data, instruments were clustered into categories, although it must be admitted, a priori, that the categorization of instruments is a difficult undertaking. For example, a broad range of instruments use strings to produce sounds, but the method by which the strings are activated can vary considerably (e.g., a hammer strike in a piano, strumming for a guitar, bowing for a violin). The categorization scheme used was based on the common notions of families of instruments (in this case, the classifications were borrowed largely from Lloyd, 1968), modified slightly to reflect not just the style of playing but also the associated learning curves. For example, pianos and accordions were classified together as keyboard (even though an accordion actually produces sounds, and sound quality, in a manner like the reed instruments), violins, viola, cello bass and fretless guitars were classified as fretless string. One might suppose that the learning of a fretless instrument differs from that of a fretted instrument. The complete categorization system is presented in Table 3.2. Note that the system includes two categories for voice: the first is for those who had private or nearly-private voice lessons, while the second is for those who had worked within a choir or chorus.

As shown in Table 3.3 and Figure 3.1, subjects reported studying (training and playing) an average 1.78±1.46[32] instruments (minimum: 0.00, maximum: 7.00), but subjects only reported training (lessons and practising) on an average of 1.53±1.39 instruments (minimum: 0.00, maximum: 7.00). The difference is due mainly to the number of subjects who taught themselves the guitar. Sixteen subjects did not train on any instrument, 20 trained on one, 17 trained on two, 6 trained on three, 1 trained on four, and 1 trained on five. When considering number of instruments studied (played and/or trained), 12 did not study any instrument (hence, 4 subjects had taught themselves one or more instruments without any formal instruction), 16 studied one, 17 studied two, 15 studied three, 9 studied three, 3 studied four, 1 studied five, 2 studied six and 2 studied 7.

---

[32] All errors are standard deviations, not standard errors.

Table 3.2: The categorical coding of instruments

Category	Instruments		
Other	any not covered below		
Voice	voice		
Voice 2	choir	chorus	
Keyboard	accordion	harmonium	piano
	calliope	harpsichord	spinet
	celesta	organ	generic "Keyboard"
	clavichord		
Woodwind	bassoon	flute	piccolo
	clarinet	fife	saxophone
	English horn	harmonica	recorder
	(alto oboe)	oboe	pan flute
Fretless String	cello	violin	fretless bass guitar
	double bass	viola	fretless guitar
	(contra bass)		
Fretted String	banjo	lute	ukelele (ukulele)
	bass guitar	mandolin	viol
	guitar	sitar	zither
Brass	bugle	French horn	tuba
	cornet	trumpet	shofar
	baritone horn	trombone	
Tuned Percussion	bells	glockenspiel	steel drum
	clarion	tubular bells	timpani
	dulcimer	marimba	xylophone
Untuned Percussion	bass drum	castanet	snare
	cymbals	drums	tom
Synthesizer	synthesizer		electronic music
Harp	harp	kithara	lyre

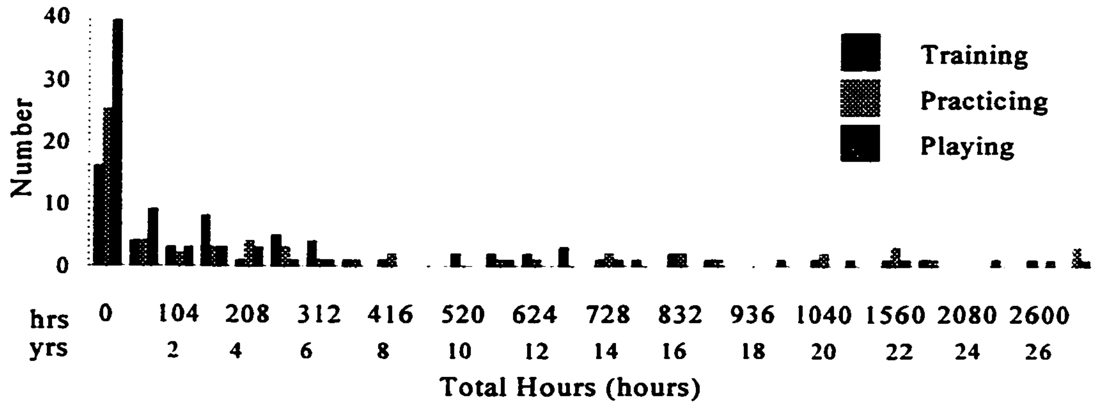
Notes: The list was not intended to be exhaustive; it was intended to capture the instruments commonly cited by subjects as well as some unusual examples.

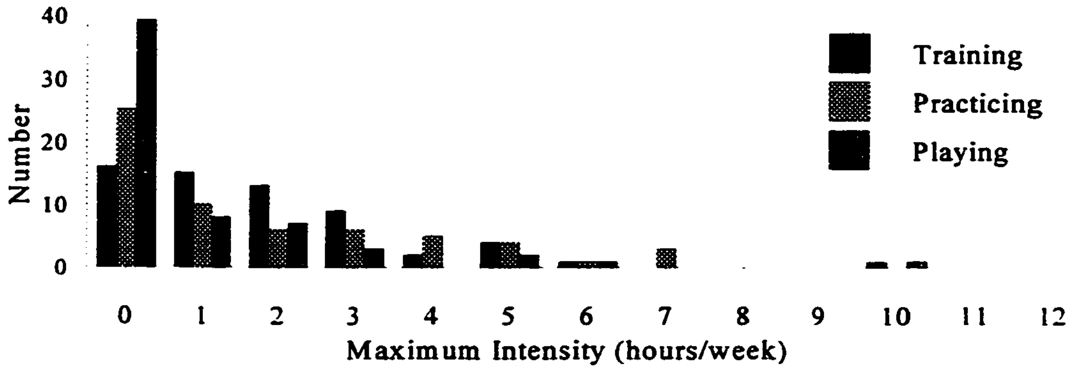
As can be seen in Table 3.3 and Figure 3.1, there is considerable variation in amount of training (and its associated degree of practising). For comparison purposes, the mean training is equivalent to 6.98±9.36 years at one hour per week, or 2.33±3.12 years at three hours per week. Also for comparison purposes, the mean practising is equivalent to 10.42±24.38 years at one hour per week, or 3.47±8.13 years at three hours per week. Note that the total times while training and the maximum intensities while training are based on all subjects (*n*=61), but the time since last training includes only those subjects who actually studied an instrument (*n*=45). This is because a total time of 0.0, or a maximum intensity of 0.0 are meaningful values for subject who never trained (i.e., no training or practice time), but the time since last training is not meaningful for a subject who did not train. The categories of keyboard (mainly piano) and woodwind (recorder, flute, clarinet and saxophone) were the most cited.

When examining the RCM grade that can be associated with training, it must be noted that not all who participate in music train in this manner. Most subjects (48) did not report any grade, one reported Grade 3, three reported Grade 4, one reported Grade 5, two reported Grade 6, one each had attained Grades 7 and 8, two had attained Grade 10 and one attained Grade 12 (ARCT). The average grade is possibly an underestimation because there were two subjects who were in the first year of music program at Dalhousie, who did not report any RCM grades[33].

---

[33] Admission to this program is based on interview and audition. Performance equivalent to about grade 8 is considered as the criterion.





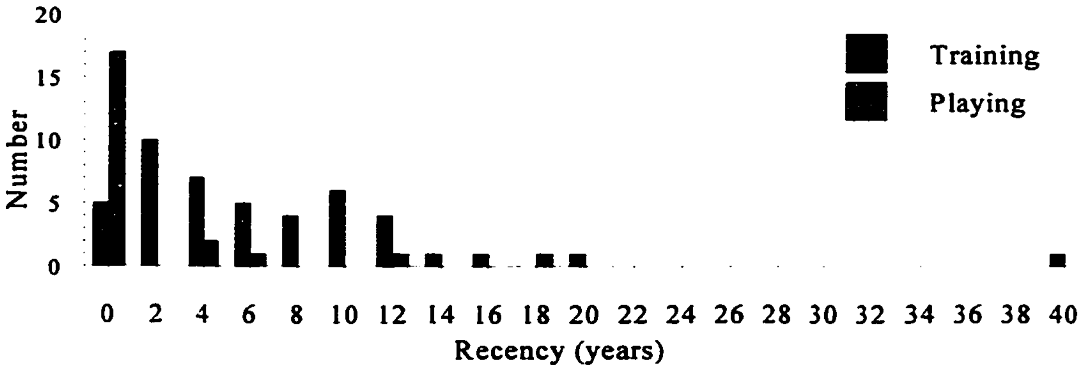


Figure 3.1: Hours of training, practicing and playing, maximum intensity of training, practicing and playing and recency of training (practicing) and playing. For hours of involvement, a second scale on the abscissa indicates the equivalent years of involvement at one hour per week.

Time devoted to playing, exclusive of training (or practising) is also shown in Table 3.3 and Figure 3.1. For comparison purposes, the mean total playing is equivalent to 3.32±11.71 years at one hour per week, or 1.11±3.90 years at three hours per week. Note again, that the total time playing, and maximum intensity of playing include all subjects (*n*=61), while the time since last playing includes only those subjects who actually played some instrument (*n*=49). Again, the maximum intensity is on one instrument only (i.e, not on all instruments in a category if multiple instruments were played simultaneously). In contrast to instruction, it was -- not surprisingly -- the category of fretted string (guitar) that was the most often cited.

Table 3.3: Training, Practicing and Playing over all subjects.

		Mean±SD	Min	Max	Units
Training	TotalInst	363.16±486.89	0.00	2569.54	hours
	MaxCatInst	302.95±431.13	0.00	2178.78	hours
	RecInst	6.16±6.78	0.00	39.00	years
	MaxSinInst	1.83±1.88	0.00	10.00	hours/ week
Practicing	TotalPrac	542.00±1267.92	0.00	7966.75	hours
	MaxCatPrac	435.35±1048.89	0.00	2178.78	hours
	MaxSinPrac	1.97±2.68	0.00	15.00	hours/ week
Grade	Grade	1.44±3.07	0.00	12.00	
Playing	TotalPlay	172.44±609.14	0.00	3754.44	hours
	MaxCatPlay	160.03±579.15	0.00	3460.78	hours
	RecPlay	1.80±4.25	0.00	16.75	years
	MaxSinPlay	0.82±1.81	0.00	10.00	hours/ week

It is no revelation that several measures are correlated with each other (see Table 3.4). Firstly, number of instruments trained was highly correlated with number of instruments played (r=0.91, p<0.01). In addition, total time training (TotalInst) was highly correlated with the maximum total time training within any single category (MaxCatInst; r=0.98, p<0.01). Total time practicing (TotalPrac) was highly correlated with maximum total time practicing within any single category (MaxCatPrac; r=0.99, p<0.01). Total time playing (TotalPlay) was highly correlated with maximum total time playing within a single category (MaxCatPlay; r=1.00, p<0.01). The implication is that most people focus on a single instrument (or instrument category). For this reason, the variables concerning maximum total time within a particular category (MaxCatInst, MaxCatPrac, MaxCatPlay) were not considered further (these variables are not included in Tables 3.4, 3.5 or 3.6).

Table 3.4 presents the correlations between training and practicing, along with the correlations of those measures with number of instruments studied or played. The important point is that the different measures are related but not redundant. When interpreting these simple correlations, it must be remembered that age is an uncontrolled factor.

Table 3.4: The correlations between the various attributes of training and playing, including the number of instruments trained and played.

	Training (Lessons)			Practice	
	TotalInst	MaxSinInst	RecInst	TotalPrac	MaxSinPrac
# Studied	0.481**	0.359**	-0.357*	0.347**	0.465**
# Played	0.430**	0.287*	-0.300*	0.394**	0.519**
TotalInst		0.761**	-0.157	0.410**	0.449**
MaxSinInst			0.113	0.125	0.342**
TotalPrac			-0.188		0.863**
MaxSinPrac			-0.132		

Notes: All correlations that involve Recency of Training and Recency of Playing are on 45 pairs: Other correlations are based on 61 pairs.
* p<0.05
** p<0.01

Table 3.5 presents the correlations between playing and other measures. Again, the important point is that the different measures are related but not identical. Note also that playing is not highly associated with amount of training although it is more associated with amount of practicing. It is arguable that both practicing and playing represent the unconstrained involvement with music, while training represents constrained involvement with music. That is, time devoted to playing or practicing may use whatever hours are available and desired, but time devoted to training is constrained by other considerations (e.g., finances, access to instruction). Generally, when examining these correlations, it must be remembered that many subjects who were currently involved with training would have no associated playing time. That is, all playing while taking lessons was considered practice. As such, playing might represent the continuation of practice after lessons had been terminated.

Table 3.5: The correlations between the various attributes of playing and other measures of music study.

	Playing		
	TotalPlay	MaxSinPlay	RecPlay
# Studied	-0.064	0.040	-0.106
# Played	0.146	0.290*	-0.329*
TotalInst	0.019	0.083	-0.168
MaxSinInst	-0.080	-0.050	0.132
RecInst	0.130	0.045	0.103
TotalPrac	0.251	0.368**	-0.355*
MaxSinPrac	0.206	0.300*	-0.303*
TotalPlay		0.898**	-0.394**
MaxSinPlay			-0.657**

Notes: All correlations that involve Recency of Training and Recency of Playing are on 45 pairs: Other correlations are based on 61 pairs.
* $p<0.05$
** $p<0.01$

These measures of training and playing were correlated against various demographic factors (Age, Sex and Handedness; see Table 3.6). The only correlations were between age and recency of training or playing.

Table 3.6: The correlations between the various attributes of music study and demographic variables.

	Age	Sex	Hand
Studied	-0.240	0.013	0.022
Played	-0.223	-0.077	-0.046
TotalInst	-0.115	0.144	0.006
MaxSinInst	-0.156	0.114	0.067
RecInst	0.791**	0.061	0.064
TotalPrac	-0.084	0.061	-0.067
MaxSinPrac	-0.101	0.037	0.038
TotalPlay	0.096	-0.237	0.025
MaxSinPlay	0.039	-0.261*	-0.081
RecPlay	-0.423**	0.291	0.145
Age		-0.010	0.159
Sex			0.162

Notes: All correlations that involve Recency of Training and Recency of Playing are on 45 pairs: Other correlations are based on 61 pairs.
* $p<0.05$
** $p<0.01$

A between-subjects ANOVA with handedness and sex as independent variables indicated a significant overall effect of sex on intensity of playing ($F(1,56)=4.79$, $p<0.03$; Females played within a single category more often than males), but no other main effects or interactions for any of the measures were observed.

Two supplementary measures of involvement with music were also obtained. The first was time spent listening to music "by choice" (Figure 3.2), which was 14.26±12.35 hours per week (minimum: 0.00, maximum 60.00). By choice was intended to capture those times in which the subject, essentially, had control over the music (and its desirability). The second was time spent listening to music "by default" (Figure 3.2) which was 5.98±5.69 hours per week (minimum: 0.0, maximum 24.00). By default refers to those occasions in which the subject had no control over the music, such as piped in music in the workplace, or the (incompatible) desires of roommates. The distinction between these two, although not perfect, was motivated by previous attempts to collect similar information: In those attempts, many subjects had strong opinions about the importance of the distinction. Interestingly, listening, either by choice or by default was not correlated with training, practicing or playing of instruments (see Table 3.7).

Table 3.7: The correlations between the listening and the various measures of music study as well as demographic variables.

	Listen$_1$	Listen$_2$
Studied	-0.139	0.130
Played	-0.100	0.142
TotalInst	0.019	0.014
MaxSinInst	-0.058	-0.114
RecInst	-0.218	-0.249
TotalPrac	-0.004	0.020
MaxSinPrac	-0.061	0.027
TotalPlay	0.116	0.045
MaxSinPlay	0.135	-0.068
RecPlay	0.017	0.023
Age	-0.081	0.019
Sex	0.061	0.051
Handedness	-0.373**	0.142
Listen 1		0.066

Notes: Listen$_1$ refers to listening by choice; Listen$_2$ refers to listening by default.
All correlations that involve Recency of Training and Recency of Playing are on 45 pairs: Other correlations are based on 61 pairs.
* $p<0.05$
** $p<0.01$

Note that those who like music enough to play an instrument would not have as many hours free to listen to music (playing, practicing, and training did not count in either category of listening). It seems that those who do not play an instrument do not fill the rest of their time with listening to music (implying that they do not like music generally). Listening by choice was correlated with handedness (r=-0.37, p<0.01) with left-handed subjects listening to more music than right-handed subjects (16.40 vs. 13.24 hours).

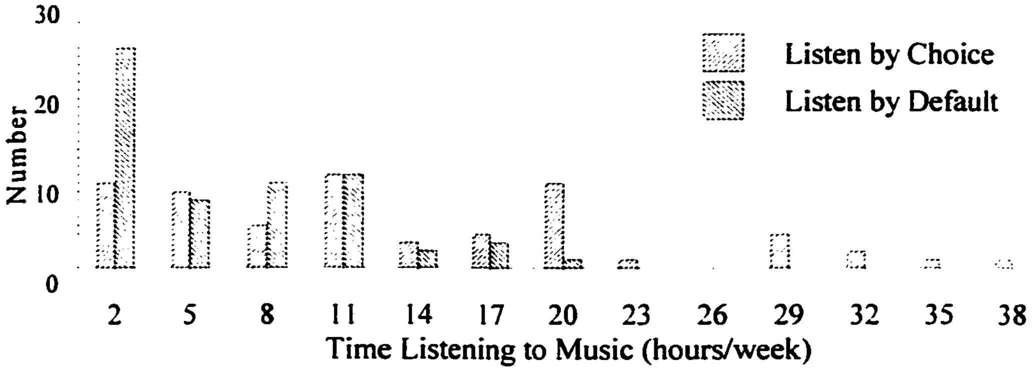


Figure 3.2: Hours per week spent listening to music by choice and by defalult

Procedure

The basic design was manifest in an eight stage experiment (see Figures 3.3 and 3.4). Stages 1, 3, 5 and 7 were designed to assess the internal representation of tonality of the subject using a modified probe tone task and as such were essentially the same, with only minor changes to the stimuli. Stage 1 was designated as practice for Stages 3, 5 and 7.

Stages 2, 4, 6 and 8 were designed to assess the boundary locations of the subjects within particular melodies. Stages 2, 4, 6 and 8 were the same, with changes to the stimuli and Stage 2 was designated as practice for Stages 4, 6 and 8.

Detailed instructions were presented to the subject at the beginning of each stage, at a pace determined by the subject, via the computer. The instructions for Stage 1 were repeated at the beginning of Stages 3, 5 and 7, but with progressively fewer details. The instructions for Stage 2 were repeated at the beginning of Stages 4, 6 and 8, again with progressively fewer details. Additional abbreviated instructions remained on screen during the each stage. The experimenter remained with the subject during the practice stages (Stages 1 and 2) to insure, by direct observation of the subject, that the instructions were understood and to provide clarification if needed. The instructions provided for each and every stage are included in Appendix: Instructions.

For all stages of the experiment (approximately 45 minutes, exclusive of debriefing), subjects were seated comfortably in front of an IBM style computer in an Industrial Acoustics single-walled sound-attenuating room.

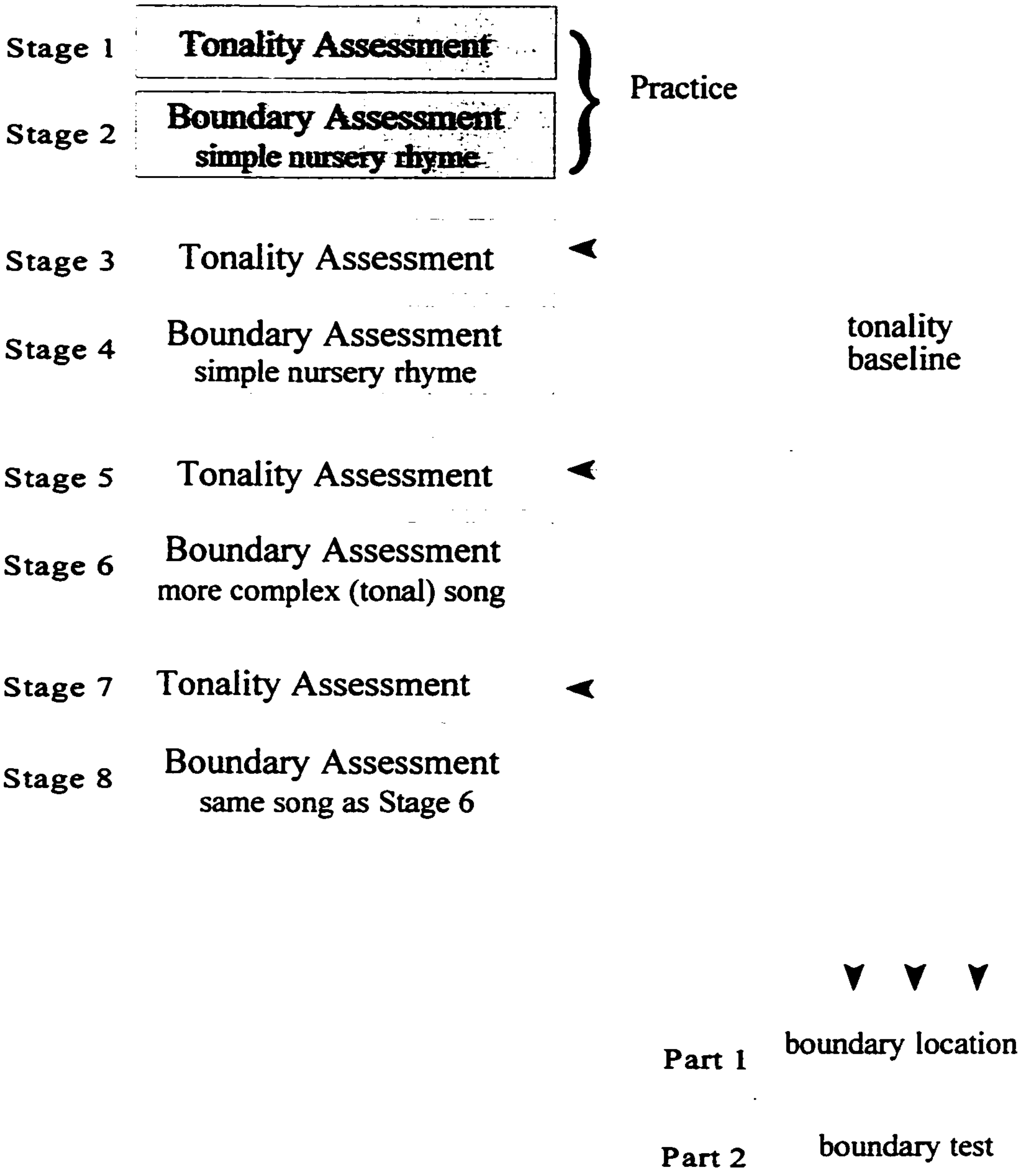


Figure 3.3: The overview of the eight-stage design. Note that stages devoted to boundary assessment are further subdivided into two parts: Boundary Location and Boundary Efficacy

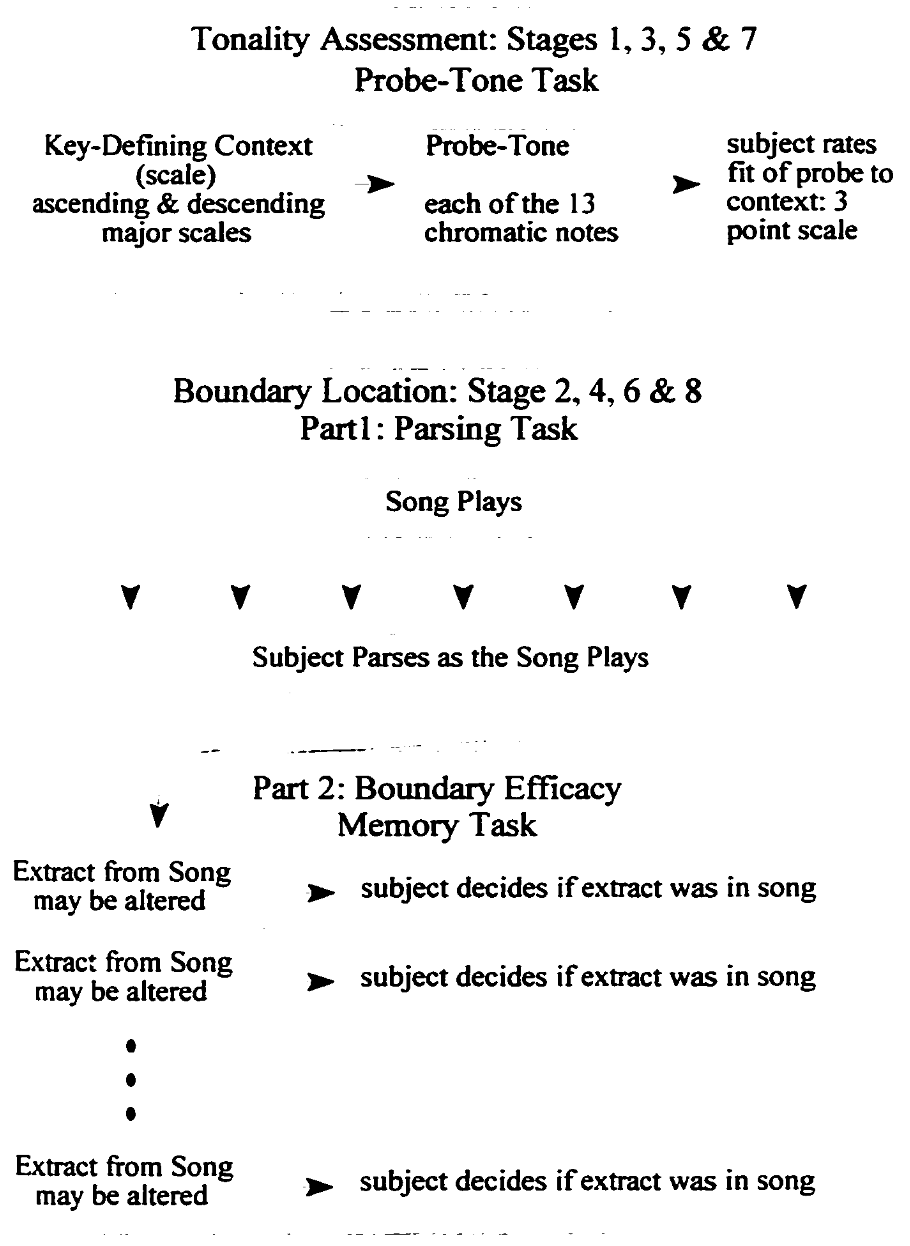


Figure 3.4: Basic design within each block. Note that the Tonality Assessment and the Boundary Assessment components are relatively independent, but Boundary Efficacy depends on Boundary Location. Hence, Tonality Assessment are considered separate stages, while Boundary Location and Boundary Efficacy are considered two parts of one stage.

Tonality Assessment

Stage 1 was designated as Tonality Assessment Practice while Stages 3, 5 and 7 were designated as Tonality Assessment 1, 2 and 3 respectively. In each of these, a modified probe tone task (Frankland & Cohen, 1990; Krumhansl, 1990) was used to determine the tonality profile of each subject. Subjects were presented with an ascending or descending major scale followed by a single probe tone drawn from the octave bounded by the scale. The scale defined a context and the subject rated the fit of the probe tone within that context on a three point scale (poor, indeterminant/uncertain, and good).

In a block of 26 trials, each of the 13 chromatic tones was presented once as a probe tone for both the ascending and descending contexts (i.e., each context -- probe-tone combination was presented only once). Within a given stage, all scale contexts were in the same key, but the presentation of the ascending and descending forms was a unique randomization for each subject, for each stage and for each block. The presentation of probe tones was also a unique randomization for each subject, for each stage, and for each block. Stage 1 (because it was Practice) had only one block of trials while Stages 3, 5, and 7 each had two blocks of trials. Key was varied across Stages such that the key of the Tonality Assessment Stage matched the key of the melody presented in the subsequent Boundary Assessment Stage.

Responses were indicated by the use of the 1, 2 or 3 of the keypad or by the use of the left, center and down arrow keys; for left-handed subjects, the use of the 1, 2 and 3 of the top row of the standard QWERTY keyboard was also provided.

Boundary Assessment

Stage 2 was designated Boundary Assessment Practice while Stages 4, 6, and 8 were designated as Boundary Assessment 1, 2 and 3 respectively. In each Stage, there were two parts. The first part was designed to assess the location of boundaries within a melody: Subjects simply listened to the melody and indicated the location of boundaries. The second part was designed to test the efficacy of those boundaries as they pertain to the formation of informational units: Subjects were given a recognition test for four-note subsequences taken from the melodies.

In the first part of each of Stages 2, 4, 6 and 8 (boundary location), a melody was presented. Subjects were instructed to press the space bar (or any key of their own choosing) any time they felt that one section of the melody had ended and a new section of the melody had begun: The term used to refer to this was "break". The analogy with the parsing of a line of speech into its componential words was stated explicitly (see Appendix 3.2 for the actual instructions). Subjects were informed that there would be three trials using the same melody, and that each subsequent repetition was intended to allow them to refine their answers. Additional verbal instructions were provided if the subject had questions or appeared confused.

At the end of the first presentation of the melody, subjects were asked to rate the familiarity of the melody on a 10-point scale, with 1 representing completely unfamiliar and 10 representing completely familiar. After rating the familiarity, subjects were asked to provide the name of the melody, if known (subjects were also informed verbally that a line of lyrics from the melody could be provided). Upon completion of this rating, subjects were provided with further clarifications if needed. Subjects then heard the melody and indicated boundaries for the second and third repetitions. Subjects initiated the presentation of each repetition of the melody at the time of their own choosing, but the timing within the melody was fixed for all subjects.

After the third repetition, subjects moved onto the memory task, again at the time of their own choosing. However, subjects were informed, at the beginning of each stage, that there would be a memory test to follow the final presentation of the melody, so that subjects would not delay a great deal before proceeding to the test part.

In the second part of each of Stages 2, 4, 6 and 8 (boundary test), subjects were presented with 16 four-note subsequences and asked to indicate whether or not each subsequence had been a part of the melody that they had just heard. Subjects were not informed of details of the construction of the subsequences, but they were informed that some subsequences were from the melody, and some were not. Subjects were explicitly instructed that the subsequence had to be an exact match if it was to be considered from the melody. Responses were indicated by the "y" and "n" keys, using the hand of their choice.

Upon conclusion of the eighth stage, subjects complete the questionnaire pertaining to musical background (see Appendix: Instructions), and were debriefed. The entire experiment lasted no more that 1 hour and generally no more than 55 minutes.

Apparatus and Stimuli

All tones within the entire experiment were created using the internal MIDI driver of a Creative Labs Sound Blaster 16, housed within an IBM AT (80286, 12 MHz) compatible computer[34]. All tones were created using the default instrument 0, mode 0, corresponding to the sound of an acoustic piano. The same computer provided instruction via a B/W monitor (Amber) and recorded responses. Programs for the presentation of stimuli and recording of responses were written in-house, using Borland's Turbo C/C++, Version 3.0, aided by the Creative Lab's Sound Blaster Developer's Kit, Version 2.0.

Tones were presented binaurally through a pair of Realistic LV 10 headphones connected directly to the audio output of the Sound Blaster 16 at a level considered comfortable by the subject. The monitor and keyboard were housed within the Industrial Acoustics single-walled sound-attenuating room, but the main computer was external to the sound-attenuating room in order to minimize noise.

Tonality Assessment

For Stage 1 (Tonality Assessment Practice), as shown in Figure 3.5, the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave C_4 to C_5 (523.25 to 1046.50 Hz). Context consisted of eight notes creating the ascending or descending C major scale, with each note having a duration of 400 ms (approximating a tempo of 150mm). Probe tones consisted of each of the 13 chromatic tones, again with a duration of 400 ms. In addition, there was a 400 ms rest (gap) between the final note of the context and the probe tone. All tones were presented at the same intensity.

---

[34] The same computer was used for the presentation of all stimuli in this thesis. A variety of computers were used for program development and data analyses.

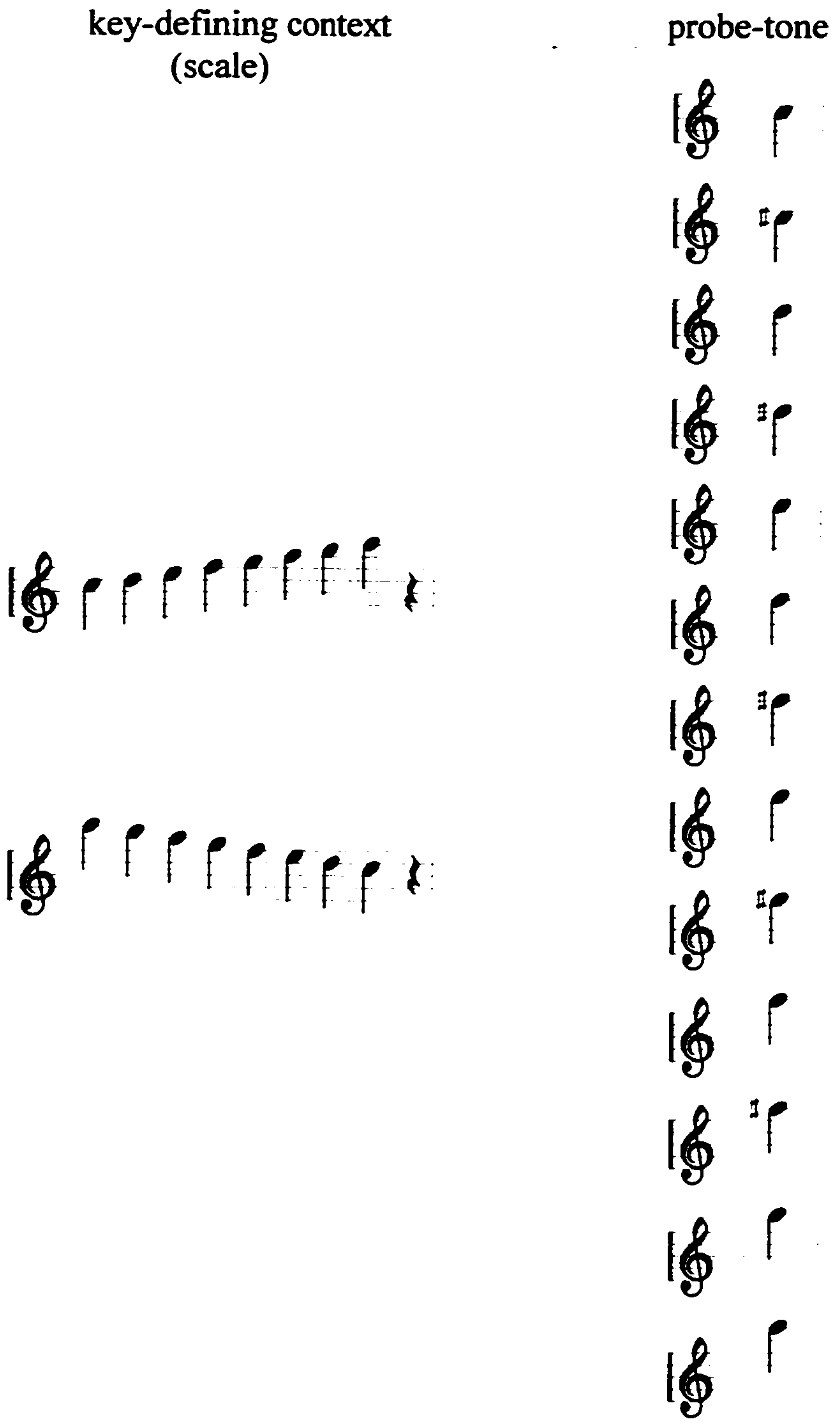


Figure 3.5: The stimuli used in Stage 1 (practice), for the asseassessment of tonality.

For Stage 3 (Tonality Assessment 1), the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave G_3 to G_4 (391.99 to 783.99 Hz; see Figure 3.6). It was intended that all other features remain as in Stage 1, but a minor error (a lower case g should have been an upper case G) resulted in the replacement of the G_4 with G_3 when G_4 was intended to be the probe tone. This error did not affect the context. The error was discovered after the 20th subject but since the error was not discernable to subjects (i.e., none commented on it during debriefing), and since the error would not affect the subsequent analysis, data collection continued unaltered so that all subjects experienced the same conditions.

For Stages 5 and 7 (Tonality Assessments 2 and 3), the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave C_3 to C_4 (261.63 to 523.25 Hz; see Figure 3.7). All other features remained the same as in Stage 1.

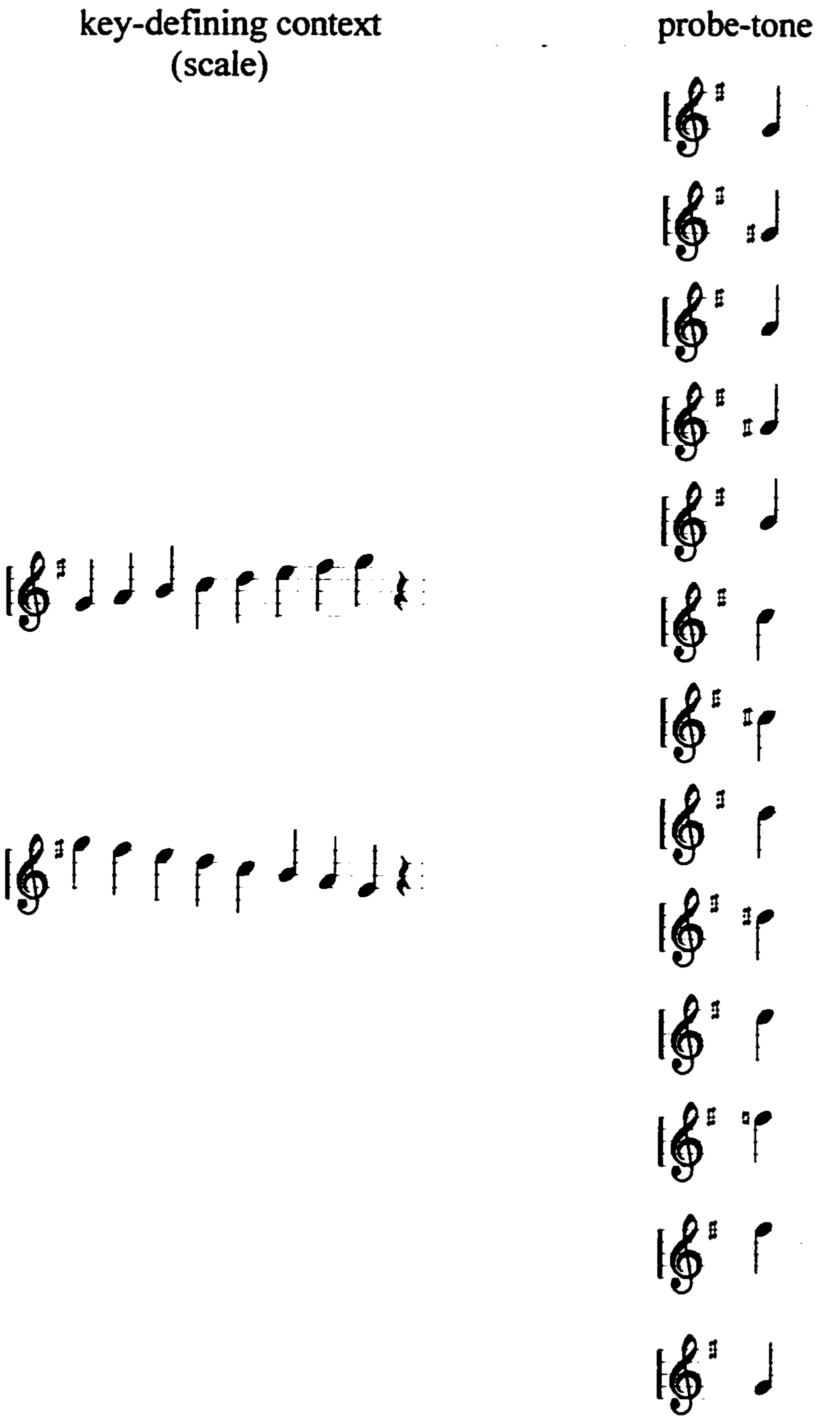


Figure 3.6: The stimuli used in Stage 3, for the assessment of tonality.

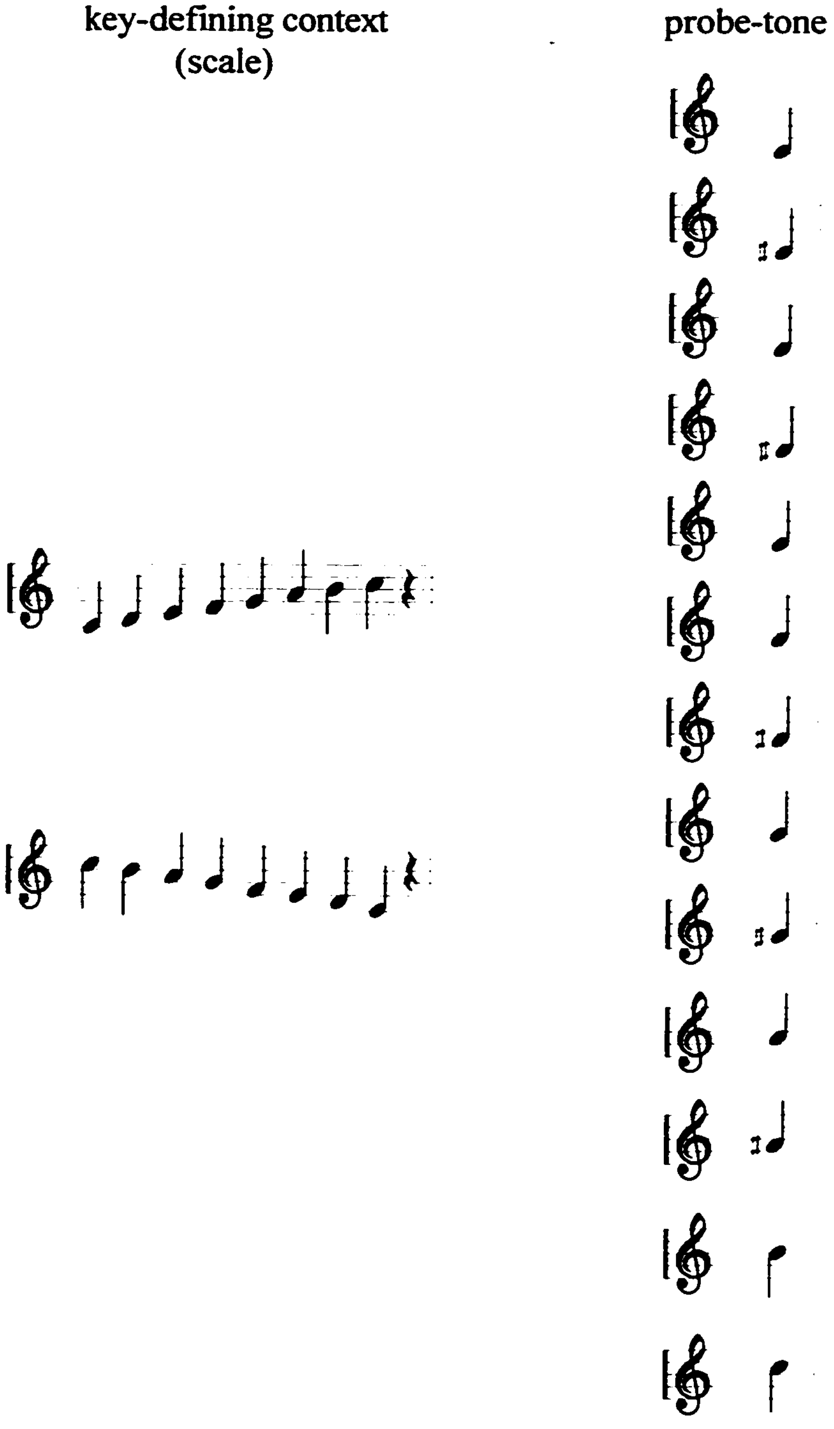


Figure 3.7: The stimuli used in Stages 5 and 7, for the assessment of tonality.

Boundary Assessment

In Stage 2 (Boundary Assessment Practice), the melody presented to subjects was "The Mulberry Bush", in the key of C (see Figure 3.8), with notes in the range G_4 to G_5 (783.99 to 1567.98 Hz), at a tempo of 135 mm, in 3/4 time. All notes were presented at the same intensity with precise computer controlled timing: No indications of beat or meter were manifest as departures from absolute timing or intensity of the notes. This melody had a moderate familiarity for subjects, achieving a mean rating (scaled from 1 to 10) of 6.01±3.13 (median: 6.75, mode: 10.00, minimum: 0.00, maximum: 10.00). Three subjects correctly labelled the tune, two by name and one by citing the first line of the melody. When using the Krumhansl key-finding algorithm (Frankland & Cohen, 1996; Krumhansl, 1990), the melody, as whole, had a key strength of r=0.83 (r^2=0.69) for the best fitting key (C-Major) and a key strength of r=0.64 (r^2=0.41) for the second best key (c-minor) yielding a q-factor of q=0.28 (see Frankland & Cohen, 1996 for a more extensive discussion). This melody could be said to be tonal.

During the assessment of boundaries, each response of a subject was tagged to the currently sounding note. That is, boundaries were associated with particular notes of the melody; boundaries were not associated with the gaps between notes (see Figure 3.9). The rationale for this lies in the uncertainty in the objective meaning of the subjective boundary. The subjective boundary could indicate the end of a functional group (i.e., the boundary was associated with the last note of a group) or the beginning of a functional group (i.e., the boundary was associated with the first note of a group) or even the actual gap between groups. Hence, there is some uncertainty in the actual location of the objective boundary: It could be after or before the associated note. In addition, there is additional uncertainty due to the reaction time of subjects.

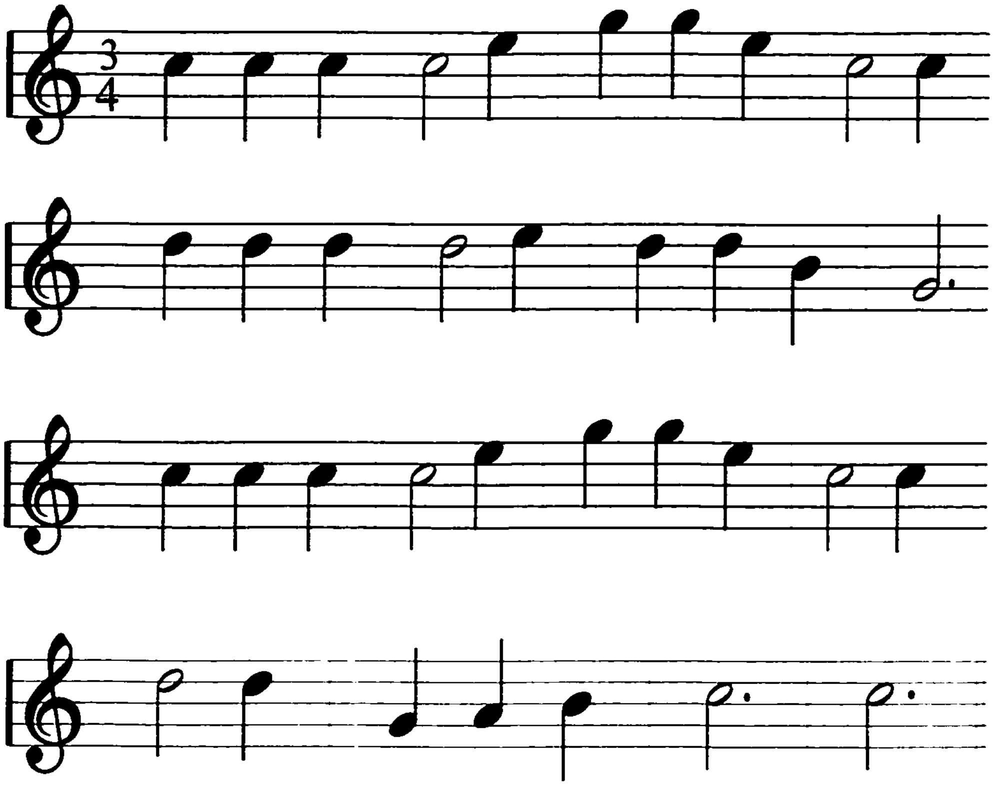



Figure 3.8: The song "The Mulberry Bush", adapted from *Nursery Songs at the Piano* (Bastien,1988), used in Stage 2, Part 1 as practice.

Test subsequences for Stage 2 were constructed online after the subject had indicated the placement of boundaries within the melody. As shown in Figure 3.9, each subsequence was created by taking an adjacent four-note span from the melody. That span consisted of four auditory events: Hence, tied notes would be considered a single auditory event, but slurred notes would not (slurs, it any, are not presented in the score because they were not included in the stimulus). A rest was also considered an auditory event (the absence of sound, with a discernable duration). The general intent was to create 16 four-note subsequences balanced across six conditions: One half of the subsequences would contain one and only one note which had been associated with a boundary (8 Boundary Test Sequences) and the other half would not contain any notes which had been associated with a boundary (8 No-Boundary Test Sequences). Furthermore, for the eight Boundary Test Sequences, one quarter (1/4) would have a boundary associated with the first note position (Condition 1), one quarter (1/4) would have a boundary associated with the second note position (Condition 2), one quarter (1/4) would have a boundary associated with the third note position (Condition 3), and the final quarter (1/4) would have a boundary associated with the fourth note position (Condition 4).

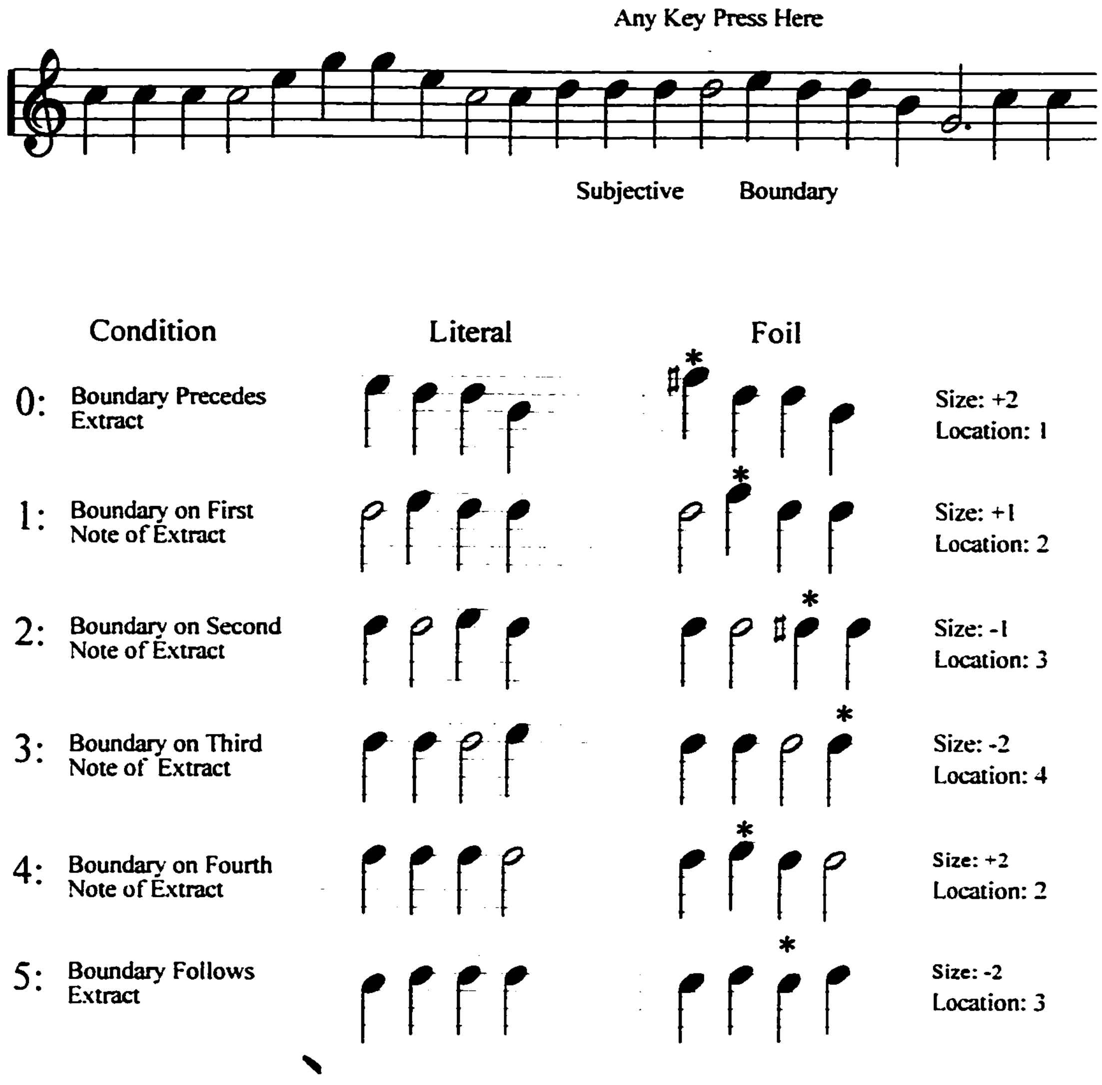


Figure 3.9: Possible extracts based on the boundary position indicated by the listener. Literal extracts were exact matches to the song. Foils changed one of the four notes of the extract by ±1 or ±2 semitones. In the experiment, foils consisted of random combinations of the size of change with the location of the change. The note changed is indicated by an asterick.

For the eight No-Boundary Test Sequences, one half (½) were designed to begin just after a note that had been associated with a boundary (Condition 0); that is, in the melody, the note just in front of the extracted subsequence would have been associated with a boundary. The other half (½) of the No-Boundary Test Sequences were designed to end just prior to a note that had been associated with a boundary (Condition 5); that is, in the melody, the note just after the extracted subsequence would have been associated with a boundary.

As stated, the objective was 16 four-note subsequences balanced across six conditions, but there were three main constraints on the achievement of this objective. The number and placement of boundaries by the subject limited the options: Since it was not desirable to use the same boundary in more than one condition, and 16 subsequences were needed, then ideally, 16 boundaries were needed. The type of melody affected the choice of subsequences: Many melodies contain multiple redundancies in the presentation of subsections (i.e., motives, phrases) and one would expect that boundaries would likely fall within the same position in each repetition. One would not want to present the same subsequence repeatedly even if those repetitions actually came from different parts of the melody. Finally, since the task was a memory recognition task, it was necessary that the decision about which sequences to present would be made quickly. However, juggling the responses of the subject with the stimulus and design of the experiment is difficult if one attempts the most general case, and this results in a very slow algorithm. In fact, such an algorithm was created for earlier versions of the presentation program and it was too slow by virtue of the number of checks needed to avoid any imaginable adverse condition. Conversely, since this experiment was intended to assess boundary formation in music, the implication is that the parameters controlling such formation are not well known, and hence, any attempt to optimize the selection of stimuli is premature. Details of the final solution can be obtained from the author, but essentially, the algorithm made one pass through the melody, and for each four-note subsection of the melody:

1) allocated each subsection of the melody to whichever condition was appropriate (with respect to boundaries in the region of the subsequence)

2) inserted each subsection into a table of subsequences for the appropriate condition (i.e., condition number)

3) once a particular condition had achieved the correct number of test sequences, further subsections that matched that condition were used to randomly replace previous subsequences in that condition

Two extra filler conditions were necessary: One extra condition (Condition -1) consisted of subsequences which did not contain any notes which had been associated with a boundary and were far from any boundaries (an extension of the No Boundary Test Sequences condition). The other extra condition (Condition -2) consisted of anything not already mentioned. In essence, this condition included subsequences that contained more than one note associated with a boundary (an extension of the Boundary Test Sequences condition). These extra conditions were used to maintain the balance between the Boundary Test and No-Boundary Test Sequence conditions if the numbers in the important conditions were insufficient. Such Filler Test Sequences were properly notated and treated as distinct in any subsequent analyses.

Finally, one half of the subsequences in each subcondition were converted into foils by increasing or decreasing the pitch of one note in the subsequence by one or two semitones. Subsequences that had been altered were designated as "foils". Subsequences that had not been altered were designated as "literals".

In Stage 4 (Boundary Assessment 1), the melody presented to subjects was "Three Blind Mice". The melody was shifted to the key of D so that for the presentation, notes ranged from D_4 to D_5 (587.33 to 1174.66 Hz). The tempo at presentation was 145 mm in 3/4 time. All notes were presented at the same intensity with precise computer controlled timing to minimize any indications of beat by changes in the relative timing or intensity of the notes.

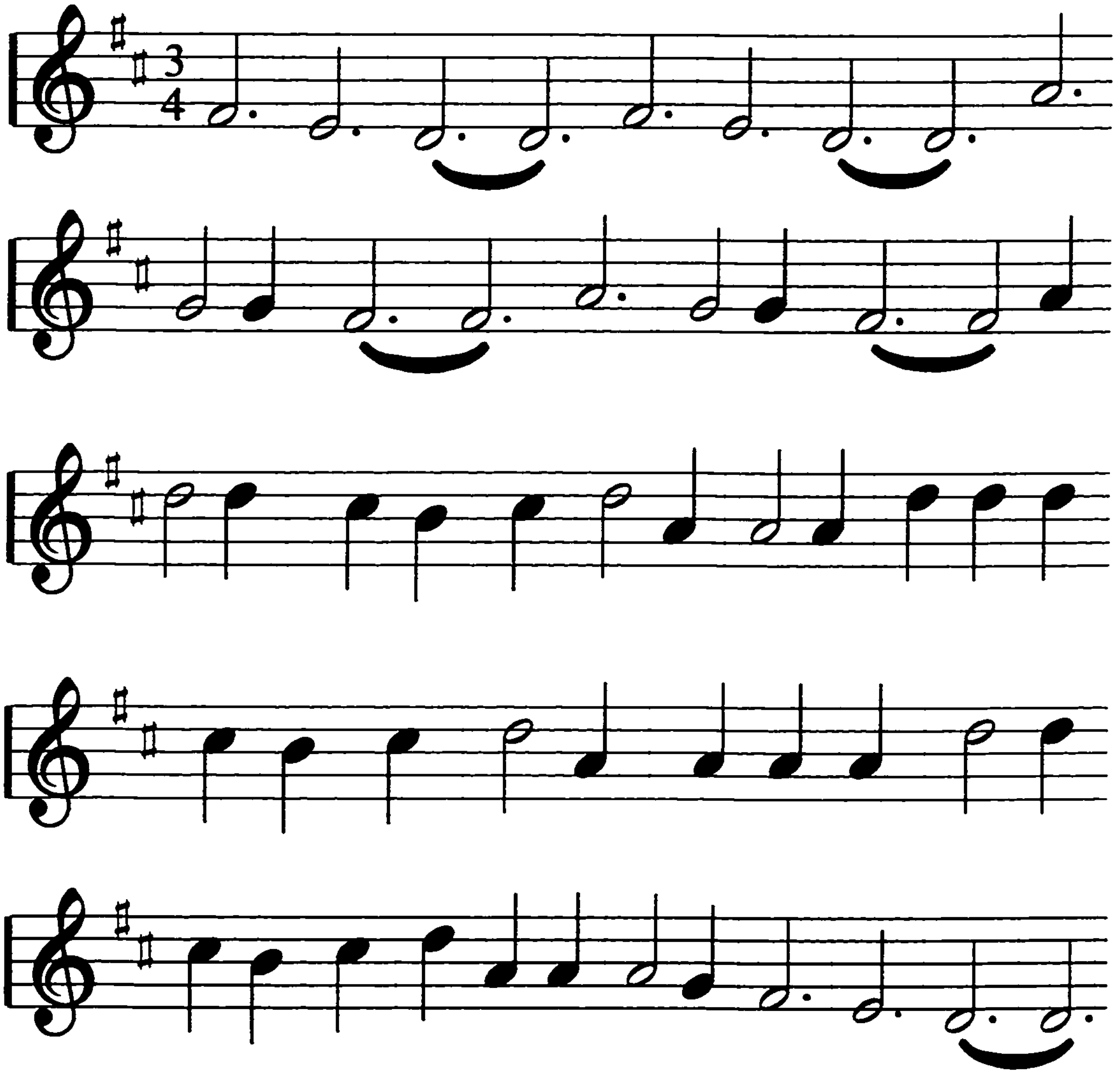



Figure 3.10: The song "Three Blind Mice", adapted from *Nursery Songs at the Piano* (Bastien,1988), used in Stage 4, Part 1.

This melody had a high familiarity for subjects, achieving a mean rating (scaled from 1 to 10) of 8.96±2.26 (median: 10.00, mode: 10.00, minimum: 1.25, maximum: 10.00). Forty-six subjects correctly identified the melody as "Three Blind Mice", two labelled it as "Hot Cross Buns", and one each labelled it as "Frere Jacques" and "Baa Baa Black Sheep". The Krumhansl key-finding algorithm (Frankland & Cohen, 1996; Krumhansl, 1990), found that for the melody as whole, the key strength was r=0.95 (r^2=0.90) for the best fitting key (G-Major) and a key strength of r=0.64 (r^2=0.41) for the second best key (b-minor) yielding a q-factor of q=0.49. This melody could be said to be unambiguous in its tonal centre. Test sequences were constructed in the same manner as Stage 2.

In Stages 6 and 8 (Boundary Assessments 2 and 3), the melody presented was "Softly Now the Light of Day" (the same extract that had been used by Boltz, 1989; see Figure 3.11). The extract was presented as notated (in the key of C) with notes ranging from B_3 to B_4 (246.94 to 493.88 Hz), at a tempo of 150 mm in 2/4 time. Again, all notes were presented at the same intensity with precise computer controlled timing so that there were no manifest indications of beat in the relative timing or intensity of the notes. In Stage 6, this melody was very unfamiliar to subjects, achieving a mean rating (scaled from 1 to 10) of 1.70±2.03 (median: 1.00, mode: 0.00, minimum: 0.00, maximum: 7.50). In Stage 8, the ratings of familiarity improved a little to 4.53±3.29 (median: 5.00, mode: 0.00, minimum: 0.00, maximum: 10.00). No one identified the tune, although one subject though it to be "that Russian guy", and another thought it to be Bach. For this melody, as a whole, the Krumhansl key-finding algorithm (Frankland & Cohen, 1996; Krumhansl, 1990), found that the key strength was r=0.71 (r^2=0.51) for the best fitting key (C-Major) and a key strength of r=0.66 (r^2=0.43) for the second best key (e-minor) yielding a q-factor of q=0.07. Although the melody could be said to be tonal, it is definitely more ambiguous that the previous two with respect to its tonal centre. The previous method of Stage 2 was used to create the test subsequences.

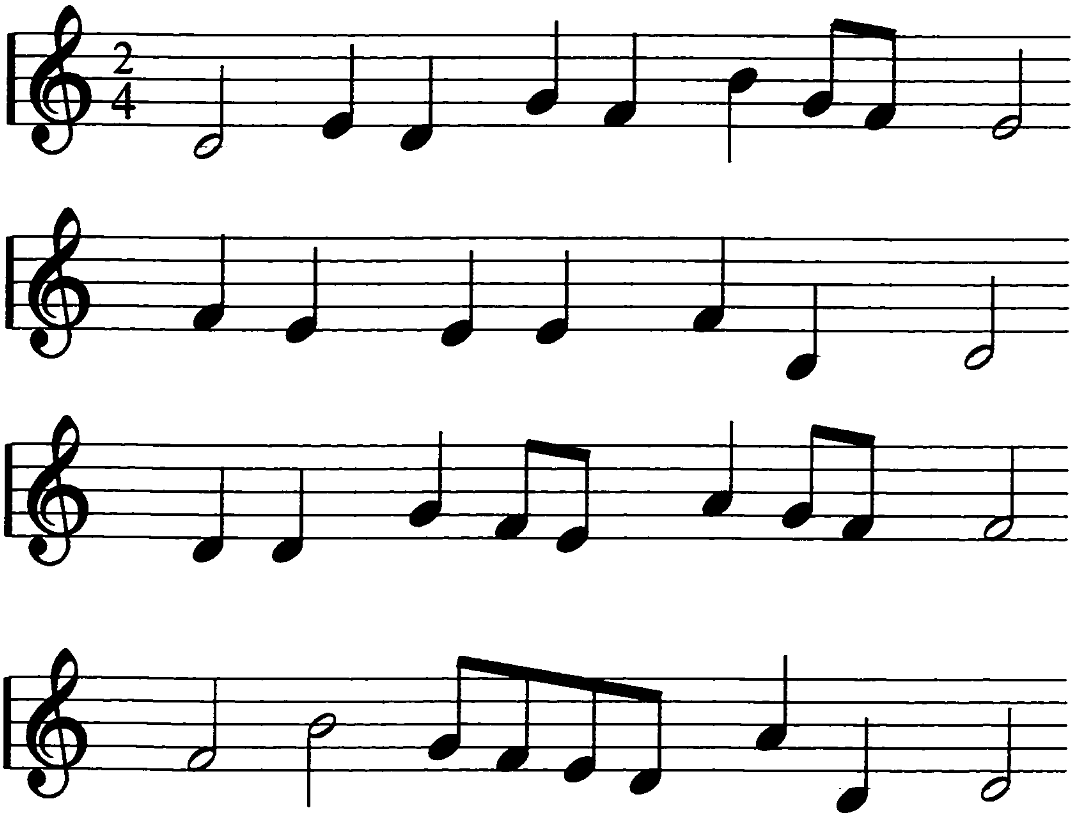



Figure 3.11: The song "Softly Now the Light of Day", by Phillip Doane, adapted from Boltz (1988), used in Stages 6 and 8, Part 1.

Annotated Results

Of the eight stages of this experiment, Stages 1, 3, 5, 7 were concerned with the assessment of tonality, while Stages 2, 4, 6 and 8 were concerned with the assessment of subjective boundaries in music. Although some relationship was expected between the two components, initially, each can be discussed in a relatively independent fashion. Because the present work is concerned with the parsing of music, the assessment of boundaries will be discussed first. This discussion includes the relationship between the empirical data and the models. The assessment of tonality will be discussed second. Finally any interactions between tonality and boundaries will be discussed.

The main focus of the boundary analysis was the assessment of the placement of boundaries, and the relationship between empirically determined boundary locations and the model of Lerdahl and Jackendoff (1983). Several other analysis were conducted to insure the integrity of any conclusion based on this analysis.

The main focus of the tonality analysis was the creation of groups of subjects homogeneous in sensitivity to tonality (as it is assessed by the probe-tone task). Again, a number of other analyses were performed as checks on the validity of this process.

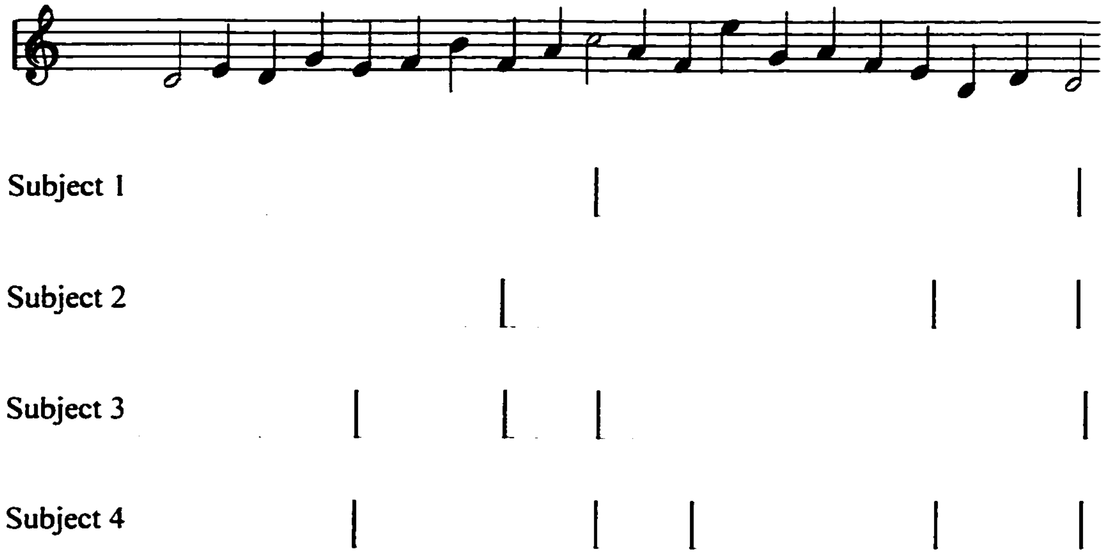
The tonality analysis was designed to create groups of subjects. The last analysis examined parsing within each tonality group, thereby assessing the relationship between the boundary formation and tonality (and diatonicism). This analysis was intended as an exploration of the possible relationships. Note that the theoretical model does not address such relationships at the level of analysis within this thesis.

General Considerations for Boundary Assessment

General Considerations for Boundary Location (Part 1 of each Stage)

In Part 1 of Stages 2, 4, 6 and 8, subjects heard a monophonic melody, repeated three times. Subjects indicated the location of boundaries within the melody, by the pressing of a key (usually the space bar, but it could be any key). The key press was associated with the currently sounding note, creating a boundary profile for each melody (see Figure 3.10 or 3.16). Because subjects heard three repetitions of each melody within one stage, subjects actually produced three boundary profiles for each melody. Each boundary profile consisted of a note-by-note map of the location of those subjective boundaries, with boundaries being coded as a 1 and the lack of a boundary being coded as a 0. Because Stage 2 was practice, it was not analysed: However, Stage 2 was used to test programs and statistical algorithms to be used in the subsequent parts.

For this work, the most important question concerns the locations of the boundaries. This question actually entails three subquestions: Where does each individual subject locate the boundaries?, are those locations stable within a single subject?, and are those locations stable across subjects? The first subquestion is answered by boundary profiles produced by each subject in response to each presentation of the melody. However, the answer to the first subquestion is interesting only if individual subjects show some consistency across repetitions of each melody. That is, although it would be interesting if every subject responded differently to every presentation of the same melody, unfortunately, such a state would render any subsequent analysis nearly impossible (either on a subject-by-subject basis, or across subjects). Hence, it is the second subquestion that is of primary interest. Do individual subjects produce the same boundary profile across repetitions of the same melody? The third subquestion rests to some degree on the first and second subquestions: If individual subjects show some consistency across repetitions, then do different subjects show the same, or similar, boundary profiles for the same melody? That is, the third subquestion address the relationship between profiles for different subjects. Several analyses were performed to answer these questions

To compare boundary profiles both within and between subjects, it is necessary to have some objective criteria for such a comparison. As shown in several examples in Figure 3.12, a boundary profile consists of a binary coding of the presence of boundaries on a note-by-note basis. Any boundary-profile comparison should consider this. Two subjects who placed boundaries on the same note event would be considered matched on that boundary. Two subjects who placed boundaries on different note events would be considered mismatched on those boundaries.

Boundary Matricies

First Subject		Second Subject 2 b	2 nb	3 b	3 nb	4 b	4 nb
1	b	1	1	2	0	2	0
	nb	2	16	2	16	3	15
2	b			2	1	2	1
	nb			2	15	3	14
3	b					3	1
	nb					2	14

Similarity Measure	1-2	1-3	1-4	2-3	2-4	3-4
Simple	0.85	0.90	0.85	0.85	0.80	0.85
Jaccard	0.25	0.50	0.40	0.40	0.33	0.50
Kulczynski 2	0.42	0.75	0.70	0.58	0.53	0.68
Sokal & Sneath 4	0.67	0.85	0.81	0.75	0.71	0.79
Phi	0.33	0.67	0.58	0.49	0.40	0.58

Figure 3.12: The effects of subjective boundary location on the degree of similarity for various boundary profiles. Note that the different subjects produce profiles with differing numbers of boundaries. b refers to boundary, nb refers to no-boundary. The example presents only 4 subjects, but in fact that analysis was extended to all 61 subjects.

Although the situation in which neither subject placed a boundary would technically constitute a match, not as much significance should be attached to such a match since it can be assumed that the co-occurrence of no boundaries would be a common event: That is, boundaries would be relatively rare events in a sequence of note events. Events that are linked to boundaries are therefore relatively more important than events that are not linked to boundaries. To weigh the co-occurrence of no boundaries equally with boundary matches and mismatches would artificially increase the degree of similarity. For example, as shown in Figure 3.12, consider a melody consisting of 20 note events. If one boundary profile has boundaries at note-event numbers 10 and 20, while a second boundary profile has boundaries at note-event numbers 8 and 20, one would intuitively like to conclude that the boundary profiles are different (see Figure 3.12). However, including the co-occurrence of no boundaries would mathematically demonstrate a high degree of similarity (18/20 note events are the same).

Even though the co-occurrence of boundaries is more important for the purposes of boundary-profile comparison, the co-occurrence of non-boundaries cannot be ignored. For example, in the previous example, ignoring the co-occurrence of no boundaries would also be an error because each note event represents the potential for a boundary, and as such, the fact that both boundary profiles do not have boundaries at 17 of the 20 locations is important information. Figure 3.12 provides a number of examples of this idea. To move beyond an intuitive sense of the similarity between two boundary profiles requires some mathematical assessment of similarity. However a proper assessment utilizing appropriate weights for the four conditions (e.g., boundary-match, two types of boundary mismatch, no-boundary-match) requires the kind of information that this study is trying to provide (i.e., the conditional probabilities for boundaries). That notwithstanding, there are several binary similarity (or proximity or dissimilarity) measures that can address this issue (see Table 3.8 & Figure 3.12). All such measures compare the number of matches of boundaries or non-boundaries to the mismatches of boundaries and non-boundaries. The methods differ in the degree of weight given to the non-boundary match conditions, as well as other details. Since the relative importance of the non-boundary match condition and the ratio of boundaries to non-boundaries were unknown (these are related issues), in this work, boundary profiles were compared using the simple similarity measure (Sm), the Jaccard similarity measure (Ja: the similarity ratio), the Kulczynski 2 similarity measure (K2), the Sokal and Sneath 4 (SS4) similarity measure and the phi similarity measure (phi: see Table 3.8; Howell, 1992; SPSS, 1988). The first two are intuitively obvious (differing in the amount of weight given to no-boundary match conditions) measures of boundary profile similarity. The second two measure provide non-directional conditional probabilities (differing in the amount of weight given to no-boundary match conditions) that average the predictability of boundary locations in one profile given the boundary locations of the other profile. The last is the binary form of the Pearson correlation statistic. Table 3.8 provides more extensive information for each of the statistics, including the formula, minimum and maximum, as well as, some indication of important (critical and chance) values for those measures. Critical values are defined as those similarity values that would indicate two boundary profiles that contain the same boundary locations 50% of the time. Chance values are defined as those similarity values that would be determined if boundary locations were determined solely by random occurrences. For those measures that include the no-boundary match conditions (i.e., the "d" of Table 3.8), three boundary to no-boundary ratios (i.e., the proportion of boundaries within the number of note events) were considered. Similarities exceeding these critical, or even these chance, values can be considered important. However, these critical and chance values should be interpreted judiciously because the proportion of note events containing boundaries is unknown, and because the computation of critical and chance values assumed equal numbers of boundaries per boundary-profile. Five different measures are used because this study is, in part, exploratory and as such, each measure provides a somewhat different assessment of similarity. Five measures are presented for the main analyses of Stage 4 and of Stages 6 & 8. Based on those results, subsequent analyses focussed on only the phi measure.

Table 3.8: Boundary profile similarity measures, their equations and their critical and chance values.

		Boundary Profile A	
		Boundary	No-Boundary
Boundary Profile B	Boundary	a	b
	No-Boundary	c	d

$$\text{simple} = \frac{a + d}{a + b + c + d} \qquad \text{Jaccard} = \frac{a}{a + b + c}$$

$$\text{Kulczynski 2} = \frac{a/(a+b) + a/(a+c)}{2} \qquad \text{phi} = \frac{ad-bc}{\sqrt{[(a+b)(a+c)(b+d)(c+d)]}}$$

$$\text{Sokal \& Sneath 4} = \frac{a/(a+b) + a/(a+c) + d/(b+d) + d/(c+d)}{4}$$

Measure	min–max	critical (and chance) values given π		
		0.875	0.667	0.250
Simple	0 – 1	0.875 (0.781)	0.833 (0.722)	0.750 (0.625)
Jaccard	0 – 1	0.333 (0.067)	0.333 (0.091)	0.333 (0.143)
Kulczynski 2	0 – 1	0.500 (0.125)	0.500 (0.167)	0.500 (0.250)
Sokal & Sneath 4	0 – 1	0.714 (0.500)	0.700 (0.500)	0.667 (0.500)
phi	-1 – 1	0.428 (0.000)	0.400 (0.000)	0.333 (0.000)

Notes: Critical and Chance values are computed for the situations in which the proportion of note events containing boundaries is 1/8 (π=0.125), 1/6 (π=0.167) and 1/4 (π=.250).

π is essentially the ratio (a+c):(b+d) for Boundary Profile A or the ratio (a+b):(c+d) for Boundary Profile B.

It was assumed that (a+c)/(b+d) = (a+b)/(c+d) (equal numbers of boundaries per boundary profile).

The first analysis looked at the consistency within each subject by determining the similarities of the first, second and third repetitions on a subject-by-subject basis: These analyses will be referred to as *consistency analyses*. If subjects performed consistently, then all measures should produce high numbers (i.e., nearer to 1.0, than to 0.0) between the repetitions; in particular, because subjects were explicitly instructed to use the first and second repetitions to refine their answers, one would expect a highest similarity between the second and third repetitions. (This prediction rests to some degree, on the learning curve for the task in conjunction with the learning curve for the melody, so it should not be taken as binding.)

The second analysis looked at similarity across subjects by determining the similarity ratings between all subjects on the third repetition. These analyses will be referred to as *similarity analyses* to distinguish them from the previous consistency analyses. Only the third repetition was used for this comparison by virtue of the aforementioned instructions. The examination of between subject consistency was augmented by a cluster analysis on these similarity ratings: Such an analysis can indicate whether or not there are any homogeneous subgroups within the complete set of 61 subjects. Cluster analyses appear in many guises, but the distinctions only matter in the event that the different versions produce different results. Several variations of the basic cluster analysis were used (average between-groups linkage, average within-groups linkage, single linkage and complete linkage), but only the average between-groups linkage will be reported since no strong differences emerged.

Before comparisons across subjects could be made, one minor pretest was performed. It was theoretically possible that individual subjects scored boundaries differently: That is, some subjects could place boundaries at the beginning of units and some could place boundaries at the end of units (see Figure 3.13). If such a difference existed, then the implication is that some subjects might be shifted by one note event with respect to other subjects. In addition, a similar shift might be produced if some subjects had longer response times (for the actual key press), thereby producing a relative shift in the location of boundaries.

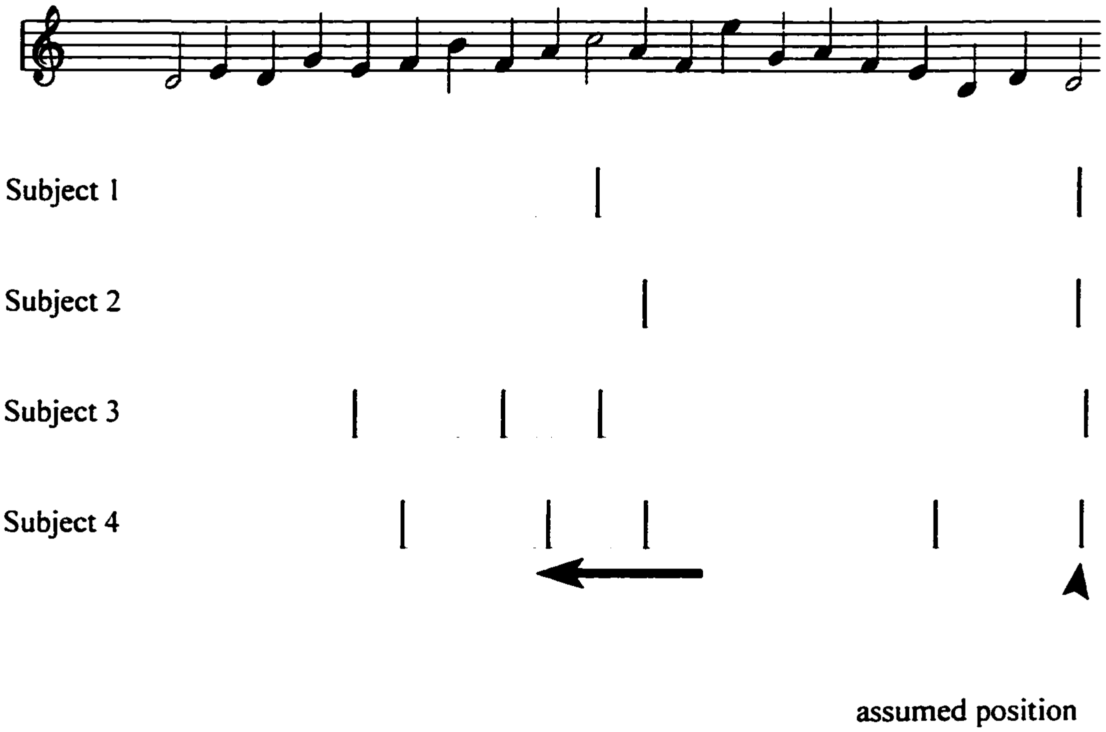


Figure 3.13: The effect of the meaning of boundaries on the similarity of profiles. In this case, all profiles would be aligned to achieve the maximum correlation before any subsequence analysis. Althought the direction of shift is arbitrary, overall the individual implementations were set so that the total amount of shifting was minimized. Shifts were only implemented if the alignment improved.

To assess this, the mean profile was created from all 61 subjects. Each individual subject was then correlated against this mean profile using relative shifts of -1, 0, and +1. The best shift for that individual was considered to be the value that maximized this correlation (this assumes that no one subject had a significant contribution to the average).

For example, if most subjects indicated boundaries at the beginning of units and a few subjects indicated boundaries at the end of units, then the average profile would match the majority of subjects (boundaries at the beginning). Hence, the majority of subjects who indicated boundaries at the beginning would have a shift of 0 with respect to the average profile. However, the few subject who indicated boundaries at the end would have a shift of -1 with respect to the average profile.

Conversely, if most subjects indicated boundaries at the end of units and a few subjects indicated boundaries at the beginning of units (i.e., the reverse case), then the average profile would match the majority of subjects (boundaries at the end). Hence, the majority of subjects who indicated boundaries at the end would have a shift of 0 with respect to the average profile. However, the few subject who indicated boundaries at the beginning would have a shift of +1 with respect to the average profile.

It is important to realize that this information (i.e., collections of subjects showing the same relative shift) would be extracted in the aforementioned cluster analysis. However, this relative shifting is not the point of the aforementioned cluster analysis. Two subjects should be considered similar if they create similar groups of notes. If two subjects had created identical groupings of notes but one of the subjects had indicated the onset of groups while the other had indicated the closure of groups, then the cluster analysis would consider them different. By performing this preliminary analysis, this problem is avoided.

In the worst case for this preliminary analysis, three distinct groups could be produced, which could then be renormalized to a shift of 0. As it happened, all subjects were maximized for a shift of 0 (all subjects were at the same shift), indicating a remarkable degree of intersubject consistency, so only those results are reported.

The third analysis created boundary profiles for each repetition, based on the groups that were extracted (if any).

The fourth analysis examined the predictability of boundaries based on the Group Preference Rules (GPR) of Lerdahl and Jackendoff (1983). The analysis was essentially the same as that of the previous chapter: The previous analysis fitted the quantified rules of Lerdahl and Jackendoff to the grammatical parsing of the lyrics and to Lerdahl and Jackendoff's example of Mozart's work. In this case, the rules were used to predict the empirically derived boundaries based on the average boundary profile. Essentially, the correlation between the empirically derived boundaries and each of the rules was obtained. Since Lerdahl and Jackendoff have argued that the individual rules should summate to create stronger boundaries, a simple multiple regression analysis was used to determine the best combination of rules for the formation of boundaries. Because the Lerdahl and Jackendoff rules were assigned to the note event beginning a unit while the choice of subjects was not known (subjects could respond to the opening or closing of a unit), it was necessary to test the match between the rules and the empirical average using three relative shifts (i.e., shifts of -1, 0, and 1). Only the best match is presented. Note that this shifting is not the same as the alignment between different subjects (above). This is the alignment of the average of all subjects to the theoretical boundaries.

General Considerations for Boundary Efficacy (Part 2 of each Stage)

The fifth analysis examined the utility of the boundary locations. If boundary construction is more than an artifact of the experiment (regardless of any intersubject consistency), then boundary formation should have some impact on subsequent recall for melodic material. Stated succinctly, retrieval should be affected by the number of units that a subject is asked to retrieve. Hence, performance when subjects are asked to retrieve two units from memory should be lower than performance when subjects are asked to retrieve only one unit from memory.

As previously described, to test this, in Part 2 of each of Stages 2, 4, 6 and 8, subjects were provided with test sequences that had been extracted from the melody and were requested to indicate whether or not each test sequence actually came from the melody. One half of the test sequences straddled a boundary. Hence, these comparisons required the retrieval of two units from memory (see discussion in the Introduction of this chapter). The other half of the test sequences were wholly contained within a single unit. Hence these comparisons required the retrieval of only one unit from memory. The extraction and construction of test sequences was done on a subject-by-subject basis, so subjects were tested for their own boundaries, not for boundaries based on the average profile.

While the simple dichotomy of one or two units allows for some predictions (cf., Peretz, 1989), there are several caveats. For the moment, assume that there exists a single "proper" parsing of a melody into its units (in the manner that a sentence may be parsed into its words). Firstly, the relationship between boundary locations in any empirical boundary profile (i.e., either from a single subject or from the average of subjects) and the true boundary locations in the melody is not yet known. It is possible that the empirical boundary location in the profile indicates either the beginning or the end of a unit. For example, as shown in Figure 3.14, a boundary location indicated on the first note of a test sequence might be the first note of a unit (if the subject uses key presses to indicate the beginning of units) or it might indicate the end of a unit (if the subject uses key presses to indicate the end of units). As such, a boundary on the first note could represent either one or two units. The situation is similar for the last (fourth) note of the test sequence: a boundary on the fourth note could imply two memory units (if boundaries indicate the beginning of units) or one memory unit (if boundaries indicated the end of units). Ideally, boundaries placed on the second or third note of the test sequence should be unambiguous (they represent two units -- possibly and probably incomplete units). The implication for analysis is that the different boundary conditions must be treated as categorically different. That is, although it seems that there is an interval scale for the location of boundaries in the test sequence, in truth it only categorical.

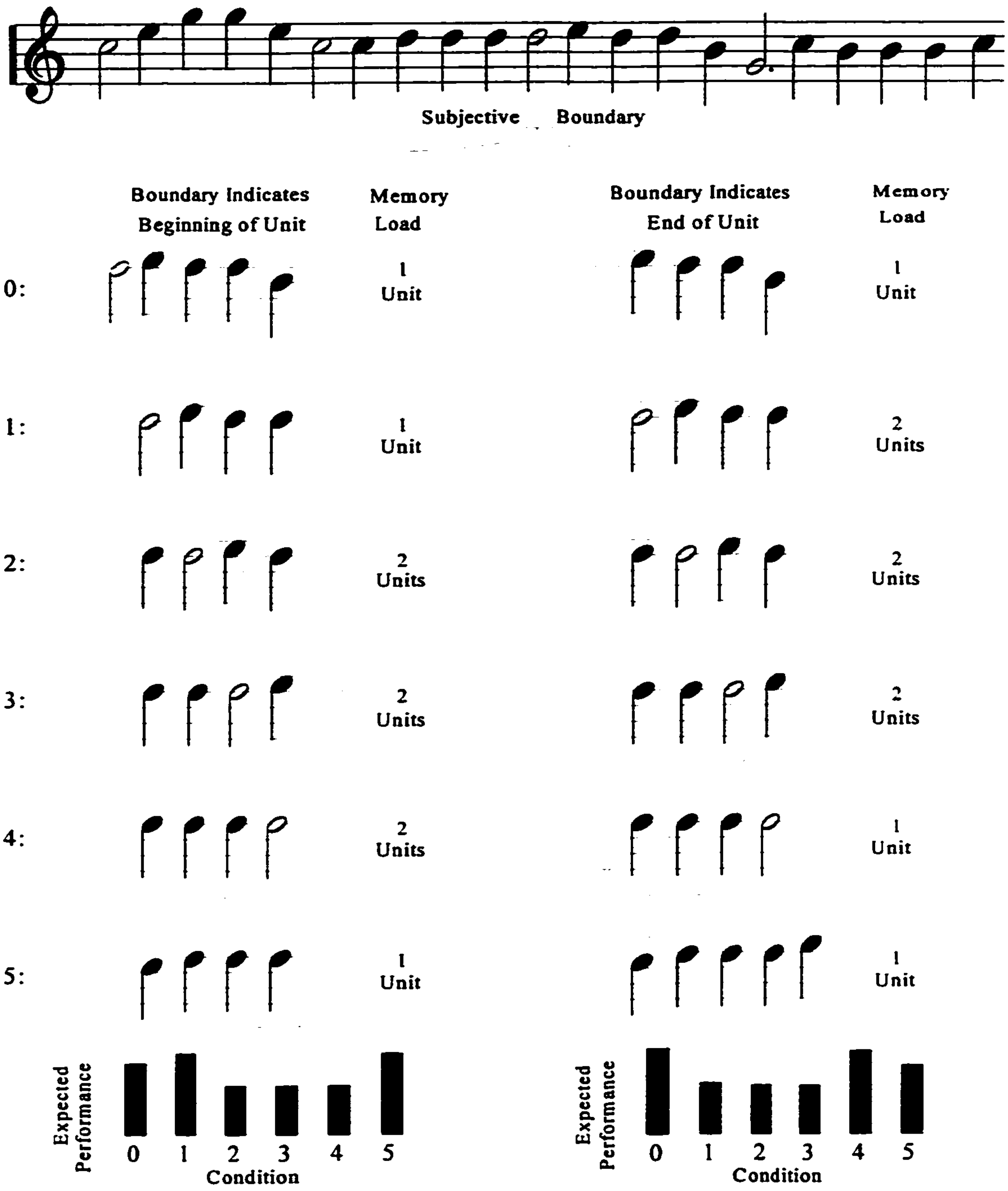


Figure 3.14: Performance implications based on memory-load considerations, when boundaries indicate the beginning or the end of a unit for each condition. Conditions 0 & 5 always represent 1 unit (higher performance) while Conditions 2 & 3 always have 2 units (lower performance). Additionally, some test sequences will start or end a memory unit, which may further enhance performance. The two lightened notes are a notational convenience: They were not presented.

There is a further reason for this constraint. Given that there may be a response delay between the detection and the recording of a boundary (i.e., detection and key-press), it is possible that the indicated boundary location does not actually represent the existence of two memory units. For example, a boundary located on the second note of a test sequence could represent a (long) delay to respond to a real boundary on the note before the test sequence. Similarly, a boundary located on the note following a test sequence may actually represent a real boundary located on the third note of the test sequence, and hence, two memory units.

For these reasons, the coding of boundary location within/around the test sequence was considered categorical. This has important consequences for the type of analysis (i.e., a regression analysis of ANOVA-type data). That is, the independent variable was considered categorical, and the six main categories were:

Condition 0: boundary on the note event preceding the test sequence
Condition 1: boundary on the first note event of the test sequence
Condition 2: boundary on the second note event of the test sequence
Condition 3: boundary on the third note event of the test sequence
Condition 4: boundary on the fourth note event of the test sequence
Condition 5: boundary on the note event following the sequence

Contrast analysis (vectors) was used to compare the various conditions. Ideally, for each subject there should be four trials within each of Conditions 0 and 5, and two trials within each of Conditions 1, 2, 3, and 4 (see the Introduction & Methods of this chapter for rationale). These 16 conditions were further divided into two Types: one half were designated as "Literals" and one half were designated as "Foils". The Literal sequences were exact four-note extracts from the melody. The Foil sequences were four-note extracts in which one randomly selected tone from the test sequence had been randomly changed by ±1 or ±2 semitones. Although the change was selected randomly, its position (variable Location: 1, 2, 3 or 4) and magnitude (variable Size: -2, -1, 1 and 2) were recorded for subsequent, secondary analyses that might be of use for the design of subsequent experiments. In addition, the change in the tonality of the note was also recorded. That is, the key of the entire melody was computed using the Krumhansl key-finding algorithm (Frankland & Cohen, 1996; Krumhansl, 1990), and then the valence of the changed note within that key was determined before and after the change. Based on past work and common notions of tonality within music theory, three levels of stability (importance) were considered: Level 1 consisted of the notes of the tonic triad (the tonic, mediant and dominant), Level 2 consisted of the remaining non-triad, diatonic notes (the supertonic, subdominant, submediant and leading tone), and Level 3 consisted of the non-diatonic notes (minor second, minor third, augmented fourth, augmented fifth and augmented sixth). Changes between these levels were tagged using a categorical scale that considered no change in level (movement within Level 1, Level 2 or Level 3), a change in one level (movement between Levels 1 and 2 or Levels 2 and 3), or a change in two levels (movement between Levels 1 and 3 only), as well as the direction of change (less tonal to more tonal or vice versa). In addition, such changes were labeled (ranked) by detectability. It was assumed, for example, that a sequence that contained a note that changed from a Level 1 to Level 3 would be obvious to the listener (i.e., easily detected as a Foil). One can surmise that a sequence that contained a note that changed from Level 3 to Level 1 would not be detected as a Foil (i.e., unstable notes can drift to more stable notes in a memory representation; cf., Frankland & Cohen, 1996; cf., Cuddy, Cohen & Mewhort, 1881; Cuddy, 1993). Full details of this coding are presented in Table 3.9, but the ranking is not intended as anything more than a subjective guide (though, based on the literature). Categories of change are treated as a categorical scale (i.e., not ordinal, not interval). In this coding, a Literal sequence was arbitrarily labeled as 0. In addition, the analysis of tonality within test sequences is considered secondary to the purpose of this experiment.

Table 3.9: The categorical coding of the changes in the tonality of the foil test sequences.

Tonal Function Before Change	Tonal Function After Change	Rank of Change
Level 1: tonic mediant dominant	Level 1: tonic mediant dominant	4
	Level 2: supertonic subdominant submediant leading tone	3
	Level 3: minor second minor third augmented fourth minor sixth minor seventh	1
Level 2: supertonic subdominant submediant leading tone	Level 1: tonic mediant dominant	7
	Level 2: supertonic subdominant submediant leading tone	5
	Level 3: minor second minor third augmented fourth minor sixth minor seventh	2
Level 3: minor second minor third augmented fourth minor sixth minor seventh	Level 1: tonic mediant dominant	8
	Level 2: supertonic subdominant submediant leading tone	9
	Level 3: minor second minor third augmented fourth minor sixth minor seventh	6

Note that all of the changes (Size, Location and Tonality) were only applicable to the Foil sequence. Hence, either Size or Location set to 0 indicated a Literal sequence (note that a Size of 0 could never be combined with a Location other than 0 and vice versa).

Because extracts were constructed on-line, the actual algorithm for test sequence construction needed two additional conditions:

Condition -1: no boundaries near the test sequence
Condition -2: more than one boundary in or adjacent to the test sequence

The negative values used in the label serve to remind that these conditions are not part of the true experimental design. Since the meaning of Condition -1 is clear, it was included in subsequent analyses. However, Condition -2 was not included in any subsequent analysis because the meaning of Condition -2 was not clear. Condition -2 would always include two boundaries, but either one or both of those boundaries could be on the ends of the sequence implying one perfectly bounded unit or either one or both of those boundaries could be within the sequence, implying two or three units. Finally, before the analysis could begin, it was necessary to check to determine if changes to the test sequence (i.e., the creation of the Foil) resulted in a "new" test sequence that represented a Literal sequence extracted from a different part of a melody.

The statistical analysis of the test sequences was divided into two components to clarify the important relationships. Initially, the analysis of the test sequence only examined proportion correct as a function of Condition, Type (Literal and Foil) and their interaction. These variables were controlled by design. Subsequent analyses further examined the performance on Foils by examining Size, Location, and Tonality, in conjunction with Condition (and Type which is implicit in Size, Location and Tonality) as well as their interactions. These analyses were considered secondary because these were not controlled explicitly in the design. In all analyses, the extraction position of the test sequence within the melody and the sequential position of the test sequence within the presentation of test sequences were treated as covariates to control for primacy and recency effects.

Boundary Location: Stage 4, Part 1

In Part 1 of Stage 4, (see Table 3.10), the mean consistency values (across all repetitions) for subject-by-subject repetitions of the melody were 0.94±0.10 for the simple similarity measure, 0.73±0.35 for the Jaccard measure, 0.83±0.26 for the Kulczynski measure 2, 0.90±0.16 for the Sokal and Sneath measure 4 and 0.79±0.32 for the phi measure. Clearly all values indicated a high degree of similarity within individual subjects: The modal value for many measures was 1.00 indicating that most subjects performed identically across the various repetitions. Furthermore, by both the standard deviations and the minima, subjects never fell to the level of chance indicated in Table 3.8. However, recall that these chance levels assumed equal numbers of boundaries per boundary profile and a particular probability for boundaries within a boundary profile.

Table 3.10: The average consistency of subjects over different repetitions of the same melody within Stage 4, Part 1, given the similarity measures of Table 3.8.

	Consistency	Mean & SD	Median	Mode	Min	Max
Simple Similarity	1 to 2	0.940±0.050	0.938	0.938	0.750	1.000
	1 to 3	0.933±0.050	0.938	0.917	0.750	1.000
	2 to 3	0.942±0.067	0.958	0.958	0.708	1.000
Jaccard Similarity	1 to 2	0.728±0.192	0.714	1.000	0.143	1.000
	1 to 3	0.710±0.205	0.700	1.000	0.143	1.000
	2 to 3	0.754±0.212	0.800	1.000	0.263	1.000
Kulczynski 2 Similarity	1 to 2	0.834±0.137	0.844	1.000	0.333	1.000
	1 to 3	0.821±0.149	0.826	1.000	0.333	1.000
	2 to 3	0.847±0.158	0.889	1.000	0.417	1.000
Sokal & Sneath 4 Similarity	1 to 2	0.899±0.088	0.905	1.000	0.596	1.000
	1 to 3	0.890±0.092	0.894	1.000	0.596	1.000
	2 to 3	0.905±0.099	0.932	1.000	0.611	1.000
Phi Similarity	1 to 2	0.796±0.167	0.806	1.000	0.174	1.000
	1 to 3	0.777±0.186	0.788	1.000	0.174	1.000
	2 to 3	0.809±0.199	0.863	1.000	0.222	1.000

To further examine the consistency ratings, the correlation between consistencies were computed across subjects. That is, the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 1 and 3 (see Table 3.11), the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3 and the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3. Regardless of the measure used, the resulting correlations were all high (and significant at a type 1 error rate of $\alpha=0.05$), indicating that the subjects who were the most consistent in the first two repetitions remained the most consistent in the third repetition.

Table 3.11: The correlations between consistency ratings over the different repetitions of the same melody within Stage 4, Part 1, given the similarity measures of Table 3.8.

Correlations	Repetitions	Consistency Between Repetitions		
		1 to 2	1 to 3	2 to 3
Simple Similarity	1 to 2	1.000	0.738**	0.650**
	1 to 3		1.000	0.766**
Jaccard Similarity	1 to 2	1.000	0.664**	0.566**
	1 to 3		1.000	0.599**
Kulczynski 2 Similarity	1 to 2	1.000	0.655**	0.597**
	1 to 3		1.000	0.652**
Sokal & Sneath 4 Similarity	1 to 2	1.000	0.672**	0.607**
	2 to 3		1.000	0.680**
Phi	1 to 3	1.000	0.672**	0.605**
	2 to 3			0.682**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

As shown in the within-subjects ANOVA of Table 3.12, consistency ratings did not change as a function of the repetitions being compared, regardless of measure used: None of the overall ANOVAs and none of the contrasts (comparing the consistencies of nearest neighbours in time and comparing the near neighbours in time to the distant neighbours in time) were significant.

Table 3.12: The ws-analysis of the change in consistency ratings for the different repetitions of the same melody within Stage 4, Part 1, given the similarity measures of Table 3.8.

WS-ANOVA		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Simple	Overall	0.003	2	0.105	120	1.509	0.225	0.025	0.008
	Ψ_1	0.000	1	0.065	60	0.157	0.693	0.002	
	Ψ_2	0.003	1	0.040	60	3.714	0.059	0.058	
Jaccard	Overall	0.041	2	1.208	120	2.060	0.132	0.033	0.017
	Ψ_1	0.016	1	0.669	60	1.473	0.230	0.024	
	Ψ_2	0.025	1	0.539	60	2.788	0.100	0.044	
Kulczynski 2	Overall	0.017	2	0.658	120	1.505	0.226	0.024	0.008
	Ψ_1	0.006	1	0.362	60	0.922	0.341	0.015	
	Ψ_2	0.011	1	0.296	60	2.218	0.142	0.036	
Sokal & Sneath 4	Overall	0.006	2	0.276	120	1.380	0.255	0.023	0.006
	Ψ_1	0.002	1	0.156	60	0.593	0.444	0.010	
	Ψ_2	0.005	1	0.120	60	2.408	0.126	0.049	
Phi	Overall	0.024	2	0.912	120	1.552	0.216	0.025	0.009
	Ψ_1	0.005	1	0.513	60	0.731	0.396	0.012	
	Ψ_2	0.017	1	0.398	60	2.610	0.111	0.042	

Notes: Ψ_1 compares Consistency from 1 to 2 with Consistency from 2 to 3 (i.e., nearest neighbours in time)
Ψ_2 compares the average of Consistency from 1 to 2 and Consistency from 2 to 3 with the Consistency from 1 to 3 (close in time to far in time)
No violations of the assumptions of the ws-anova were observed.

When comparing across subjects (Table 3.13), average similarity ratings were lower, as would be expected, with means of 0.90±0.07 for the simple similarity measure, 0.60±0.20 for the Jaccard measure, 0.75±0.16 for the Kulczynski measure 2, 0.85±0.10 for the Sokal and Sneath measure 4 and 0.69±0.21 for the phi measure. These values are still above the critical and chance levels specified in Table 3.8. Only the minimum value for the Jaccard similarity measure dips below the aforementioned chance value.

Table 3.13: The average similarity ratings between different subjects for the same melody, within Stage 4, Part 1, Repetition 3 only, given the similarity measures of Table 3.8.

Similarity Measure	Mean & SD	Median	Mode	Min	Max
Simple	0.900±0.074	0.917	0.938	0.625	1.000
Jaccard	0.602±0.201	0.636	0.667	0.100	1.000
Kulczynski 2	0.752±0.159	0.800	0.833	0.261	1.000
Sokal & Sneath 4	0.846±0.102	0.876	0.899	0.514	1.000
Phi	0.686±0.205	0.737	0.787	0.029	1.000

Overall, the central tendencies in the similarity analyses (mean, mode, median) indicate that subjects perform more-or-less the same. However the minima on both the Jaccard and phi measures (note that in both the consistency and the similarity analyses, it is the Jaccard and phi measures that demonstrated the greatest range) imply that a few subjects are somewhat different -- not distinctly different, but different enough to be a consideration in later analyses.

To further explore the relationship between different subjects, a cluster analysis was performed on each of the similarity matrices (i.e., using each the similarity measures) determined from the third repetition of the melody. One can use the previously defined critical values to define a cut line for the creation of subgroups. However, as might be expected given the high degree of similarity between different subjects, no distinct groups of subjects were found. That is, the cluster profiles were indicative of a situation in which individual subjects were gradually added to an homogeneous group: Using such a cut line did not create two or more distinct groups of subjects. Only the cluster analysis for the phi measure is presented in Figure 3.15 because all were very similar (see the Synopsis later in this chapter).

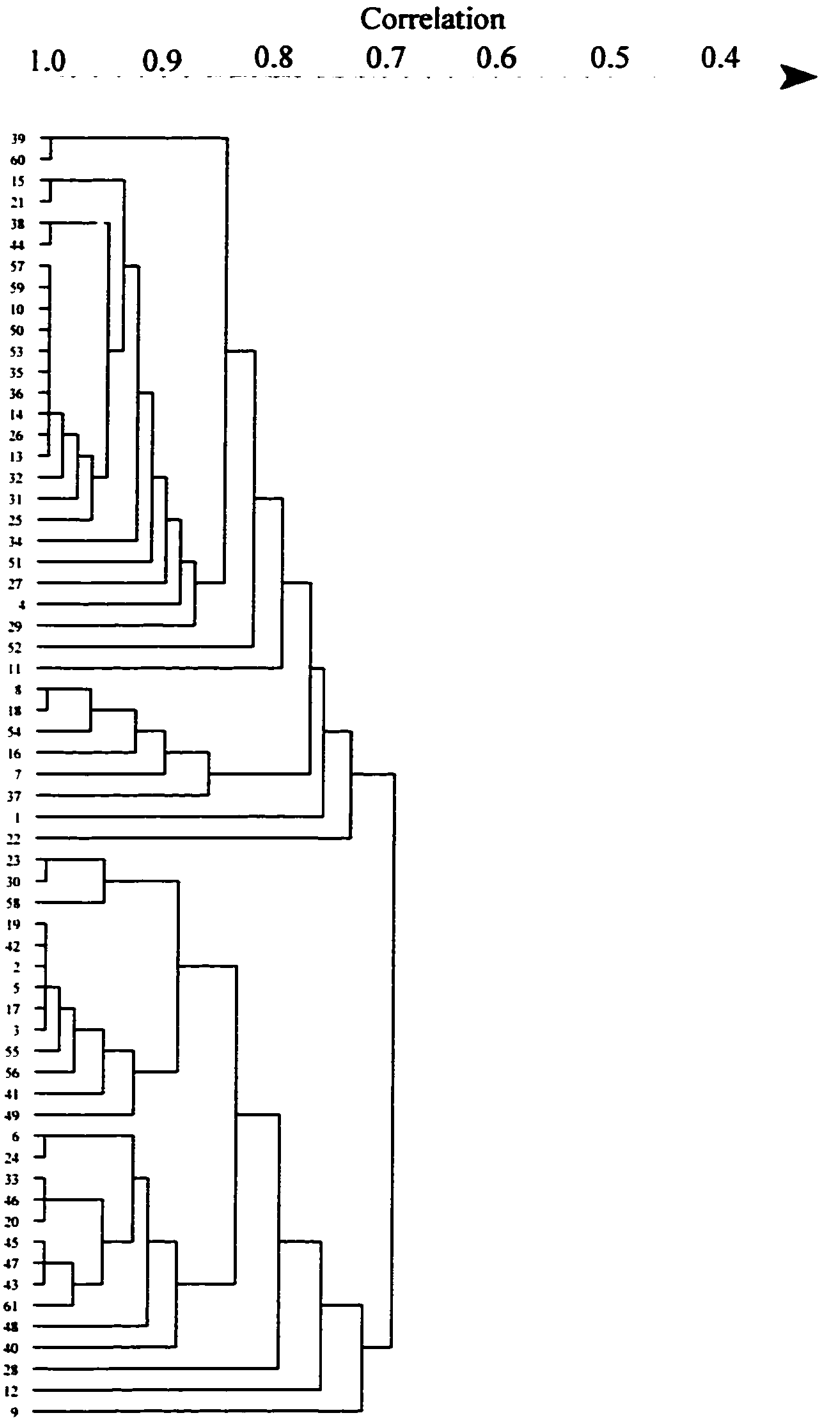


Figure 3.15: Cluster analysis of phi similarities within Stage 4 (scaling is set to enable comparisons with later analyses).

All the results (consistency analysis, similarity analysis and cluster analysis) imply that all subjects are fairly similar. This is more impressive when one considers that subjects exhibited a large range of training (assessed by numerous measures) and a large range of involvement with music (assessed by various measures of playing and time spent listening to music).

Given that all subjects can be treated as one group, it is finally possible to create the average boundary profile. The average boundary profile for each repetition is shown in Figure 3.16 (the melody is split for ease of presentation). Subjects placed strong boundaries (i.e., many subjects placed boundaries at these points) at note events 4, 7, 11, and 16, with weaker boundaries at note events 25 and 36 (tied notes represent one event because they represent a single sound). It was arbitrarily decided that there should be a boundary at the end of the melody for all subjects, since most did so, and many explicitly asked if they should during the Practice Stage of the experiment. Generally the agreement between subjects is high at the beginning but lower near the end. The detailed descriptive statistics for this profile are provided in Tables 3.14 through 3.16 (various measures pertaining to the locations of boundaries) for each repetition of the melody. Over all repetitions, the mean number of boundaries was fairly constant at 8.21±0.21. The modal number of boundaries remained constant at 9.00. This number of boundaries represents a mean number of note events *between* boundaries equal to 6.11±0.09; the mean number of note events within a unit should therefore be 7.11 because each boundary should be associated with either the beginning or the end of a unit (i.e., boundaries are located on particular note events). The size of individual units ranged from 0.00 note events to 36.00 note events. However, on a subject-by-subject basis, the most common size was 3.00 note events (implying 4.00 note events when the boundary is included), and the least common sizes were 4.00, 5.00, 6.00 and 7.00 (implying 5.00, 6.00, 7.00, and 8.00 when boundaries are included; see Tables 3.14 to 3.16).

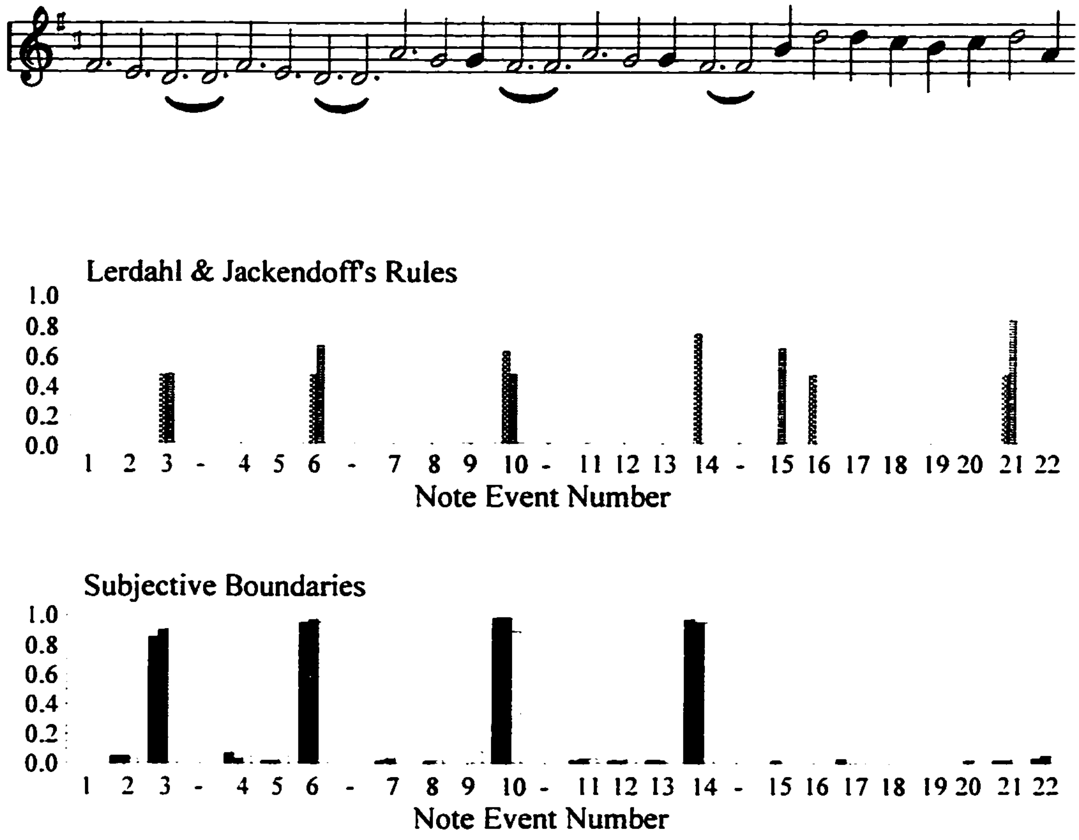


Figure 3.16a: The first part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries based on the average of all subjects (first, second and third repetition of the song). The predicted and empirical boundaries have been aligned by shifting the predicted boundary profile by one note unit to the left.

▨ **Rule 2a** ▤ **Rule 3a** ■ **First Presentation**
▩ **Rule 2b** ▧ **Rule 3d** ■ **Second Presentation**
☐ **Third Presentation**

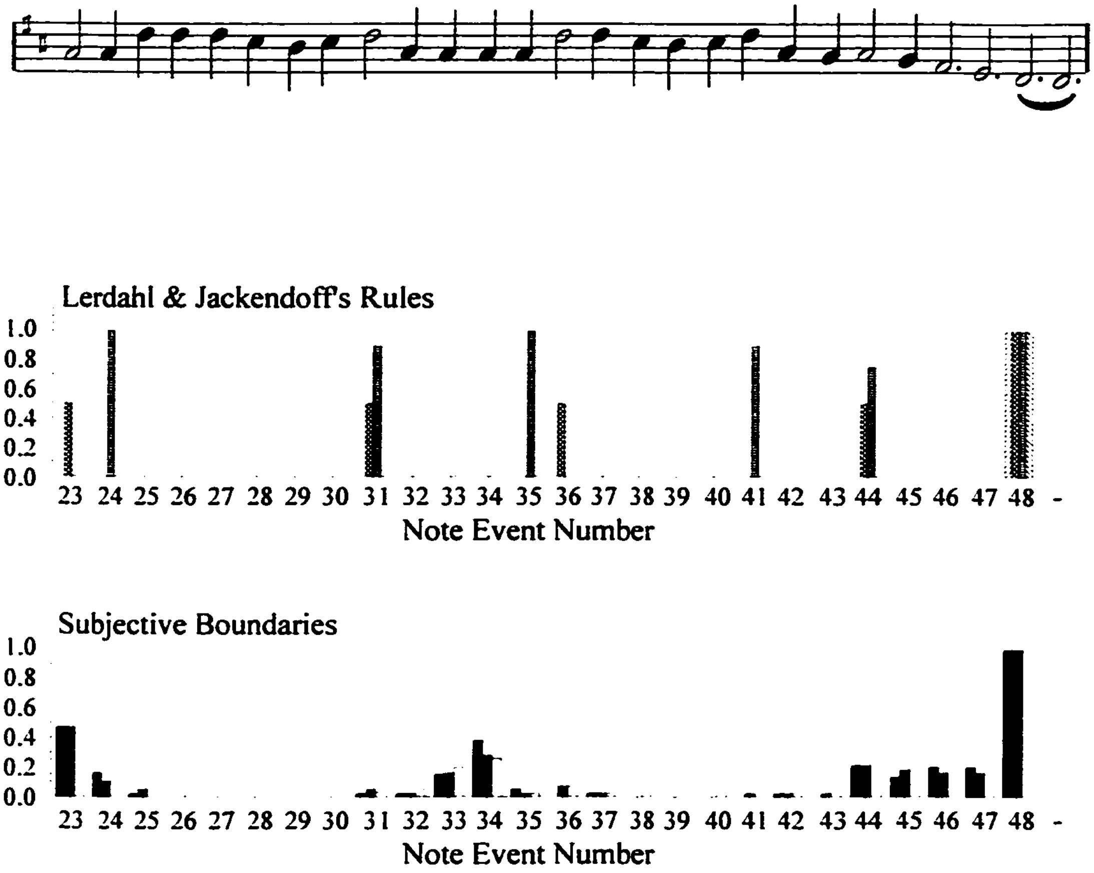


Figure 3.16b: The second part of the melody used in Stage 4, Part 1 and the predicted boundary locations given Lehrdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries based on the average of all subjects (first, second and third repetition of the song).

▨ **Rule 2a** ▩ **Rule 3a**
▦ **Rule 2b** ▧ **Rule 3d**

■ **First Presentation**
■ **Second Presentation**
☐ **Third Presentation**

For the purpose of analysis, four different measures of boundary location were available: total number of boundaries, minimum distance between boundaries, maximum distance between boundaries and mean distances between boundaries. Since mean and total represent (non-linear) transforms of each other, one can be dropped. The decision was made to drop total number of boundaries since this is somewhat arbitrary (i.e., it depends on the length of the melody being tested): the mean distance between boundaries was retained. Minimum distance, maximum distance and mean distance have some utility for interpretation and comparison between different melodies. For example, experts within a domain tend to use larger units. This could manifest in a higher minimum (if experts do not use small units), and/or a higher mean (if, on average, experts use larger units) and/or a higher maximum (if the largest units of experts are bigger than the largest units of non-experts). At this point, it seem prudent to retain all three.

Table 3.14: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 1 (the first presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	0.885±0.709	1.000	1.000	0.000	3.000
Maximum	18.623±9.000	13.000	10.000	9.000	33.000
Mean	6.210±1.938	5.714	4.875	1.938	14.667
Number with 0	0.475±1.286	0.000	0.000	0.000	9.000
Number with 1	1.082±0.586	1.000	1.000	0.000	3.000
Number with 2	1.082±0.614	1.000	1.000	0.000	3.000
Number with 3	2.016±0.591	2.000	2.000	0.000	3.000
Number with 4	0.082±0.277	0.000	0.000	0.000	1.000
Number with 5	0.000±0.000	0.000	0.000	0.000	0.000
Number with 6	0.033±0.180	0.000	0.000	0.000	1.000
Number with 7	0.066±0.250	0.000	0.000	0.000	1.000
Number with 8	0.492±0.536	0.000	0.000	0.000	2.000
Number with 9	0.492±0.649	0.000	0.000	0.000	2.000
Number with ≥10	1.246±0.505	1.000	1.000	0.000	2.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Number with "xx" is the mean number of times subjects created units containing xx note events (exclusive of the actual boundaries).

Table 3.15: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 2 (the second presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	0.836±0.522	1.000	1.000	0.000	3.000
Maximum	19.295±5.000	15.000	10.000	5.000	33.000
Mean	6.093±1.807	5.714	4.875	1.611	10.750
Number with 0	0.475±1.479	0.000	0.000	0.000	10.000
Number with 1	1.115±0.551	1.000	1.000	0.000	3.000
Number with 2	1.082±0.000	1.000	1.000	0.000	2.000
Number with 3	2.066±0.655	2.000	2.000	0.000	3.000
Number with 4	0.115±0.451	0.000	0.000	0.000	3.000
Number with 5	0.082±0.420	0.000	0.000	0.000	3.000
Number with 6	0.049±0.284	0.000	0.000	0.000	2.000
Number with 7	0.066±0.250	0.000	0.000	0.000	1.000
Number with 8	0.557±0.646	0.000	0.000	0.000	3.000
Number with 9	0.377±0.553	0.000	0.000	0.000	2.000
Number with ≥10	1.131±0.499	1.000	1.000	0.000	2.000

Notes: See Table 3.13

Table 3.16: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 3 (the third presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	0.934±1.328	1.000	1.000	0.000	10.000
Maximum	19.066±9.847	13.000	10.000	7.000	35.000
Mean	6.030±2.934	4.875	4.875	1.611	22.500
Number with 0	0.787±1.854	0.000	0.000	0.000	11.000
Number with 1	1.180±0.592	1.000	1.000	0.000	4.000
Number with 2	1.148±0.573	1.000	1.000	0.000	4.000
Number with 3	1.967±0.752	2.000	2.000	0.000	3.000
Number with 4	0.115±0.370	0.000	0.000	0.000	2.000
Number with 5	0.033±0.180	0.000	0.000	0.000	1.000
Number with 6	0.033±0.180	0.000	0.000	0.000	1.000
Number with 7	0.098±0.351	0.000	0.000	0.000	2.000
Number with 8	0.557±0.563	1.000	1.000	0.000	2.000
Number with 9	0.377±0.489	0.000	0.000	0.000	1.000
Number with ≥10	1.148±0.441	1.000	1.000	0.000	2.000

Notes: See Table 3.14.

From the associated correlation matrices (Tables 3.17, 3.18, 3.19), one can see that the minimum number of notes between boundaries (the smallest grouping size for each subject) was not correlated with the maximum number of notes between boundaries (the largest grouping size for each subject). A positive correlation would imply that some subjects used bigger units than other subjects. That is, a positive correlation would imply that those who had the highest maximum, also had the highest minimum. It is interesting that this is not true because, a priori, one might predict that musically trained listeners would on average tend to use larger units (i.e., their chunks would be larger) than untrained listeners. Hence, a priori, one might expect a subgroup of subjects (with high training) to have both high maxima and high minima, thereby producing some correlation between any two of minimum, maximum or mean distances.

Table 3.17: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 1 (the first presentation of the melody).

	Correlations Between Measures		
	Minimum	Maximum	Mean
Minimum	1.000		
Maximum	0.277	1.000	
Mean	0.607**	0.787**	1.000
Number with 0	-0.469**	-0.134	-0.369*
Number with 1	-0.458**	-0.211	-0.432**
Number with 2	-0.476**	-0.421**	-0.603**
Number with 3	0.084	-0.284	-0.310*
Number with 4	0.473**	0.037	0.190
Number with 5	0.000	0.000	0.000
Number with 6	-0.232	-0.154	-0.274
Number with 7	-0.239	-0.244	-0.294
Number with 8	-0.068	-0.460**	-0.256
Number with 9	-0.056	-0.514**	-0.380*
Number with ≥10	0.034	-0.203	-0.018

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.18: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 2 (the second presentation of the melody).

	Correlations Between Measures		
	Minimum	Maximum	Mean
Minimum	1.000		
Maximum	-0.035	1.000	
Mean	0.279	0.846**	1.000
Number with 0	-0.523**	-0.110	-0.432**
Number with 1	-0.397**	-0.231	-0.533**
Number with 2	-0.014	-0.193	-0.195
Number with 3	0.178	-0.238	-0.243
Number with 4	-0.060	-0.193	-0.233
Number with 5	-0.090	-0.237	-0.176
Number with 6	-0.169	-0.205	-0.330*
Number with 7	-0.300*	-0.122	-0.342*
Number with 8	0.325*	-0.626**	-0.467**
Number with 9	0.045	-0.495**	-0.315*
Number with ≥10	0.084	-0.068	0.118

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.19: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 3 (the third presentation of the melody).

	Correlations Between Measures		
	Minimum	Maximum	Mean
Minimum	1.000		
Maximum	0.264	1.000	
Mean	0.870**	0.655**	1.000
Number with 0	-0.304*	-0.092	-0.370*
Number with 1	-0.409**	-0.134	-0.475**
Number with 2	-0.272	-0.315*	-0.405**
Number with 3	-0.269	-0.308*	-0.288
Number with 4	-0.018	-0.204	-0.122
Number with 5	-0.061	-0.180	-0.227
Number with 6	-0.061	-0.171	-0.124
Number with 7	0.050	-0.175	-0.042
Number with 8	-0.129	-0.698**	-0.393**
Number with 9	-0.141	-0.584**	-0.343*
Number with ≥10	0.245	-0.056	0.187

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

A within-subjects ANOVA was performed on the minimum, maximum and mean measures of boundary location to see if there were any changes as a function of repetition (see Table 3.20). There were none. The associated correlation matrix (Table 3.20) implies that individuals were more consistent over the last two repetitions.

Table 3.20: The analysis of the changes in the boundary statistics across repetitions of the same melody, for Stage 4, Part 1.

WS-ANALYSIS		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Minimum Distance Between Boundaries	Overall	0.295	2	77.705	120	0.228	0.802	0.004	0.000
	Ψ_1	0.074	1	44.426	60	0.095	0.759	0.002	
	Ψ_2	0.221	1	33.279	60	0.399	0.530	0.007	
Maximum Distance Between Boundaries	Overall	14.240	2	4.05.093	120	0.211	0.809	0.004	0.000
	Ψ_1	5.975	1	2118.523	60	0.169	0.682	0.003	
	Ψ_2	8.265	1	1916.568	60	0.259	0.613	0.005	
Mean Distance Between Boundaries	Overall	1.016	2	379.857	120	0.161	0.852	0.003	0.000
	Ψ_1	0.987	1	259.740	60	0.228	0.635	0.004	
	Ψ_2	0.030	1	120.117	60	0.015	0.904	0.000	

Correlation Within Measures			Repetition Number		
			1	2	3
Minimum Distance Between Boundaries	Repetition Number	1	1.000	0.308*	0.381**
		2		1.000	0.104
Maximum Distance Between Boundaries	Repetition Number	1	1.000	0.649**	0.612**
		2		1.000	0.663**
Mean Distance Between Boundaries	Repetition Number	1	1.000	0.484**	0.388**
		2		1.000	0.572**

Notes: The test on the Maximum Distance between Boundaries was the only test to not violate the Mauchly sphericity test, but in no case did any corresponding multivariate test, Pillais, Hotellings Wilks or Roys, indicate a different conclusion.

* Significant (one tailed) at $p<0.05$

** Significant (one tailed) at $p<0.01$

Given that there exists an average boundary profile that can be said to represent the parsing by individual subjects, the question shifts to one of the prediction of those boundaries based upon the structure of music. The empirically derived average boundary profile was correlated against each of Lerdahl and Jackendoff's (1983) rules for parsing.

As was mentioned earlier (cf., the analysis of Chapter 2), the correlation was calculated for three values of relative shift (i.e., -1, 0, and 1) between the empirically derived boundary profile and the theoretical boundary profiles based on each rule. In this aligning process, the theoretically derived boundary profiles (and the grammatical coding) were shifted to the left, not shifted, and to the right and the correlations that were determined from each shift are presented in Table 3.21. Interestingly, the best match occurred when the theoretically derived boundary profile was shifted by one note event to the right for both Rules 2b (Attack-point proximity) and 3a (Register change). The profile for the grammatical coding of lyrics did not have to be shifted (0 shift). This implies that subjects placed their boundaries at the ends of units. In the construction of the theoretically derived boundary profiles, Lerdahl and Jackendoff's rules were used to place boundaries at the beginning of units. The grammatical coding of the lyrics was used to place boundaries at the end of units. The intercorrelations between the rules and the grammatically coding of the lyrics are repeated for convenience in Table 3.21.

Rule 2b (Attack-point) was generally successful at predicting boundaries (average r=0.77; r^2=0.59), while Rule 3a (Register change) was marginally successful at predicting boundaries (average r=0.34; r^2=0.12). The grammatical coding of the words in the melody was also correlated with boundary location (r=0.59, r^2=0.35), implying that subjects may have been using their knowledge of the lyrics to assist in the placement of boundaries.

Table 3.21: The correlations between the rules of Lerdahl and Jackendoff and the empirically derived boundary profiles for each repetition of the melody in Stage 4, Part 1, as well as the correlations between the rules. The rules were shifted to align to the empirical data.

	Relative Shift	Repetition Number		
		1	2	3
Repetition 1		1.000	0.996**	0.991**
Repetition 2			1.000	0.994**
Repetition 3				1.000
Rule 2b	+1	0.758**	0.776**	0.762**
	0	-0.176	-0.178	-0.173
	-1	-0.116	-0.113	-0.113
Rule 3a	+1	0.347*	0.345*	0.336*
	0	-0.203	-0.171	-0.177
	-1	-0.086	-0.089	-0.077
Grammatical Coding of Lyrics	+1	-0.118	-0.125	-0.091
	0	0.590**	0.593**	0.584**
	-1	-0.048	-0.068	-0.077

Correlations Between Measures				
	Rule 2a	Rule 2b	Rule 3a	Rule 3d
Rule 2b	1.000	—	—	—
Rule 2b		1.000	0.369**	—
Rule 3a			1.000	—
Rule 3d				1.000
Grammar	—	0.429**	0.064	—

Notes: Repetitions are the average boundary profile for each repetition (61 subjects).
Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.
Grammatical is the grammatical coding of words/phrases within the melody.
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

The profiles themselves are much more informative (see Figure 3.16). In Figure 3.16, it is the theoretical boundary profiles based on the rules that is shifted to match the boundary profile since it is the latter that is considered the more accurate: that is, theory is aligned with reality, not vice versa. Rule 2b (Attack-Point) is quite successful for the earlier part of the melody (note events 3, 6, 10 and 14), but performs less well near the end (false predictions at note events 16, 21, 31, and 36). The subjective boundaries near the end are, however fewer and much weaker (i.e., less agreement between subjects). Rule 3a (Register Change), although marginally successful from a statistical standpoint, is obviously abysmal for prediction. Note that Rules 2a (Slur/Rest) and 3d (Length Change) did not actually apply within this particular piece of music, and that Rules 3b (Dynamic change) and 3c (Articulation change) were not applicable within this study since they were controlled at the level of the stimulus.

The final analysis used multiple regression to determine the best mix of rules (using the single optimal shift) for each empirically derived boundary profile at each repetition of the melody (see Table 3.22). For each repetition, only Rule 2b (Attack-point) was needed to explain boundary locations, producing an average R^2=0.59. The complete equation for each stage represents the forced entry of both rules, providing the maximal prediction possible. A stepwise solution demonstrated that only Rule 2b would be necessary. That is, the small improvement in prediction that Rule 3a provided was not worth the loss in df (in the error term). In Table 3.22, the change R^2 for the stepwise solution is shown for those terms that were included in the stepwise equation. Note that the statistics for the change in R^2 for the first term entered (Rule 2b) are the same as the statistics for R^2 for the entire equation at that point (it is the entire equation at that point).

Table 3.22: The regression equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) in the average boundary profile for each repetition of the melody, in Stage 4, Part 1.

Regression Equations for Each Repetition: Stage 4, Part 1						
Rep	Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
1	B = 0.03±0.03 + (0.86±0.12)*Rule2b	0.575	0.000	0.575	62.179	0.000
	- (0.01±0.01)*Rule3d			0.575	30.419	0.000
2	B = 0.03±0.03 + (0.88±0.12)*Rule2b	0.602	0.000	0.602	69.630	0.000
	- (0.02±0.09)*Rule3a			0.603	34.111	0.000
3	B = 0.04±0.03 + (0.86±0.11)*Rule2b	0.580	0.000	0.580	63.650	0.000
	- (0.02±0.09)*Rule3a			0.581	31.202	0.000

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.

Each equation represents the maximal equation when all rules are used to predict boundary location. Within each repetition, the rules above the dotted line are those that were included using a forward stepwise regression analysis. Rules below the dotted line were not included in the stepwise approach.

Because there was a high similarity between different repetitions, this analysis was repeated using the average of the boundary profiles from Repetitions 2 and 3. To the extent that the results based on the average replicate the results based on the individual repetitions, the average can be used in future analyses. Generally, the results are the same (see Table 3.23).

Table 3.23: The correlations and equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) using the average of the boundary profiles for the second and third repetitions, in Stage 4, Part 1.

	Shift	Stage 4
Rule 2b	+1	0.770**
	0	-0.176
	-1	-0.113
Rule 3a	+1	0.340*
	0	-0.174
	-1	-0.083
Grammatical Coding of Lyrics	+1	-0.108
	0	0.590**
	-1	-0.073

Regression Equation for Average of Repetitions 2 and 3: Stage 4					
Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
B = 0.08±0.06 + (1.74±0.24)*Rule2b	0.593	0.000	0.593	67.028	0.000
- (0.04±0.18)*Rule3a			0.593	32.847	0.000

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.

The previous similarity analyses implied that most subjects are very similar, but given the minima of both the Jaccard and phi measures, it is important to check if this high degree of similarity actually applies in the comparison of the theoretical and empirical boundary profiles. The simplest manner in which to accomplish this is to examine the relationship between the theoretical boundary profile and the empirical boundary profile of each subject at each repetition. Unfortunately, the empirical boundary profile of any one subject at any one repetition does not contain a lot of information (i.e., may be susceptible to variation). To circumvent this, the empirical boundary profiles from the second and third repetitions were averaged for each subject (given the previous consistency analysis, this averaging is valid): This profile was called the individual boundary profile (i.e., the individual boundary profile consists of the average response of a single subject). The individual boundary profile for each subject was then compared to each of the rules.

The average correlation between the individual boundary profiles and Rule 2b was r=0.68±0.10. Similarly, for Rule 3a, the average correlation was r=0.30±0.07. Finally, for the grammatical coding of the lyrics, the average correlation was r=0.68±0.10. These are slightly, but not dramatically, lower than those obtained overall (compare Tables 3.29 and 3.24).

Table 3.24: Statistics concerning the correlations between the individual boundary profiles and the prediction of boundaries given the rules of Lerdahl and Jackendoff in Stage 4, Part 1.

Descriptive Statistics Across Individual Subjects					
	Mean±SD	Median	Mode	Min	Max
Rule 2b	0.668±0.096	0.705	0.736	0.313	0.772
Rule 3a	0.296±0.070	0.305	0.350	0.037	0.474
Grammar	0.511±0.122	0.520	0.520	0.153	0.718

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

To finish this analysis, the correlation between the individual boundary profile and each rule was assumed to be measure of utility (i.e., utilities for Rules 2b and 3a). That is, if an individual subject "used" a particular rule, then that subject's individual boundary profile would be correlated with the rule. Conversely, if an individual subject did not "use" the rule, then that subject's individual boundary profile would not be correlated with the rule. These correlations were called utilities. The utilities for different rules for different subjects were then compared.

If different subjects used different rules, then there would be no correlations between the utilities of the different rules. On the other hand, if subjects tended to use all rules in proportion (i.e., those who used Rule 2b also used Rule 3a), then there would be a correlation between the utilities. Finally, if some subjects used one rule while other subjects used a different rule, then the correlations between utilities would be negative (see Figure 3.20) The results indicated that those who used Rule 2b the most, also used Rule 3a the most (r=0.64 p<0.01). The same can be said for the grammatical coding of the lyrics (see Table 3.25).

Table 3.25: The correlations, across subjects, between the utilities of the different rules for predicting boundaries in the individual boundary profiles.

Correlations Between Measures			
	Rule 2b	Rule 3a	Grammar
Rule 2b	1.000	0.636**	0.411**
Rule 3a		1.000	0.472**

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.
* Significant (one tailed) at p<0.05
** Significant (one tailed) at p<0.01

To complement this analysis, all subjects were ranked in the degree to which they used the different rules (i.e., the aforementioned utilities). The utilities were then plotted in order of rank (one plot for Rule 2b and one plot for Rule 3a). Figure 3.17 plots the value for Rule 2b (in order of rank) and then overlays the corresponding values for Rule 3a and the grammatical coding of the lyrics. In a separate panel, Figure 3.17 also plots the value of Rule 3a (in order of rank) and then overlays the corresponding values for Rule 2b and the grammatical coding of lyrics. One can see the relationship between the two rules. Figure 3.18 presents the some simplified interpretations of these plots (using idealized data). By inspection of Figure 3.17, one can see that in general, utilities for the two rules change together. Those who used Rule 2b also used Rule 3a. That is, different subjects did not use different rules. Rather all subjects used the two rules in rough proportion.

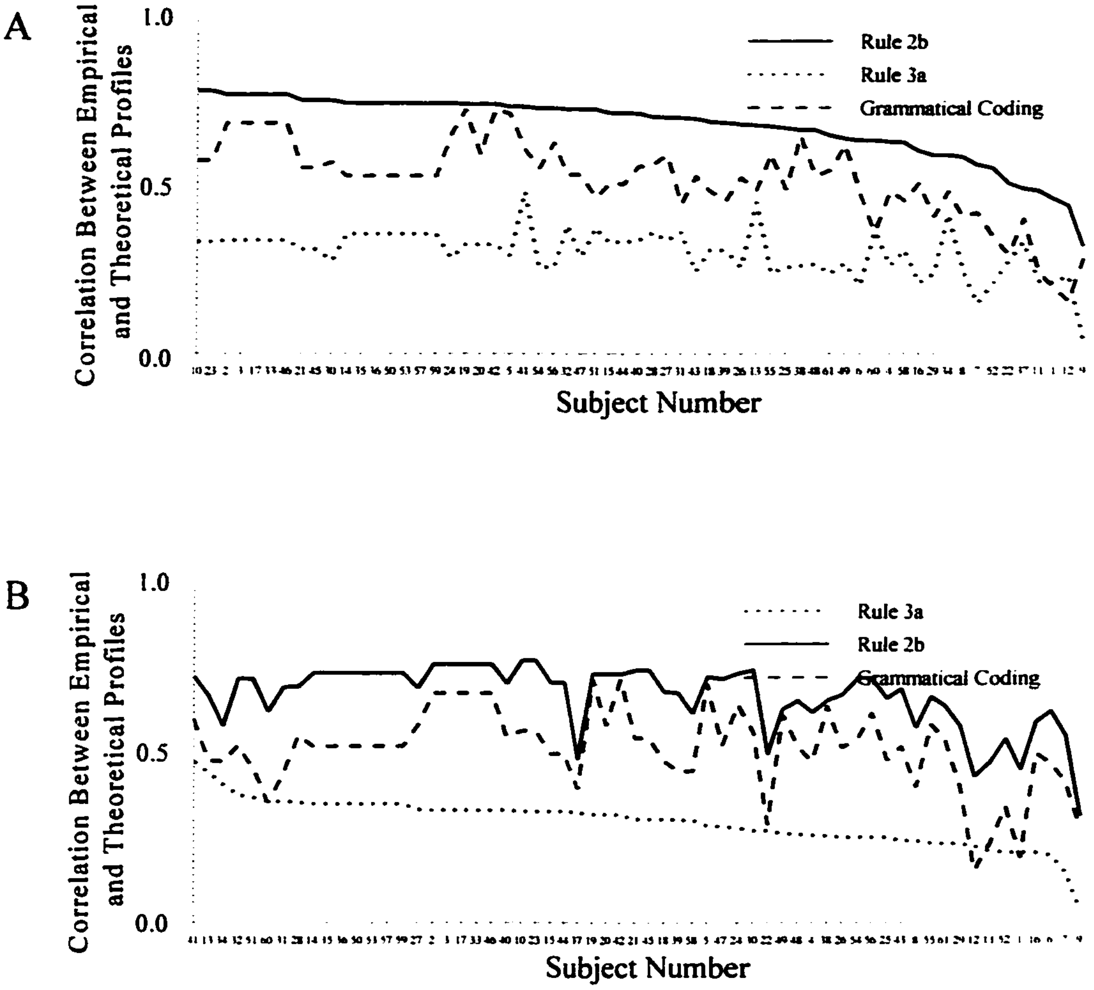


Figure 3.17: The relationship between individual empirical boundary profiles and the theoretical rules of Lerdahl and Jackendoff (1983). In A, subjects are ranked by Rule 2b and in B, subjects are ranked by Rule 3a. Note that until the last few subects, there is a gradual decrease in the utility each rule. Also note the the utility of Rule 2b is generally correlated with the utility of Rule 3a. Compare with Figure 3.18.

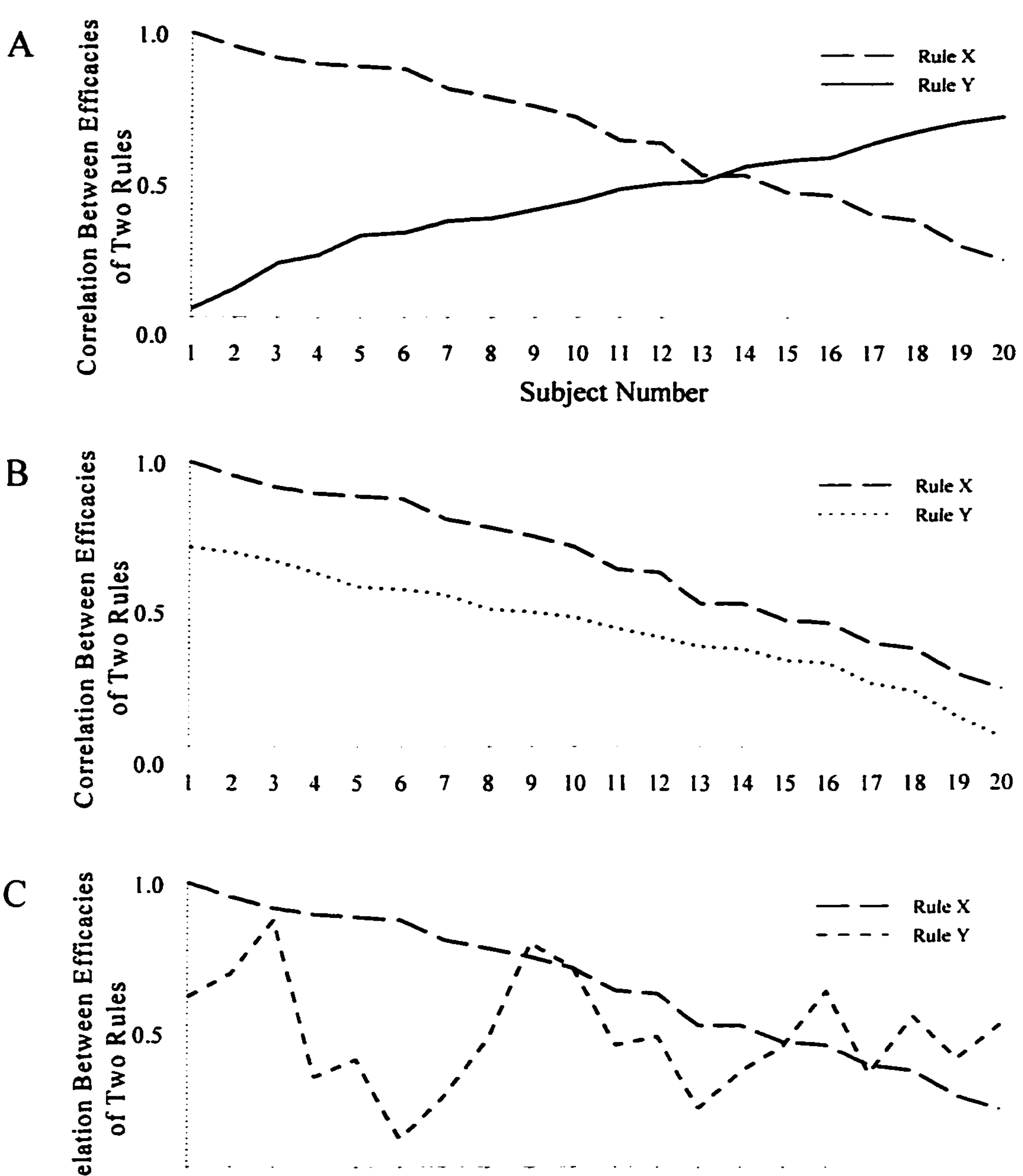


Figure 3.18: Interpretations for the correlations between the hypothetical utilities of different rules across subjects. In A, B and C, subjects are ranked by their use of Rule X. Note that in Panel A, those subjects who use Rule X the most are not the ones who used Rule Y the most. Panel B depicts the situation in which subjects use the both rules proportionally. Panel C depicts the situation in which theuse of Rule X is not related to the use of Rule Y.

In addition to demonstrating the relationship between rule utilities, this presentation allows one to see if there are groups of subjects who used rules differently. If some subjects used the rules while others did not, then a sharp discontinuity should be seen in the plot: No such discontinuity can be seen in Figure 3.17 (except perhaps in the last few subjects), implying that the degree to which individual subjects use each rule is a continuum.

Boundary Efficacy: Stage 4, Part 2

Given that all subjects placed boundaries at, more or less, the same locations, the analysis of test sequences is simplified. That is, even if the meaning of a subjectively located boundary is uncertain (i.e., is it the end of a unit, or the beginning of the next unit?), even if Lerdahl and Jackendoff's (1983) predictions are not completely verified, it seems that all subjects position boundaries using similar subjective criteria. Since the comparison of the empirically derived boundary profiles with the theoretically derived boundary profiles of Lerdahl and Jackendoff implies that subjects use boundaries to indicate the end of units, the interpretation of each of the conditions is tentatively defined (see Figure 3.12).

Each subject was presented with 16 test sequences that varied on two basic dimensions: seven Conditions (-1, 0, 1, 2, 3, 4 and 5) and two Types (Literal and Foil), but Type:Foil was further divided into Size, Location and Tonality. For Foil sequences, Location varied over 1, 2, 3 and 4 while Size varied over 2, 1, -1 and -2. Both Size and Location equal to 0 represented the Literal type. Tonality actually assessed the degree of change in tonality when the Foil sequence was created from its parent Literal. To create this variable, the Krumhansl key-finding algorithm was used to assess the key of the entire melody, as it was presented (i.e., for the melody of Stage 4, the key was d-major, as can be seen in Figure 3.16). To fashion the Foil, one note of a Literal sequences had been changed. This change affected the role of that single note within the tonal hierarchy (of the whole melody), and it was this change that was categorized (see Table 3.9). Assuming that the listener has a fair approximation of the true melody stored in memory, changing any note that had been a tonic to a non-diatonic note should be obvious; on the other hand, changing a non-diatonic note to a diatonic note might be less noticeable (or perhaps changing a diatonic, non-triad, note to another diatonic, non-triad, note). In essence, any change that resulted in a large deviation from the prevailing tonality should be more detectable than any change that did not result in a large change in tonality (remember that all changes are relatively small in terms of absolute pitch). To simplify matters, changes in tonality were categorized using only three levels of tonality: tonic triad notes, other diatonic (non-tonic triad) notes and non-diatonic notes. Changes between these three levels were tagged in a nine level scheme. The nine level scheme tries to crudely categorized the detectability of such changes using the work of Jones (1981, cf., Frankland & Cohen, 1996) as a guide.

The analysis was based on 16 test sequences from 61 subjects: essentially 976 test sequences. Ideally, for each subject there were to have been four test sequences in Conditions 0 and 5, and two in each of Conditions 1, 2, 3 and 4, implying totals of 244, 122, 122, 122, 122 and 244 test sequences in each of Conditions 0, 1, 2, 3, 4 and 5 respectively. From this, 26 test sequences were lost due to invalid responses (anticipatory responses or invalid key presses). In addition, of the 950 sequences that remained, 90 fell into Condition -1, which was undesirable, but useable, since the role of boundaries within these sequences could be delineated. A further 176 trials fell into Condition -2, which was undesirable and unusable, since the role of boundaries within these sequences is equivocal. The reason such a large number fell into Condition -2 can be seen in Figure 3.16: Subjects placed many boundaries in the initial part of the melody so the number of positions from which valid four note sequences could be extracted were reduced. This left 166, 100, 79, 111, 134 and 94 in Conditions 0, 1, 2, 3, 4 and 5 respectively. In addition, by design one half the sequences were to be Literals, and one half were to be Foils. Of the 774 valid and useful sequences, 434 were Literals (497 when including Condition -2), and 340 were Foils (453 when including Condition -2). The slight imbalance in the Types can be explained by the fact that on 46 occasions, the act of creating a Foil test sequence actually resulted in the creation of a Literal sequence from a different part of the melody. In general, the intended balancing worked.

The best insights are obtained from the raw data, but those insights are supported by rather complex analyses. In the initial analysis, Condition and Type were analyzed as categorical variables. In the secondary analyses, Condition, Type, Size, Location and Tonality, were all analyzed as categorical variables, despite any manifest interval scaling. That is, although a Location of 4 is twice as large as a Location of 2 and four times a Location of 1, the psychology of the importance of changes at these locations is not known: In addition, the meaning of a Location of 0 relative to the other locations is equivocal. Similarly, although a Size of -2 is twice that of a Size of -1, one cannot assume that they are perceived subjectively in this manner. From what is known about the psychophysical scaling (cf., Shepard, 1982) of the size of intervals in a musical context, it is almost certain that they are not (this is the rationale for the tonality analysis). Changes in Tonality, as tagged in the nine level scheme of Table 3.9, must be treated as categorical since there is not enough literature to state which should truly be more detectable within this context (i.e., within sequences extracted, then modified, then compared with the original melody).

Since the design used an unbalanced within-subjects approach, the only type of analysis that could be used was that of multiple regression, but treating all independent variables as categorical and essentially doing a within-subjects ANOVA from the perspective of multiple regression. Therefore, six vectors were used to code Condition, one vector was used to code Type, three vectors were used to code Size, three vectors were used to code Location and eight vectors were used to code Tonality: Note that, properly eight vectors would be required for tonality, but design constraints limited the necessary number to only four; the melody did not contain any non-diatonic notes so transitions from non-diatonic notes to any others were not possible and the size of changes were limited to ±2 semitones, so changes from triad tone to triad tone were not possible. These vectors represented the equivalent of individual contrasts within the more traditional ANOVA framework. Since orthogonality had been lost by design (unequal numbers per cell and the expected inequality between Foils and Literals), non-orthogonal vectors were used when they provided more information.

The Analysis of Condition, Type and Their Interactions

Six vectors (contrasts) for the effect of Condition were created (used in both the initial and secondary analyses) to compare (the contrasts are presented later with their analyses in Table 3.29):

Boundary on Note 2 versus Boundary on Note 3
Boundary on Note 2 or 3 versus Boundary on Note 4
Boundary on Note 2 or 3 versus Boundary on Note 1
Boundary on Note 0 or 5 versus Boundary on Note 1, 2, 3 or 4
Boundary before Sequence versus Boundary After Sequence
Boundary beside Sequence versus Boundary Far from Sequence

In this notation, Note 0 is a shorthand for boundaries on the note preceding the sequence and Note 5 is a shorthand for boundaries on the note following the sequence. Type was analyzed by a single vector, and the interaction of Condition and Type obviously required 6 vectors. In addition to these effects, there were 60 subject vectors (using dummy coding) and four covariate vectors (first and second order for the serial position within the melody -- with 0 defined as the middle of the melody; first and second order for the serial position of the test sequence within the test phase -- with 0 defined as the middle of the set of test sequences). Using a hierarchical analysis, subject effects were entered first, followed by the covariates, the main effects of Condition and Type, and finally the interaction. Four analyses were performed, each representing a slightly different approach to the problem of the unbalanced design. The first used the interaction of the subject vectors with the effects vectors as the error term. This approach is generally considered the most appropriate, and paradoxically, is considered unnecessary (cf., Cohen & Cohen, 1995; Keppel & Zeddick, 1992). The second analysis is the more common, simplified form of the first analysis, in which the residual is used as the error term for all effects (effectively, all subjectXeffects interactions are considered equivalent). The third analysis examined the incremental change in R^2 as the different terms were added to the equation. Finally, the fourth analysis tested the part correlations (i.e., the semipartials) for each effect. The fourth is often considered the most analogous to the balanced design (cf., Cohen & Cohen, 1994; Howell, 1995; Keppel & Zeddick, 1992).

The means of Condition (recognition memory as a function of the position of the boundary within the sequence), Type (Literals verse Foils) and their interaction is presented in Table 3.26.

Table 3.26: The means, standard deviations and cell counts for Condition, Type and ConditionXType in Stage 4, Part 2. Note 1 refers to the condition in which there was a boundary on the first note of the test sequence (similar for Note 2, 3 and 4). After refers to a test sequence that had a boundary on the note just following the sequence. Before refers to a test sequence that had a boundary just before the sequence. Far refers to a test sequence that did not contain and was not adjacent to boundaries.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.80±.41	.92±.28	.71±.46	.67±.48	.61±.49	.78±.42	.70±.47	.77±.42
	59	106	59	48	49	80	33	434
Foil	.68±.48	.63±.49	.78±.42	.48±.51	.66±.48	.91±.29	.79±.41	.72±.45
	31	60	41	31	62	54	61	340
Cond	.76±.43	.81±.39	.74±.44	.59±.49	.64±.48	.83±.34	.76±.43	.75±.44
	90	166	100	79	111	134	94	774

Essentially, the Condition is significant (see Tables 3.27 and 3.28 for the Omnibus analyses using the aforementioned approaches, and Table 3.29 for the individual contrasts on the two main effects and the interaction). However, note that the corresponded $\Delta R^2=0.03$ is not large implying that the effect is not particularly strong (about 3% of the variance; the part correlation [a.k.a.: semi-partial] is about the same). It must be remembered that test sequence construction is a quasi-experimental variable and as such, there must be many factors affecting retention (e.g., various structure that would affect test sequence complexity; various structures that affect the comparison of the test sequence to the "memory" of the melody). For example, the covariate Melody Position refers to the extraction point within the melody of the test sequence (a second order covariate looking for recency or primacy effects). It is significant, demonstrating that factors extraneous to the purpose of the thesis can have major impacts on test sequence retention. Unfortunately, melodic complexity is not yet well enough defined to allow it to be used as a covariate (conversely, this work contributes to the notion of melodic complexity). Boundary location is only one factor among many that might affect performance. As such, boundary location per se cannot be expected to account for a large part of the variation in performance.

Table 3.27: The regression analysis of Condition, Type and ConditionXType in Stage 4, Part 2, using both the specific error term, F_{sp} and the pooled error term. F_{po}.

Hierarchical Regression Analysis: Anova Model

Effect	SS	df	MS	F_{sp}	$p(F_{sp})$	F_{po}	$p(F_{po})$
Subjects	10.511	59					
Melody Pos.	5.154	2	2.577	8.453	<0.001	14.883	<0.001
Sequence Pos.	0.168	2	0.084	0.276	>0.250	0.485	>0.250
Condition	4.670	6	0.778	5.275	<0.001	4.495	<0.001
Type	0.201	1	0.201	0.885	>0.250	1.161	>0.250
ConditionXType	5.473	6	0.912	6.992	<0.001	5.268	<0.001
Pooled Residual	120.682	697	0.173				
CondXSubj	51.183	347	0.148				
TypeXSubj	13.403	59	0.227				
CondXTypeXSubj	24.390	187	0.130				
Residual	31.706	104	0.305				
Total	146.859	773					

Notes: Effects are listed in their order of entry into the analysis.
For F_{sp}, the error term is the interaction of the effect with subjects.
For F_{po}, the error term is the residual.

Note that the two analyses (Tables 3.27 and 3.28) produce slightly different dfs for some terms. The difference is due to the philosophy that the ANOVA approach uses all vectors (contrasts) that were included in the model, while the regression approach uses only those vectors (contrasts) that actually contributed to the term under consideration (i.e., that were actually entered into the equation). In fact, the ANOVA approach was constructed by hand from the regression approach.

Table 3.28: The regression analysis of Condition, Type and ConditionXType in Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.072	59					
Melody Pos.	0.035	2	13.986	0.000			
Sequence Pos.	0.001	2	0.456	0.634			
Condition	0.032	6	4.337	0.000	0.036	5.097	0.000
Type	0.001	1	1.119	0.291	0.000	0.267	0.605
ConditionXType	0.037	6	5.268	0.000	0.037	5.268	0.000
Pooled Residual	0.822	694					
CondXSubj	0.349	347					
TypeXSubj	0.091	59					
CondXTypeXSubj	0.166	187					
Residual	0.216	101					
Total	1.000	770					

Notes: Effects are listed in the order in which they enter the equation.
For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.
For R^2_{part}, the error term includes all effects in the full model.

Observe that Condition 4 produced the highest level of performance, which would be consistent with the idea that boundaries indicate the end of a unit (therefore, the four notes of the test sequences would be extracted from a single unit and those four notes represent the tail end of the unit; cf., Figure 3.12). Condition 0 produced high performance as well, which would be consistent with boundaries indicating the closure of a unit (therefore the four notes of the sequence would be extracted from a single unit that represents the beginning of the unit). These conclusions are supported by the tests on Contrasts 2, 4 and 5 (see Table 3.29), and are consistent with intuitions based on Figure 3.16. Conditions 1, 2 and 3 produced the lowest performance consistent with the notion that each of these situations represents two units: Contrast analysis (Contrasts 1 and 3) indicated that these conditions (1, 2 and 3) are equivalent. Contrast 6 indicated that the average of the before and after conditions was not different from the far condition (i.e., no difference between all condition in which the boundary was not in the test sequence).

Given Table 3.26, it is clear that the two types of test sequences (Literals & Foils) were not significantly different (this can be seen in both the ΔR^2 and the part correlation). Overall subjects are equally accurate when deciding that a test sequence was in a familiar melody as they are deciding that a test sequence was not in that same melody.

Table 3.29: The vector (contrast) analysis on the effect of Condition, Type and ConditionXType in Stage 4, Part 2. Both the vectors (contrasts) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are listed. The *F* values represent tests of the part correlations.

Contrast	Far	Before	Note 1	Note 2	Note 3	Note 4	After
Means	.76±.43	.81±.39	.74±.44	.59±.49	.64±.48	.83±.38	.76±.43
Cond Ψ_1	0	0	0	1	-1	0	0
Cond Ψ_2	0	0	0	1	1	-2	0
Cond Ψ_3	0	0	-2	1	1	0	0
Cond Ψ_4	0	-1	1	1	1	1	-1
Cond Ψ_5	0	1	0	0	0	0	-1
Cond Ψ_6	-2	1	0	0	0	0	1

Contrast	Literal	Foil
Means	.77±.42	.72±.45
Type Ψ_1	1	-1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Cond Ψ_1	-0.001	-0.038	-0.041	1.200	0.274
Cond Ψ_2	-0.141	-0.156	-0.170	20.701	0.000
Cond Ψ_3	-0.077	0.045	0.049	1.706	0.192
Cond Ψ_4	0.082	0.081	0.089	5.591	0.018
Cond Ψ_5	0.054	-0.063	-0.070	3.407	0.065
Cond Ψ_6	0.060	-0.046	-0.051	1.822	0.178
Type Ψ_1	0.057	-0.018	-0.020	0.267	0.605
TypeΨ_1 X CondΨ_1	0.034	0.034	0.037	0.952	0.329
TypeΨ_1 X CondΨ_2	0.048	0.039	0.042	1.285	0.257
TypeΨ_1 X CondΨ_3	0.034	-0.018	-0.020	0.267	0.605
TypeΨ_1 X CondΨ_4	0.086	0.023	0.026	0.456	0.500
TypeΨ_1 X CondΨ_5	0.150	0.131	0.142	14.349	0.000
TypeΨ_1 X CondΨ_6	0.103	0.053	0.058	2.379	0.123

Notes: The interaction contrasts are each Condition contrast multiplied by the Type contrast.

Tables 3.27, 3.28 and 3.29 show the significant interaction between Condition and Type. The indication is that the pattern across Conditions -1, 0 and 5 changes as a function of Type: for Literals these means are 0.70±0.40 versus 0.92±0.28, and 0.69±0.47 respectively, while for the Foils these means are 0.68±0.48 versus 0.63±0.49 and 0.79±0.41 (see interaction Contrasts 5 and 6). The effect size of the interaction is about same magnitude as that of Condition (ΔR^2=0.04; similar for the part correlation).

Mean differences between corresponding conditions are presented in Table 3.30. Some conditions seem to elicit a tendency to "yes" answers (it was in the melody), while others seem to have a tendency to "no". That is, when the boundary was before the sequence (hence the sequence was the opening of a unit), subjects tended to say "yes" thereby getting Literals correct and Foils incorrect. However, when the boundary was on Note 4, the response seems to be weighted to a "no" thereby getting the Foils correct and the Literals incorrect. The effect is not very pronounced (the correlation between corresponding Literal and Foil conditions is only r=0.10).

Table 3.30: Differences between corresponding Literal and Foil Type Conditions.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.80±.41	.92±.28	.71±.46	.67±.48	.61±.49	.78±.42	.70±.47	.77±.42
	59	106	59	48	49	80	33	434
Foil	.68±.48	.63±.49	.78±.42	.48±.51	.66±.48	.91±.29	.79±.41	.72±.45
	31	60	41	31	62	54	61	340
Cond	.12±.43	.29±.39	-.07±.44	.19±.49	-.05±.48	-.13±.34	-.09±.43	.05±.44

To check the integrity of the analysis, raw residuals, studentized residuals, Leverage, Cooks distance and Mahalanonbis' distance were examined (cf., Cohen & Cohen, 1994; Howell, 1992; Stevens, 1996). No outliers were detected. It should be noted that the covariate for melody position was significant while the covariate for sequence position was not. That is, performance was slightly higher for extracts taken from the end of the melody (the covariate examined second order relationships between melody position and performance collapsed over Literals and Foils). Performance did not differ between those extracts that were presented immediately after the conclusion of the melody and those that were presented long after the conclusion of the melody, suggesting no decay of the long term representation of the melody.

The Analyses of Location, Size and Tonality and Their Interactions

Three vectors (contrasts) were created to analysis the effect of Location (the contrasts are presented later with their analyses in Table 3.34):

Change at Note 1 versus Change at Note 2
Change at Note 3 versus Change at Note 4
Change at Note 1 or 2 versus Change at Note 3 or 4

The idea is to compare note changes on the two early positions, note changes on the two later positions, and finally the changes on the early positions to the changes on the late positions. It should be noted that the original contrast for Type should separate Location 0 from the remaining conditions. That is, Location equal to 0 was Type:Literal and Locations equal to 1, 2, 3 or 4 were variations of Type:Foil.

In a similar manner, three vectors (contrasts) were created to analyze the effect of Size (these contrasts are presented with their analyses in Table 3.38):

Change of Size -2 versus Change of Size 2
Change of Size -1 versus Change of Size 1
Change of Size -2 or 2 versus Change of Size -1 or 1

If only the absolute magnitude of the change matters when creating a Foil sequence (for subsequent performance), then neither the first nor the second contrast should be significant, while the third should. One might have such a prediction for the subjects who are not sensitive to diatonicism (or tonality). Such subjects might be sensitive to proximity.. Again, it should be noted that the original contrast for Type will separate the Size 0 condition from the remaining.

Finally, the effect of Tonality was coded by eight vectors (these contrasts are presented with their analyses in Table 3.42). In these vectors, the transition from a tonic triad note to a non-diatonic note was considered a baseline against which all other transitions were compared (i.e., the set corresponds to that of Dunnett contrast set [cf., Howell, 1992], considering Level 1 to Level 3 transitions as the control). It was assumed that such a transition should be the most obvious to detect (therefore, the highest level of performance) since such a transition represents moving from the most stable notes to the least stable notes. In addition, because the melody did not contain any accidentals, any accidentals in a test sequence should be an obvious marker as a Foil. However, since there are a number of possible effects and associated explanations, this simple coding will allow for multiple interpretations. The vectors (contrasts) were designed to test the following transitions:

Triad to Non-Diatonic versus Diatonic to Non-Diatonic
Triad to Non-Diatonic versus Triad to Diatonic
Triad to Non-Diatonic versus Triad to Triad
Triad to Non-Diatonic versus Diatonic to Diatonic
Triad to Non-Diatonic versus Non-Diatonic to Non-Diatonic
Triad to Non-Diatonic versus Diatonic to Triad
Triad to Non-Diatonic versus Non-Diatonic to Triad
Triad to Non-Diatonic versus Non-Diatonic to Diatonic

Note that of this set, only the first, second, fourth and sixth contrasts (in italics) actually apply because the melody did not contain any non-diatonic notes (hence, there could not be any transitions from non-diatonic notes to other notes). In addition, because the size of the change was limited to ±2 semitones, it was not possible for a transition to move from triad tone to triad tone. The original contrast for Type separates the Literal conditions from the others.

Each effect, Size, Location and Tonality, is initially presented in isolation. There are two reasons for this. Firstly, there was no attempt to balance the number of presentations of each of these effects (Location and Size were balanced at the time of presentation by a random number generator, but the numbers were not explicitly set a priori; Tonality was not balanced at all), so there is considerable overlap between the various vectors: Hence order of entry might hide some relationships. By presenting each in isolation, the effect of each can be known. Secondly, each of these effects must be treated as an addition to the previous analysis of Condition, Type and ConditionXType. By presenting each in isolation, the ConditionXeffect interactions can be clearly seen. Note that each of Location, Size, and Tonality will not interact with Type (in the design sense) since each of Location, Size and Tonality is merely a breakdown of the Type:Foil condition. In these secondary analyses, the subject vectors were entered first, the covariates second, Condition third, Type fourth, ConditionXType fifth, each effect and finally each interaction of Condition with the effect. Because the ANOVA analyses are essentially redundant with the regression analysis, only the regression analyses are presented[35]. The regression analyses are presented because they contain more information (i.e., the part analysis).

For Location, means are presented in Table 3.31 To facilitate the comparison, Literals are included in the table for comparison.

Table 3.31: The means, standard deviations and cell counts for Condition, Location (Literals correspond to Location equal 0) and ConditionXLocation, in Stage 4, Part 2.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Loc
Liter	.80±.41	.92±.28	.71±.46	.68±.47	.61±.49	.78±.42	.70±.47	.77±.42
	59	106	59	48	49	80	33	434
Loc 1	.71±.49	.15±.37	.71±.46	—	.85±.37	1.0±.00	.23±.44	.57±.50
	7	20	28	0	20	6	13	94
Loc 2	.67±.58	.83±.38	.92±.28	1.0±.00	.84±.37	1.0±.00	.88±.35	.88±.33
	3	24	13	2	19	12	8	81
Loc 3	.67±.58	1.0±.00	—	.47±.51	.29±.47	.50±.71	1.0±.00	.48±.51
	3	1	0	19	14	2	3	42
Loc 4	.67±.49	.93±.26	—	.40±.52	.44±.53	.88±.33	.95±.23	.80±.40
	18	15	0	10	9	34	37	123
Cond	.76±.43	.81±.39	.74±.44	.59±.49	.64±.48	.83±.38	.76±.43	.75±.44
	90	166	100	79	111	134	94	774

Location (considering the three vectors that coded for differences between Locations 1, 2,

---

[35] The SS_{tot} are included for those who wish to recreate the ANOVA analyses. To create the SS for each term, the ΔR^2 for each term is multiplied by the SS_{tot}, and the df are adjusted to represent the entire set of vectors (contrasts). The usual computations of the ANOVA are then completed. This was done for each analyses; there were subtle differences, but none that affected significance or interpretation.

3 and 4; i.e., not including Literals) did produce significant differences. Table 3.32 provides the omnibus analysis of Location and the interaction between Condition and Location. The magnitude of the Location effect (ΔR^2=0.04) is comparable to that of Condition (though the part correlation is reduced) while the magnitude of the ConditionXLocation interaction is larger (ΔR^2=0.06; same for the part).

Table 3.33: The regression analysis of Condition, Type and ConditionXType with the addition of Size (of change) and ConditionXSize, in Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.072	59					
MelodyPos.	0.035	2	13.986	0.000			
Sequence Pos.	0.001	2	0.456	0.634			
Condition	0.032	6	4.337	0.000	0.021	3.277	0.004
Type	0.001	1	1.119	0.291	0.000	0.042	0.838
ConditionXType	0.037	6	5.268	0.000	0.009	1.516	0.170
Location	0.037	3	10.985	0.000	0.019	6.003	0.001
CondXLocation	0.063	15	3.925	0.000	0.063	3.925	0.000
Pooled Residual	0.722	679					
Total	1.000	773					

SS_{tot} = 146.859

Notes: Effects are listed in the order in which they enter the equation.

For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.

For R^2_{part}, the error term includes all effects in the full model.

As expected, changes on the first note interfered with performance, but strangely, changes on the third note were more detrimental. The interaction of condition and location was significant. By inspection of Table 3.31, one can see that changes on the first note annihilated performance when the boundary preceded or followed the test sequence but changes on the first note did not affect performance when the boundary was contained within the sequence. Changes on the second note when the second note was a boundary were obvious to subjects while changes on first, third and fourth notes when they were boundaries were not as obvious. However, it is important to realize that since Location was not explicitly controlled, some cells in the interaction had no test sequences. Their means are not necessarily the 0 that appears in the table. Hence, the first curiosity regarding the changes on the second note when the second note was also the boundary might be due to chance (n=2). Contrast analysis (Table 3.33; see LocΨ_3) indicated a large effect of the first two positions (changes on Notes 1 or 2) in comparison to the last two positions (changes on Notes 3 and 4), but as can be seen by the means, this analysis (planned) hides the real pattern: Lower performance for changes on Notes 1 and 3. Of the significant interaction contrasts, LocΨ_1 X CondΨ_4 implies that the difference between changes on the first note are changes on the second note varies according to whether of not the boundary is within the sequence. Similarly, LocΨ_3 X CondΨ_4 implies the same thing for Locations 1 and 2 combined verse Locations 3 and 4 combined.

Table 3.33: The vector (contrast) analysis on the effect of Location in Stage 4, Part 2. Both the vectors (contrasts; Condition vectors have been presented previously) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are listed. The *F* values represent tests of the part correlations.

Contrast	Note 1	Note 2	Note 3	Note 4
Means	.57±.50	.88±.33	.48±.41	.80±.40
Loc Ψ_1	1	-1	0	0
Loc Ψ_2	0	0	1	-1
Loc Ψ_3	1	1	-1	-1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Loc Ψ_1	-0.167	-0.132	-0.153	16.316	0.000
Loc Ψ_2	-0.123	-0.030	-0.036	0.864	0.353
Loc Ψ_3	-0.007	0.004	0.005	0.017	0.892
LocΨ_1 X CondΨ_2	0.022	-0.012	-0.014	0.127	0.721
LocΨ_1 X CondΨ_3	0.035	0.032	0.038	0.975	0.324
LocΨ_1 X CondΨ_4	-0.108	-0.115	-0.134	12.497	0.000
LocΨ_1 X CondΨ_5	-0.064	0.004	0.005	0.014	0.905
LocΨ_1 X CondΨ_6	-0.175	0.045	0.053	1.890	0.170
LocΨ_2 X CondΨ_1	0.023	0.039	0.045	1.407	0.236
LocΨ_2 X CondΨ_3	-0.062	0.024	0.029	0.562	0.454
LocΨ_2 X CondΨ_4	0.009	-0.001	-0.001	0.001	0.978
LocΨ_2 X CondΨ_5	0.045	-0.005	-0.005	0.020	0.887
LocΨ_2 X CondΨ_6	-0.049	0.012	0.014	0.030	0.718
LocΨ_3 X CondΨ_1	-0.034	-0.000	-0.001	0.000	0.989
LocΨ_3 X CondΨ_2	0.100	0.007	0.008	0.044	0.834
LocΨ_3 X CondΨ_4	-0.238	-0.099	-0.116	9.304	0.002
LocΨ_3 X CondΨ_5	0.007	0.002	0.003	0.006	0.939
LocΨ_3 X CondΨ_6	-0.128	0.006	0.007	0.037	0.848

Notes: The interaction contrasts are each Condition contrast multiplied by each Location contrast.

The means for Size are presented in Table 3.34. To facilitate comparison, the means for Literals are included in the table.

Table 3.34: The means, standard deviations and cell counts for Condition, Size (Literals correspond to Size equal 0) and ConditionXSize, in Stage 4, Part 2.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Loc
Liter	.80±.41	.92±.28	.71±.46	.68±.47	.61±.49	.78±.42	.70±.47	.77±.42
	59	106	59	48	49	80	33	434
Siz -2	.50±.71	.67±.58	—	—	.84±.37	.87±.35	—	.82±.39
	2	3	0	0	19	15	0	39
Siz -1	.62±.50	.54±.50	.84±.37	—	—	.90±.30	.88±.35	.70±.46
	21	41	19	0	0	21	8	110
Siz 1	1.0±.00	.68±.57	.73±.46	.52±.51	.62±.49	.86±.38	.96±.20	.72±.45
	2	3	22	21	34	7	26	115
Siz 2	.83±.41	.92±.28	—	.40±.52	.44±.53	1.0±.00	.59±.50	.68±.47
	6	13	0	10	9	11	27	76
Cond	.76±.43	.81±.39	.74±.44	.59±.49	.64±.48	.83±.38	.76±.43	.75±.44
	90	166	100	79	111	134	94	774

The omnibus analysis of Size and the interaction of Size with Condition is presented in Table 3.35 while Table 3.36 provides the analyses of the individual contrasts. Note that Size was not significant by any analysis which is consistent with the range shown in Table 3.34. ConditionXSize was significant and comparable in magnitude to the effect of Condition alone.

Table 3.35: The regression analysis of Condition, Type and ConditionXType with the addition of Size (of change) and ConditionXSize, in Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations. R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{\text{part}})$	$p(R^2_{\text{part}})$
Subjects	0.072	59					
MelodyPos.	0.035	2	13.986	0.000			
Sequence Pos.	0.001	2	0.456	0.634			
Condition	0.032	6	4.337	0.000	0.024	3.527	0.002
Type	0.001	1	1.119	0.291	0.001	0.531	0.466
ConditionXType	0.037	6	5.268	0.000	0.018	2.602	0.017
Size	0.003	3	0.814	0.487	0.001	0.293	0.869
ConditionXSize	0.034	12	2.488	0.003	0.034	2.488	0.003
Pooled Residual	0.784	682					
Total	1.000	773					

$SS_{\text{tot}} = 146.859$

Notes: Effects are listed in the order in which they enter the equation.

For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.

For R^2_{part}, the error term includes all effects in the full model.

Table 3.36: The vector (contrast) analysis on the effect of Size, in Stage 4, Part 2. Both the vectors (contrasts; Condition vectors have been presented previously) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are provided. The *F* values represent tests of the part correlations.

Contrast	2	1	-1	-2
Means	.82±.39	.70±.46	.72±.45	.68±.47
Size Ψ_1	1	0	0	-1
Size Ψ_2	0	1	-1	0
Size Ψ_3	1	-1	-1	1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Size Ψ_1	0.059	0.020	0.022	0.342	0.559
Size Ψ_2	-0.013	0.020	0.023	0.362	0.548
Size Ψ_3	0.028	-0.008	-0.007	0.052	0.821
SizeΨ_1 X CondΨ_2	0.069	0.061	0.069	3.258	0.072
SizeΨ_1 X CondΨ_4	-0.044	-0.010	-0.011	0.082	0.775
SizeΨ_1 X CondΨ_6	-0.004	-0.043	-0.049	1.636	0.201
SizeΨ_2 X CondΨ_2	0.025	0.028	0.032	0.678	0.411
SizeΨ_2 X CondΨ_3	0.025	0.000	0.000	0.000	0.994
SizeΨ_2 X CondΨ_5	-0.036	-0.011	-0.012	0.101	0.751
SizeΨ_2 X CondΨ_6	-0.123	-0.026	-0.030	0.606	0.437
SizeΨ_3 X CondΨ_1	-0.019	-0.008	-0.009	0.051	0.821
SizeΨ_3 X CondΨ_2	0.017	-0.023	-0.026	0.475	0.491
SizeΨ_3 X CondΨ_4	-0.017	0.005	0.006	0.024	0.876
SizeΨ_3 X CondΨ_5	0.163	0.119	0.133	12.239	0.001
SizeΨ_3 X CondΨ_6	0.024	-0.012	-0.013	0.120	0.730

Notes: The interaction contrasts are each Condition contrast multiplied by each Size contrast.

Note that Sizes -2 and 2 (i.e., the greatest change) did not generally produce the best performance (regardless of significance). This implies that size is not an important factor in performance. The interaction of Condition and Size was significant, which is likely due to the change in pattern for the before and after conditions (see the second last interaction contrast), but this should be viewed with some caution, given the number of empty cells (see Table 3.34).

For Tonality, the means are presented in Table 3.37. To facilitate comparison, the means for Literals are also presented in the table.

Table 3.37: The means, standard deviations and cell counts for Condition, Tonality (Literals correspond, arbitrarily, to Tonality equals 10) and ConditionXTonality, Stage 4, Part 2.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Loc
Liter	.80±.41	.92±.28	.71±.46	.68±.47	.61±.49	.78±.42	.70±.47	.77±.42
	59	106	59	48	49	80	33	434
Ton 1	.89±.33	.40±.50	.78±.42	.80±.45	.89±.32	.89±.32	.95±.21	.76±.43
	9	32	40	5	18	18	22	144
Ton 2	.50±.71	1.0±.00	—	—	1.0±.00	1.0±.00	.56±.53	.85±.36
	2	5	0	0	1	17	9	34
Ton 3	.78±.44	.91±.29	1.0±.00	.42±.50	.56±.50	.82±.39	.91±.29	.70±.36
	9	22	1	26	41	17	23	139
Ton 5	—	—	—	—	—	1.0±.00	.00±.00	.14±.38
	0	0	0		0	1	6	7
Ton 7	.45±.52	.00±.00	—	—	.50±.71	1.0±.00	1.0±.00	.50±.52
	11	1	0	0	2	1	1	16
Cond	.76±.43	.81±.39	.74±.44	.59±.49	.64±.48	.83±.38	.76±.43	.75±.44
	90	166	100	79	111	134	94	774

Notes: Ton 1 = Triad to non-diatonic
Ton 2 = Diatonic, non-triad to non-diatonic
Ton 3 = Triad to diatonic, non-triad
Ton 5 = Diatonic, non-triad to diatonic, non-triad
Ton 7 = Diatonic, non-triad to triad
: Only those tonality conditions that could exist, by design, are included.

Tonality (considering the four vectors that coded for the different changes in tonality; i.e., not including Literals) did produce significant differences that generally followed predictions, with an effect size comparable to that of Condition or Location (ΔR^2=0.03). Tonality 1 (changes from a triad to a non-diatonic note) and Tonality 2 (changes from a diatonic [non-triad] to a non-diatonic note) produced the highest performance. Tonality 5 (changes from a diatonic [non-triad] to a diatonic [non-triad] note) produced the lowest performance. Tonality 3 (changes from a triad to a diatonic [non-triad] note) Tonality 7 (changes from diatonic [non-triad] to triad) were intermediate. The omnibus effects are presented in Table 3.38.

Table 3.38: The regression analysis of Condition, Type and ConditionXType with the addition of Tonality (of change) and ConditionXTonality, in Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.072	59					
Melody Pos.	0.035	2	13.986	0.000			
Sequence Pos.	0.001	2	0.456	0.634			
Condition	0.032	6	4.337	0.000	0.021	3.277	0.004
Type	0.001	1	1.119	0.291	0.000	0.042	0.838
ConditionXType	0.037	6	5.268	0.000	0.009	1.516	0.170
Tonality	0.035	4	7.623	0.000	0.019	6.003	0.001
CondXTonality	0.053	15	3.276	0.000	0.063	3.925	0.000
Pooled Residual	0.734	678					
Total	1.000	773					

$SS_{tot} = 146.859$

Notes: Effects are listed in the order in which they enter the equation.

For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.

For R^2_{part}, the error term includes all effects in the full model.

Note that changes that resulted in a non-diatonic note within the sequence produced generally higher performance. Also note that although there are, in principle, nine possible levels of Tonality (see Table 3.9), only five levels could exist in this task. Only four vectors were needed to code for Tonality (the full set of 8 is depicted in Table 3.42: only contrasts 1, 2, 4 and 6 were actually needed). The contrast analysis is presented in Table 3.39.

Table 3.39: The vector (contrast) analysis on the effect of Tonality, in Stage 4, Part 2. Both the vectors (contrasts; Condition vectors have been presented previously) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are provided. The *F* values represent tests of the part correlations.

Contrast	1	2	3	4	5	6	7	8	9
Means	.76±.43	.85±.36	.70±.46	—	.14±.38	—	.50±.52	—	—
Ton Ψ_1	1	-1	0	0	0	0	0	0	0
Ton Ψ_2	1	0	-1	0	0	0	0	0	0
Ton Ψ_3	1	0	0	-1	0	0	0	0	0
Ton Ψ_4	1	0	0	0	-1	0	0	0	0
Ton Ψ_5	1	0	0	0	0	-1	0	0	0
Ton Ψ_6	1	0	0	0	0	0	-1	0	0
Ton Ψ_7	1	0	0	0	0	0	0	-1	0
Ton Ψ_8	1	0	0	0	0	0	0	0	-1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Ton Ψ_1	-0.013	-0.079	-0.091	5.710	0.017
Ton Ψ_2	0.041	-0.053	-0.062	2.643	0.105
Ton Ψ_4	0.043	0.138	0.159	17.666	0.000
Ton Ψ_6	0.039	0.020	0.023	0.370	0.543
TonΨ_1 X CondΨ_2	0.039	-0.031	-0.036	0.882	0.348
TonΨ_1 X CondΨ_4	-0.050	-0.004	-0.005	0.015	0.903
TonΨ_1 X CondΨ_5	-0.185	-0.096	-0.112	8.542	0.004
TonΨ_1 X CondΨ_6	-0.018	0.038	0.044	1.300	0.255
TonΨ_2 X CondΨ_1	-0.013	-0.017	-0.019	0.258	0.612
TonΨ_2 X CondΨ_2	0.089	-0.045	-0.052	1.872	0.172
TonΨ_2 X CondΨ_3	0.091	0.019	0.022	0.322	0.571
TonΨ_2 X CondΨ_4	-0.176	-0.036	-0.042	1.186	0.277
TonΨ_2 X CondΨ_5	-0.126	-0.044	-0.051	1.755	0.186
TonΨ_2 X CondΨ_6	-0.074	0.009	0.010	0.068	0.794
TonΨ_4 X CondΨ_2	-0.015	0.070	0.082	4.541	0.034
TonΨ_6 X CondΨ_1	-0.046	0.013	0.015	0.158	0.691
TonΨ_6 X CondΨ_2	-0.011	0.012	0.014	0.136	0.713
TonΨ_6 X CondΨ_5	-0.160	0.043	0.050	1.705	0.192
TonΨ_6 X CondΨ_6	0.023	-0.015	-0.017	0.201	0.654

Notes: The interaction contrasts are each Condition contrast multiplied by each Location contrast. Since the melody did not contain any accidentals, Contrasts 5, 7 and 8 could not exist because Tonality 6, 8 and 9 did not exist. Since changes were limited to ±2 semitones, Contrast 3 could not exist.

The interaction ConditionXTonality was also significant (see Table 3.38) with an effect about the same magnitude as the ConditionXLocation interaction (ΔR^2=0.05; part about the same). As with size, performance for boundaries before was the reverse of boundaries after (the third interaction contrast). Again, this interaction should be viewed with some caution since there is a large disparity in cell counts.

In addition to the previous individual analyses, an overall analysis (Table 3.40) was conducted in which the subject vectors were entered first, the covariates second, Condition third, Type fourth, ConditionXType fifth, Location sixth, Size seventh, Tonality eighth (since Location, Size and Tonality represent a finer breakdown of one level of Type, and as such are logically orthogonal to Type and the ConditionXType interaction), and finally each interaction of Condition with Location, Size and Tonality. Three-way interactions were not analysed due to an overabundance of empty cells. Tonality was entered last because it was not controlled to the same extent as Location or Size. Size was entered just before Tonality (rather than before Location) because Size and Tonality represented a competition of sorts (Size and Tonality are both aspects of pitch or frequency, while Location is separate from the two). The same contrasts (vectors) for Condition, Type, ConditionXType, Location, Size, Tonality, ConditionXLocation, ConditionXSize and ConditionXTonality were used.

This analysis would be expected to have its greatest impact on the magnitudes of the part correlations, since part correlations are effects corrected for all other terms in the analysis. However, the part analysis is not included since it did not add any new insights. Generally, as would be expected, compared to previous analysis, there was a loss in the number of significant effects (significant reflects many things, including the magnitude of the part correlation and the df for the analysis as a whole).

Table 3.40: The regression analysis of Condition, Type, ConditionXType, Location, Size and Tonality and the interactions of Condition with Location, Size and Tonality, in Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.072	59					
Melody Pos.	0.035	2	13.986	0.000			
Sequence Pos.	0.001	2	0.456	0.634			
Condition	0.032	6	4.337	0.000	0.016	2.608	0.017
Type	0.001	1	1.119	0.291	0.000	0.053	0.818
ConditionXType	0.037	6	5.268	0.000	0.016	2.554	0.019
Location	0.037	3	10.985	0.000	0.007	2.252	0.081
Size	0.023	3	6.985	0.000	0.002	0.576	0.631
Tonality	0.017	4	3.386	0.004	0.004	1.000	0.407
CondXLocation	0.046	15	2.975	0.000	0.036	2.316	0.003
ConditionXSize	0.012	8	1.483	0.160	0.015	1.825	0.070
CondXTonality	0.012	13	0.875	0.579	0.012	0.875	0.572
Pooled Residual	0.674	651					
Total	1.000	773					

$SS_{tot} = 146.859$

Notes: The effects are listed in the order in which they entered the equation.

In this overall analysis, Condition, ConditionXType and ConditionXLocation emerged as the major factors, with a nod to Location and the ConditionXSize interaction. Together, the various terms accounted for about 23% of the variance (10% is accounted for by differences between subjects and the position in the melody from which the extract was taken). This is most clearly seen in the analysis of the part correlation for the effects (i.e., R^2_{part}): This analysis is thought to be the most comparable to that of the balanced ANOVA.

When significances were tested hierarchically, Condition, ConditionXType, Location, Size, Tonality and ConditionXLocation all increase the predictability of the equation. Note that Tonality is significant despite being the last main effect added -- despite being entered after Size. In addition, it is the contrasts coding for Size that get dropped in the final analysis -- not the contrasts coding for Tonality, implying that Tonality has more to add than Size.

To present the same results from a perspective that highlights the link between the empirical boundaries and the test sequences, Figure 3.19 presents the melody, the empirically determined boundaries and the results for test sequences as a function of the extraction point of the test sequence from the melody.

Before averaging the performance of different subjects, the mean of each subject was subtracted from each score of that subject (equivalent in function to the within-subjects analysis: the subjects terms of the previous analyses). This adjustment removes consideration of individual differences in mean level of responding (this is important because the subjects effect was fairly substantial in all previous analysis: i.e., different subjects had different mean performances).

This presentation captures something similar to the analysis of Condition (i.e., Tables 3.26 through 3.29). One can see the performance of test sequences relative to nearby boundaries. This is what the different conditions coded for (the relative position of the boundary within or near the test sequence). However, the presentation is not directly related to the analysis of Condition. Firstly, the previously excluded Condition -2 (2 or more boundaries in or around the sequence) is included. Secondly, this presentation fails to consider individually placed empirical boundaries as they relate to individually tailored test sequences. Hence, a test sequence at Note 16 of the melody might represent Condition 0 to one subject (boundary in front of sequence) and Condition -1 to another subject (boundary far from test sequence). As such, test sequence performance as a function of melody position may be collapsing different conditions. In other words, this analysis is only useful to the extent that the average profile represents any single subject. Given that the average profile is a fair representation of the individual subject, this presentation helps to elucidate various results. This presentation also demonstrates the effect of the covariate Melody Position.

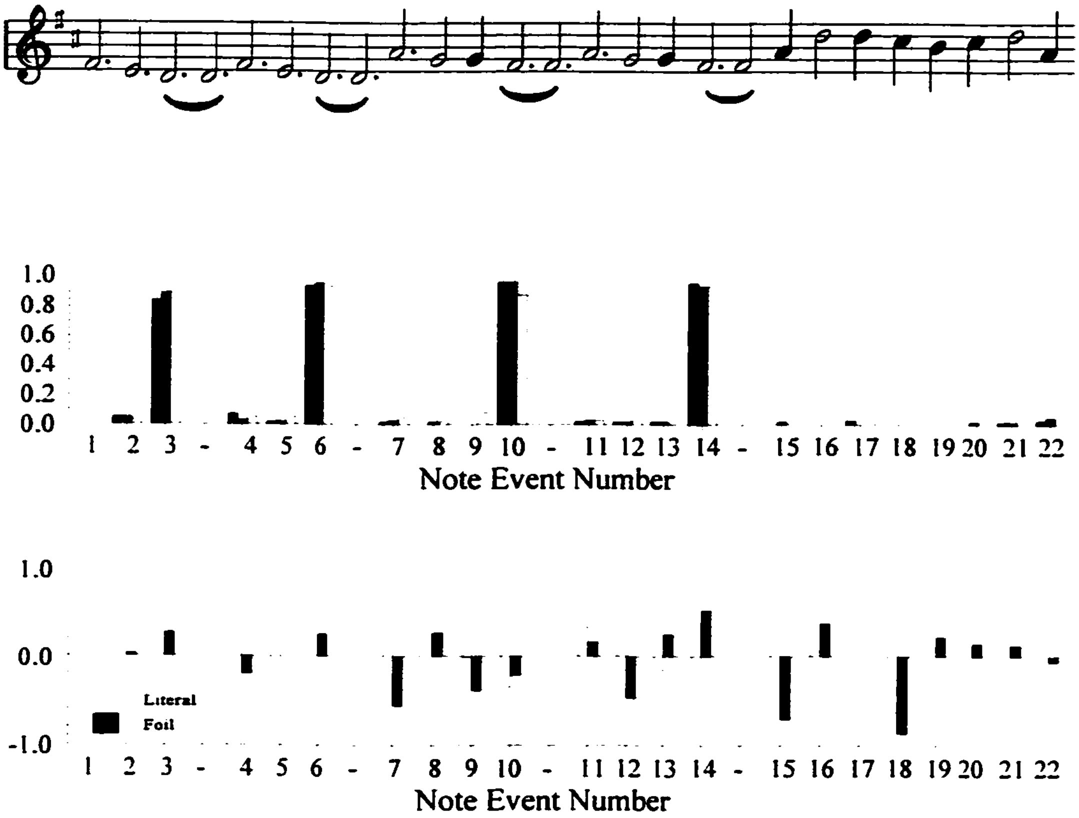


Figure 3.19a: The first part of the melody used in Stage 4, Part 1, and the location of subjective boundaries based on the average of all subjects (first, second and third repetition of the song). The boundaries placed by subjects can be compared with performance on test sequences extracted from the song. Test sequences are notated by the position of the first note (i.e., a sequence extracted at position 13 likely has a boundary on the second note of the sequence).

■ **First Presentation**
■ **Second Presentation**
□ **Third Presentation**

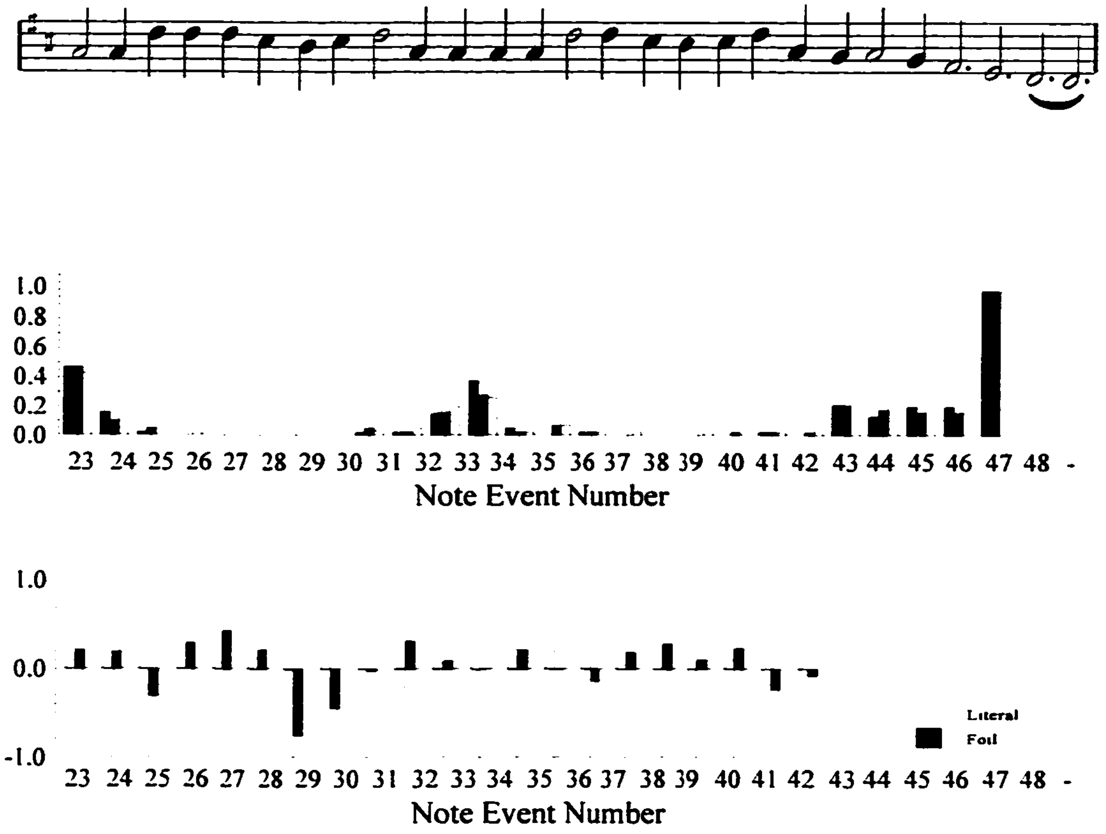


Figure 3.19b: The second part of the melody used in Stage 4, Part 1, and the location of subjective boundaries based on the average of all subjects (first, second and third repetition of the song). The boundaries placed by subjects can be compared with performance on test sequences extracted from the song. Test sequences are notated by the position of the first note (i.e., a sequence extracted at position 23 likely has a boundary on the second note of the sequence, but this is not assured).

■ First Presentation
■ Second Presentation
☐ Third Presentation

Boundary Location: Stages 6 and 8, Part 1

In Part 1 of both Stages 6 and 8, the melody "Softly Now the Light of Day" was presented to subjects. Subjects parsed the same melody on six occasions (3 in each stage), but subjects were not explicitly informed that they would received the same melody in both stages. Hence, layered on the analysis within each stage is a comparison between the two stages. Hence, in addition to the previous analyses, there are comparisons across sessions (Stage 6 and 8). Again, the analyses are completed and presented for all measures to demonstrate any (subtle) differences that might be apparent.

Over both stages (all six repetitions), the average consistency was 0.89±0.08 for the simple similarity measure (Table 3.41).

Table 3.41: The average consistency of subjects over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the Simple similarity measure of Table 3.8.

Consistency Measure		Mean & SD	Median	Mode	Min	Max
Simple Similarity	1 to 2	0.878±0.086	0.882	0.912	0.647	1.000
	1 to 3	0.872±0.085	0.882	0.882	0.618	1.000
	2 to 3	0.893±0.087	0.912	0.941	0.588	1.000
	1 to 4	0.892±0.066	0.912	0.912	0.677	1.000
	1 to 5	0.876±0.075	0.882	0.912	0.647	0.971
	1 to 6	0.882±0.071	0.882	0.882	0.706	1.000
	2 to 4	0.897±0.078	0.912	0.882	0.618	1.000
	2 to 5	0.897±0.092	0.912	0.941	0.618	1.000
	2 to 6	0.901±0.089	0.912	0.853	0.588	1.000
	3 to 4	0.899±0.078	0.912	0.912	0.647	1.000
	3 to 5	0.908±0.089	0.912	1.000	0.618	1.000
	3 to 6	0.898±0.101	0.941	1.000	0.588	1.000
	4 to 5	0.900±0.090	0.912	0.912	0.588	1.000
	4 to 6	0.900±0.083	0.912	1.000	0.677	1.000
	5 to 6	0.907±0.090	0.941	1.000	0.677	1.000

The average consistency was 0.60±0.22 for the Jaccard measure (Table 3.42).

Table 3.42: The average consistency of subjects over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the Jaccard similarity measure of Table 3.8.

Consistency Measure		Mean & SD	Median	Mode	Min	Max
Jaccard Similarity	1 to 2	0.529±0.219	0.500	0.333	0.182	1.000
	1 to 3	0.501±0.191	0.462	0.429	0.133	1.000
	2 to 3	0.610±0.216	0.600	0.667	0.235	1.000
	1 to 4	0.549±0.187	0.500	0.500	0.200	1.000
	1 to 5	0.511±0.175	0.500	0.571	0.200	0.833
	1 to 6	0.524±0.198	0.500	0.429	0.167	1.000
	2 to 4	0.608±0.198	0.625	0.500	0.222	1.000
	2 to 5	0.636±0.231	0.625	1.000	0.188	1.000
	2 to 6	0.643±0.243	0.667	1.000	0.177	1.000
	3 to 4	0.615±0.220	0.571	0.500	0.200	1.000
	3 to 5	0.674±0.233	0.667	1.000	0.273	1.000
	3 to 6	0.651±0.266	0.667	1.000	0.111	1.000
	4 to 5	0.634±0.222	0.667	1.000	0.222	1.000
	4 to 6	0.626±0.245	0.556	1.000	0.167	1.000
	5 to 6	0.677±0.242	0.667	1.000	0.154	1.000

For the Kulczynski measure 2, the average consistency was 0.74±0.17 (Table 3.43).

Table 3.43: The average consistency of subjects over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the Kulczynski 2 similarity measure of Table 3.8.

Consistency Measure		Mean & SD	Median	Mode	Min	Max
Kulczynski 2 Similarity	1 to 2	0.690±0.171	0.686	0.733	0.325	1.000
	1 to 3	0.670±0.156	0.667	0.600	0.317	1.000
	2 to 3	0.743±0.166	0.750	1.000	0.382	1.000
	1 to 4	0.710±0.142	0.700	0.675	0.343	1.000
	1 to 5	0.749±0.156	0.774	0.917	0.367	1.000
	1 to 6	0.789±0.172	0.804	1.000	0.429	1.000
	2 to 4	0.685±0.140	0.675	0.733	0.386	0.917
	2 to 5	0.761±0.179	0.800	1.000	0.324	1.000
	2 to 6	0.763±0.206	0.804	1.000	0.208	1.000
	3 to 4	0.690±0.157	0.700	0.733	0.292	1.000
	3 to 5	0.765±0.187	0.800	1.000	0.303	1.000
	3 to 6	0.765±0.169	0.800	1.000	0.364	1.000
	4 to 5	0.751±0.165	0.750	1.000	0.333	1.000
	4 to 6	0.756±0.186	0.729	1.000	0.292	1.000
	5 to 6	0.788±0.181	0.800	1.000	0.268	1.000

For the Sokal and Sneath measure 4, the average consistency was 0.84±0.11 (Table 3.44).

Table 3.44: The average consistency of subjects over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the Sokal & Sneath 4 similarity measure of Table 3.8.

Consistency Measure		Mean & SD	Median	Mode	Min	Max
Sokal & Sneath 4 Similarity	1 to 2	0.807±0.111	0.812	0.841	0.578	1.000
	1 to 3	0.796±0.102	0.798	0.766	0.526	1.000
	2 to 3	0.838±0.109	0.849	1.000	0.553	1.000
	1 to 4	0.822±0.090	0.820	0.841	0.601	1.000
	1 to 5	0.842±0.102	0.858	0.950	0.553	1.000
	1 to 6	0.865±0.114	0.880	1.000	0.574	1.000
	2 to 4	0.805±0.093	0.812	0.841	0.575	0.950
	2 to 5	0.848±0.119	0.874	1.000	0.531	1.000
	2 to 6	0.849±0.134	0.883	1.000	0.536	1.000
	3 to 4	0.809±0.098	0.812	0.841	0.554	1.000
	3 to 5	0.851±0.121	0.883	1.000	0.506	1.000
	3 to 6	0.851±0.114	0.880	1.000	0.530	1.000
	4 to 5	0.844±0.105	0.841	1.000	0.569	1.000
	4 to 6	0.846±0.118	0.828	1.000	0.554	1.000
	5 to 6	0.864±0.119	0.883	1.000	0.530	1.000

Finally, for the phi measure, the average consistency was 0.67±0.22 (Table 3.45).

Table 3.45: The average consistency of subjects over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the Phi similarity measure of Table 3.8.

Consistency Measure		Mean & SD	Median	Mode	Min	Max
Phi Similarity	1 to 2	0.606±0.224	0.622	0.679	0.155	1.000
	1 to 3	0.584±0.204	0.549	0.531	0.052	1.000
	2 to 3	0.673±0.219	0.698	1.000	0.106	1.000
	1 to 4	0.637±0.180	0.622	0.679	0.199	1.000
	1 to 5	0.680±0.204	0.717	0.897	0.106	1.000
	1 to 6	0.728±0.229	0.749	1.000	0.147	1.000
	2 to 4	0.600±0.187	0.610	0.679	0.150	0.897
	2 to 5	0.693±0.238	0.746	1.000	0.061	1.000
	2 to 6	0.695±0.269	0.766	1.000	0.070	1.000
	3 to 4	0.610±0.198	0.610	0.679	0.107	1.000
	3 to 5	0.699±0.243	0.766	1.000	0.013	1.000
	3 to 6	0.697±0.228	0.749	1.000	0.059	1.000
	4 to 5	0.684±0.211	0.679	1.000	0.139	1.000
	4 to 6	0.688±0.237	0.653	1.000	0.107	1.000
	5 to 6	0.725±0.239	0.766	1.000	0.061	1.000

Although these values are generally lower than those found in Stage 4, all values indicate a high degree of consistency within individual subjects. This is somewhat more impressive when one recalls that Stages 6 and 8 were presented as different melodies (i.e.. subjects were not informed that the melody of Stage 6 would be represented in Stage 8). It is also evident that the values increase towards the end of Stage 8 implying the subjects had learnt the novel melody. As with Stage 4, Part 1, the modal value for many measures wa 1.000 indicating that many subjects performed identically across the various repetitions. Furthermore, by both the standard deviations and the minimums, individual subjects never fell to the level of chance indicated in Table 3.8. Again, one must remember that these chance levels assumed equal numbers of boundaries per boundary profile and a particular probability for boundaries within a boundary profile.

As in Stage 4, the consistency ratings were further examined by computing the correlation between consistencies across subjects. That is, in Stage 6, the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 1 and 3, the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3 and the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3. The same pattern was followed in Stage 8; in addition, all the relationships between Stages 6 and 8 were computed, but to simplify the presentation, these were collapsed into a single set of statistics per measure. Regardless of the measure used, the resulting correlations were all significant, most at a type 1 error rate of $\alpha=0.05$, indicating that the subjects who were the most consistent in the first two repetitions remained the most consistent in the third repetition (Table 3.46). However, these numbers were, as the consistency rating in general, not as high as those in Stage 4.

The within-subjects ANOVA of the change in the consistency measure between repetitions is shown in Table 3.47 (overall and contrast analysis). Again, the entire set of interrelations was analyzed, but only the important relations are presented (changes in consistency within Stage 6, changes in consistency within Stage 8 and the changes in consistency between Stages 6 and 8 collapsed).

Table 3.46: The correlations between consistency ratings over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the measures of Table 3.8.

	Stage 6 Correlations			Stage 6 to Stage 8	Stage 8 Correlations		
		1 to 3	2 to 3			4 to 6	5 to 6
Simple Similarity	1 to 2	0.708**	0.564**	0.620±0.245 0.419 1.000	4 to 5	0.744**	0.709**
	1 to 3		0.656**		4 to 6		0.768**
Jaccard Similarity	1 to 2	0.659**	0.326*	0.458±0.233 0.181 1.000	4 to 5	0.709**	0.594**
	1 to 3		0.446**		4 to 6		0.636**
Kulczynski 2 Similarity	1 to 2	0.596**	0.370*	0.458±0.223 0.187 1.000	4 to 5	0.606**	0.579**
	1 to 3		0.425**		4 to 6		0.675**
Sokal & Sneath 4	1 to 2	0.629**	0.419*	0.503±0.225 0.168 1.000	4 to 5	0.641**	0.620**
	1 to 3		0.473**		4 to 6		0.699**
Phi	1 to 2	0.635**	0.415*	0.500±0.226 0.170 1.000	4 to 5	0.646**	0.616**
	1 to 3		0.474**		4 to 6		0.696**

Notes: The stage 6 to 8 correlation is the mean±sd, minimum and maximum correlations between measures in Stages and 8.
1 to 2 et cetera are correlations between repetitions.

For the all measures, consistency ratings changed overall (including Stage 6 and 8). However, for no measure were there any changes within Stage 8 (see Contrasts 3 and 4 in Table 3.47). Consistency ratings were different within Stage 6 (Contrasts 1 and 2 of Table 3.47) and consistency ratings were different between Stages 6 and 8 (Contrast 5 of Table 3.47). The implication is that during Stage 6, subjects learned the melody so that by the time they had reached Stage 8, their parsing of the melody was relatively stable. This is more impressive when one realizes that every subject rated the melody as unfamiliar in Stage 6 (some subjects expressed confusion about whether or not they should rate the melody as unfamiliar in Stage 8 when they had just heard it in Stage 6).

Table 3.47: Analysis of change in consistency within Stages 6 & 8 and between Stages 6 & 8, using measures cited in Table 3.8.

WS-ANOVA		SS$_{IV}$	df	SS$_{ERR}$	df	F	p	η^2	ω^2
Simple		0.092	14	1.567	840	3.512	0.000	0.055	0.040
Stage 6	Ψ_1	0.006	1	0.146	60	2.580	0.113	0.041	
	Ψ_2	0.006	1	0.086	60	4.101	0.047	0.064	
Stage 8	Ψ_3	0.001	1	0.105	60	0.783	0.380	0.013	
	Ψ_4	0.000	1	0.078	60	0.338	0.563	0.006	
Stage 6 to 8	Ψ_5	0.034	1	0.187	60	10.993	0.002	0.155	
Jaccard		1.788	14	11.577	840	9.266	0.000	0.134	0.119
Stage 6	Ψ_1	0.099	1	1.063	60	5.615	0.021	0.086	
	Ψ_2	0.105	1	0.446	60	14.104	0.000	0.190	
Stage 8	Ψ_3	0.038	1	0.746	60	3.059	0.085	0.049	
	Ψ_4	0.016	1	0.612	60	1.601	0.211	0.026	
Stage 6 to 8	Ψ_5	0.514	1	1.267	60	24.361	0.000	0.289	
Kulczynski 2		0.882	14	7.575	840	6.988	0.000	0.104	0.089
Stage 6	Ψ_1	0.052	1	0.669	60	4.675	0.035	0.072	
	Ψ_2	0.058	1	0.377	60	9.317	0.003	0.124	
Stage 8	Ψ_3	0.012	1	0.478	60	1.529	0.221	0.025	
	Ψ_4	0.010	1	0.413	60	1.506	0.225	0.024	
Stage 6 to 8	Ψ_5	0.286	1	0.825	60	20.779	0.000	0.257	
Sokal & Sneath 4		0.332	14	3.405	840	5.850	0.000	0.089	0.074
Stage 6	Ψ_1	0.020	1	0.305	60	4.001	0.050	0.063	
	Ψ_2	0.022	1	0.176	60	7.570	0.008	0.112	
Stage 8	Ψ_3	0.004	1	0.219	60	1.218	0.274	0.020	
	Ψ_4	0.004	1	0.184	60	1.153	0.287	0.019	
Stage 6 to 8	Ψ_5	0.113	1	0.372	60	18.27	0.000	0.233	
Phi		1.189	14	10.625	840	6.712	0.000	0.101	0.086
Stage 6	Ψ_1	0.074	1	0.956	60	4.652	0.035	0.072	
	Ψ_2	0.076	1	0.511	60	8.954	0.004	0.130	
Stage 8	Ψ_3	0.018	1	0.680	60	1.586	0.213	0.026	
	Ψ_4	0.012	1	0.565	60	1.314	0.256	0.021	
Stage 6 to 8	Ψ_5	0.385	1	1.181	60	19.560	0.000	0.246	

Notes: Ψ_1 (& Ψ_3) compare Consistency 1 to 2 (4 to 5) with 2 to 3 (5 to 6).

Ψ_2 (& Ψ_4) compare the average Consistency of 1 to 2 (4 to 5) and 2 to 3 (5 to 6) with Consistency 1 to 3 (4 to 6).

Ψ_5 compares the averages within Stage 6 to those within Stage 8.

In examining these results, it is important to note that although the results are significant, the size of the effect is quite small (ω^2 and η^2). Given that this is within-subjects design (this is an experimental variable), one would expect these values to be much closer to 1.0 for an important effect. That is, within-subjects designs (with 61 subjects, and no intervening time) have lots of power, and as such, would be expected to produce strong effects. On this basis, these changes are not too important, despite significance (i.e., treat ω^2 and η^2 in the manner of r^2: a correlation of 0.04 within an experimental design would not be impressive).

When comparing across subjects (Table 3.48), average similarity ratings were lower both in comparison to the within-subjects consistency analysis and in comparison to Stage 4, with means of 0.87±0.09 for the simple similarity measure, 0.55±0.22 for the Jaccard measure, 0.71±0.17 for the Kulczynski measure 2, 0.82±0.11 for the Sokal and Sneath measure 4 and 0.63±0.23 for the phi measure. These values, representing the averages of the mean similarity measures for Stages 6 and 8, are still above the critical and chance levels specified in Table 3.8.

Table 3.48: The average similarity ratings between different subjects for the same melody, within Stages 6 and 8, Part 1, Repetition 3 only, given the similarity measures of Table 3.8.

WS-ANOVA		Mean & SD	Median	Mode	Min	Max
Simple Similarity	Stage 6	0.871±0.087	0.882	0.912	0.559	1.000
	Stage 8	0.877±0.088	0.882	0.853	0.500	1.000
Jaccard Similarity	Stage 6	0.538±0.214	0.500	0.500	0.000	1.000
	Stage 8	0.561±0.231	0.500	1.000	0.105	1.000
Kulczynski 2 Similarity	Stage 6	0.702±0.169	0.700	0.917	0.000	1.000
	Stage 8	0.718±0.177	0.733	1.000	0.202	1.000
Sokal & Sneath 4 Similarity	Stage 6	0.811±0.109	0.812	0.950	0.433	1.000
	Stage 8	0.821±0.115	0.828	1.000	0.424	1.000
Phi Similarity	Stage 6	0.616±0.219	0.622	0.897	-0.133	1.000
	Stage 8	0.636±0.231	0.643	1.000	-0.151	1.000

In contrast to Stage 4, the minimum values for all the measures dip below the aforementioned chance value, although for some measures, this is only during Stage 6. As shown in Table 3.49, all measures show a significant change in similarity from Stage to Stage 8. However, in this analysis it must be remembered that there were 61*60/2=1830 values for each stage (in each stage, there was a similarity assessment for each subject with every other subject). The corresponding pairings for each stage were treated as two levels of a within-subjects variable. Hence, there was a huge amount of power (and consequently, significance). However, given the ω^2 this change is not important because one must remember that the Stage 6 to Stage 8 transition is a well controlled experimental variable (only the time changes: all other aspects of the stimulus are identical) so if time per se mattered, one would expect a large ω^2: None are remotely large.

Table 3.49: The analyses of change in similarity ratings for the same melody, from Stages 6 and 8, Part 1, Repetition 3 only, given the similarity measures of Table 3.8.

WS-ANOVA	SS	df	SS	df	*F*	*p*	η^2	ω^2
Simple	0.036	1	5.292	1829	12.541	0.000	0.007	0.006
Jaccard	0.481	1	43.986	1829	19.983	0.000	0.011	0.010
Kulczynski 2	0.252	1	30.411	1829	15.128	0.000	0.008	0.008
Sokal & Sneath 4	0.951	1	11.672	1829	14.900	0.000	0.008	0.008
Phi	0.360	1	46.608	1829	14.108	0.000	0.008	0.007

Correlation	Stage 6 to Stage 8
Simple	0.620**
Jaccard	0.518**
Kulczynski 2	0.447**
Sokal & Sneath 4	0.491**
Phi	0.498**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

To further explore the relationship between different subjects, a cluster analysis was performed on each of the similarity matrices (i.e., using each of the similarity measures) determined from the third repetition of the melody in Stages 6 and 8 only (only the phi measure of Stages 6 and 8 is presented in Figure 3.20; other clusterings were very similar which would be expected given that the modal values of similarity were often 1.0). Using the previous critical values (Table 3.8) to define a cut line for the creation of subgroups did not produce distinct groups of subjects. The cluster profiles were indicative of a situation in which individual subjects were gradually added to an homogeneous group: It is best that all subjects be treated as one homogeneous group.

It is possible to compare the clustering for Stage 4 (Figure 3.15) with the clustering for boundary location in Stages 6 and 8 (Figure 3.20): Note also that the clustering in Stage 8 is even more consistent with the interpretation of a single homogeneous group than it is in Stage 6.

Given that all subjects can be treated as one group, it is possible to create the average boundary profile for each repetition. The average boundary profile for each repetition is shown in Figure 3.21 (again, the figure is split for ease of viewing). Subjects placed strong boundaries at locations 9, 16 and 25, with weaker boundaries (i.e., fewer subjects) at locations 26 and 27. The basic statistics for these profiles are shown in Tables 3.50 to 3.55. Over all repetitions, the mean number of boundaries was fairly constant at 6.22±0.41. The modal number of boundaries remained constant at 5.00. This represented a mean number of note events *between* boundaries equal to 6.59±1.30; the mean number of note events within a unit should therefore be 7.59 because each boundary should be associated with either the beginning or the end of a unit. The size of individual units ranged from 0.00 note events to 32.00 note events. However, on a subject by subject basis, the most common unit size was 8.00 note events followed by 7.00 and 6.00 (implying 9.00, 8.00 and 7.00 note events, respectively, when the boundary is included), and the least common unit sizes were 9.00, and 10.00 or more (implying 10.00, and 11.00 or more when boundaries are included). These observations are most clear in the modal values for the various unit lengths. Most unit lengths have a modal value of 0 indicating that most subjects did not use that length at all. For example, for the sixth repetition, only unit lengths of 6, 7 and 8 have modal values other than 0.

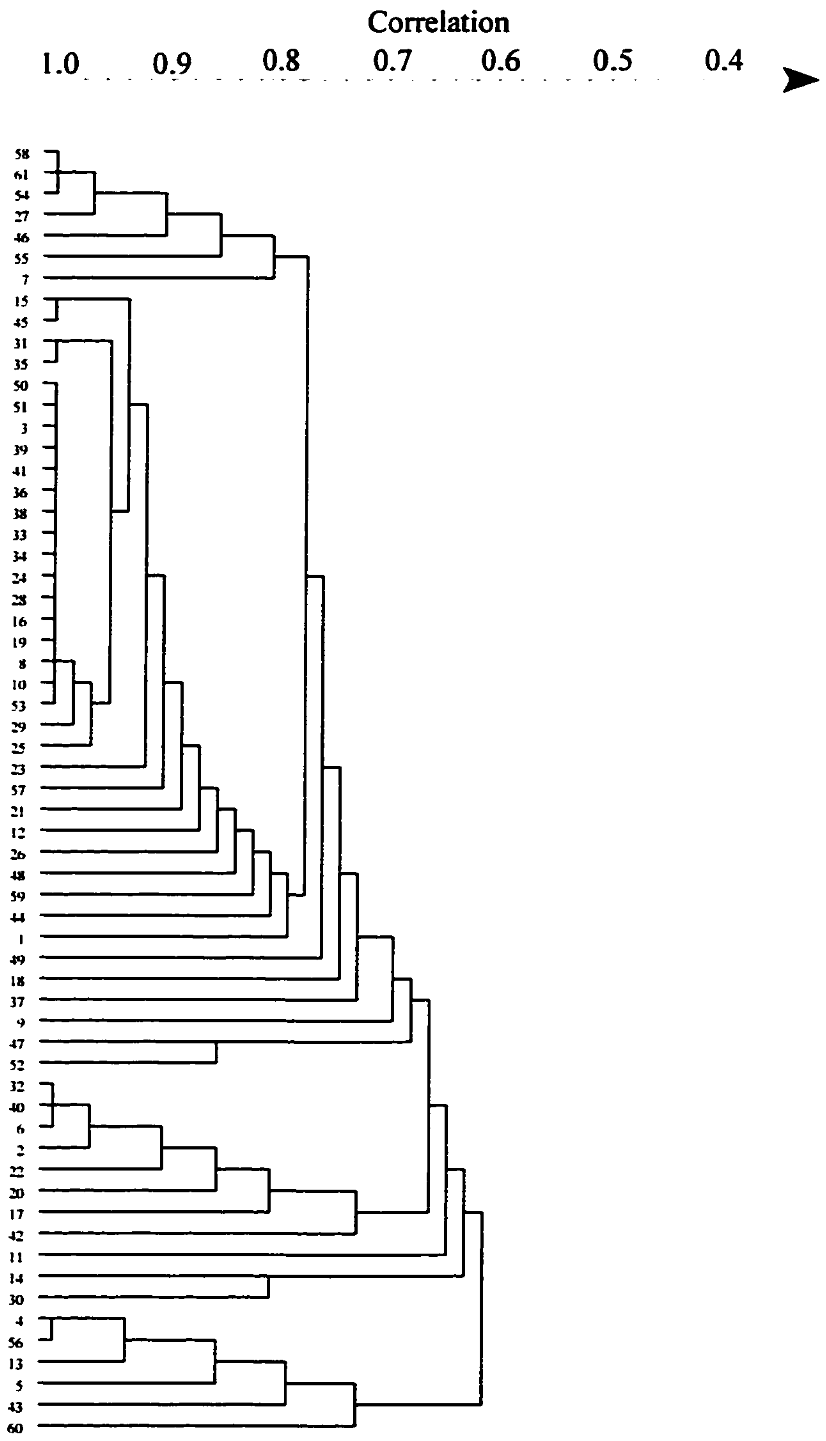


Figure 3.20a: Cluster analysis of phi similarities within Stage 6. Compare with Figure 3.15 and note that the minimum clustering slightly lower than in Stage 4.

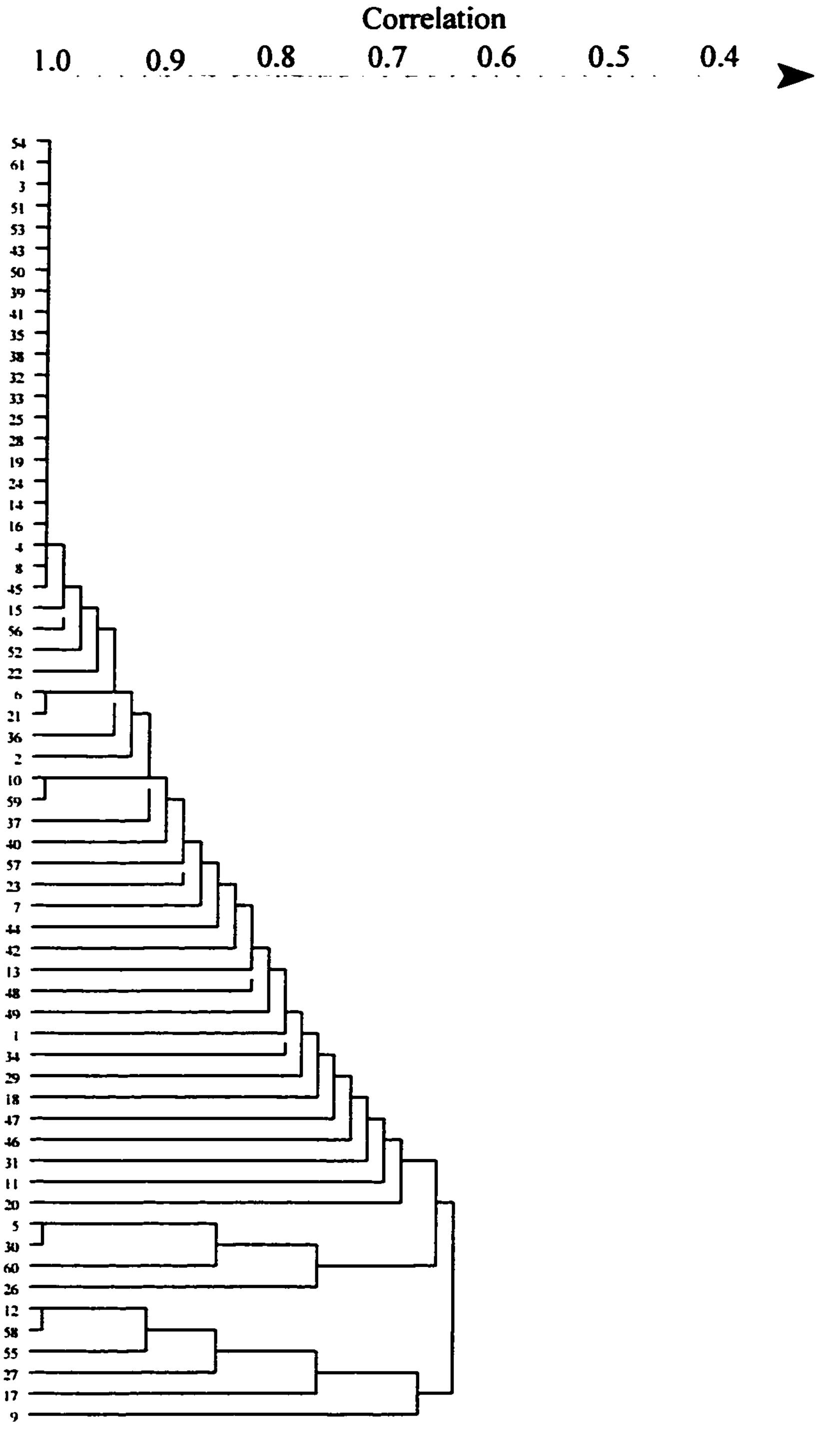


Figure 3.20b: Cluster analysis of phi similarities within Stage 8. Compare with Figure 3.15 and 3.20a and note that the minimum clustering is lower than in Stage 4 but higher than in Stage 6.

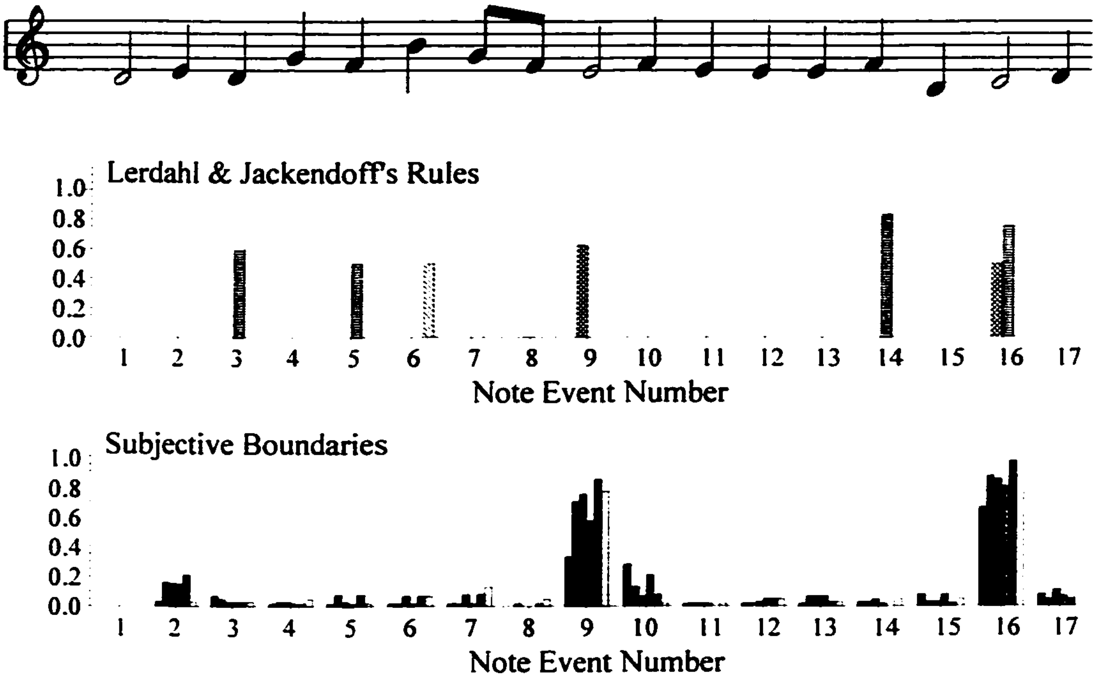


Figure 3.21a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared to the location of the subjective boundaries based upon the average of all subjects for the first through sixth presentations of the song..

▨ Rule 2a ■ Rule 3a ■ First Presentation ■ Fourth Presentation
▩ Rule 2b ▧ Rule 3d ■ Second Presentation ■ Fifth Presentation
☐ Third Presentation ☐ Sixth Presentation

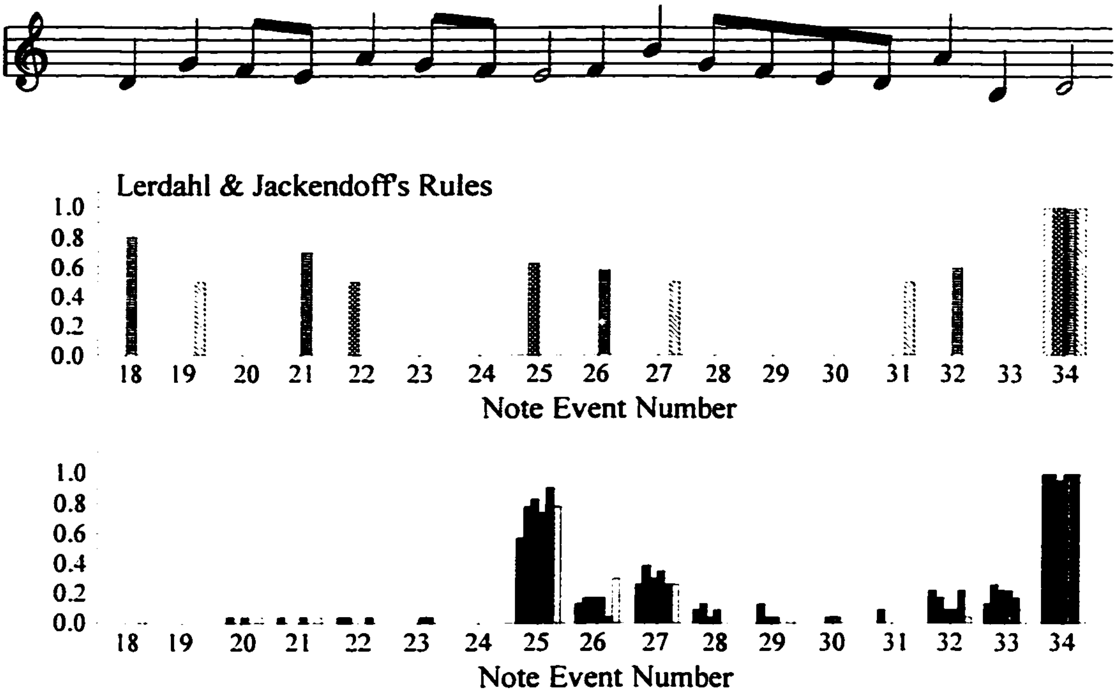


Figure 3.21b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of the subjective boundaries based on the average of all subjects for the first through sixth presentations of the song..

▨ Rule 2a
▩ Rule 2b
▦ Rule 3a
▧ Rule 3d
■ First Presentation
■ Second Presentation
□ Third Presentation
■ Fourth Presentation
■ Fifth Presentation
□ Sixth Presentation

Table 3.50: The important statistics for the location of boundaries within Stage 6, Part 1, Repetition 1 (the first presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	5.541±8.557	4.000	0.000	0.000	32.000
Maximum	13.000±7.414	9.000	8.000	5.000	32.000
Mean	9.094±7.548	7.250	7.250	2.300	32.000
Number with 0	0.426±0.694	0.000	0.000	0.000	3.000
Number with 1	0.443±0.807	0.000	0.000	0.000	3.000
Number with 2	0.131±0.340	0.000	0.000	0.000	1.000
Number with 3	0.230±0.529	0.000	0.000	0.000	2.000
Number with 4	0.197±0.477	0.000	0.000	0.000	2.000
Number with 5	0.557±0.671	0.000	0.000	0.000	2.000
Number with 6	0.344±0.574	0.000	0.000	0.000	2.000
Number with 7	0.508±0.622	0.000	0.000	0.000	2.000
Number with 8	0.967±0.912	1.000	1.000	0.000	3.000
Number with 9	0.131±0.340	0.000	0.000	0.000	1.000
Number with ≥10	0.492±0.566	0.000	0.000	0.000	2.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Number with "xx" is the mean number of times subjects created units containing xx note events (exclusive of the actual boundaries).

Table 3.51: The important statistics for the location of boundaries within Stage 6, Part 1, Repetition 2 (the second presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	2.459±3.314	0.000	0.000	0.000	14.000
Maximum	9.689±3.901	8.000	8.000	5.000	26.000
Mean	5.993±3.014	5.600	7.250	1.750	15.500
Number with 0	0.820±1.057	1.000	0.000	0.000	5.000
Number with 1	0.525±0.788	0.000	0.000	0.000	3.000
Number with 2	0.279±0.710	0.000	0.000	0.000	4.000
Number with 3	0.328±0.676	0.000	0.000	0.000	2.000
Number with 4	0.279±0.488	0.000	0.000	0.000	2.000
Number with 5	0.426±0.670	0.000	0.000	0.000	2.000
Number with 6	0.852±0.727	1.000	1.000	0.000	3.000
Number with 7	0.656±0.629	1.000	1.000	0.000	2.000
Number with 8	0.967±0.816	1.000	1.000	0.000	3.000
Number with 9	0.082±0.277	0.000	0.000	0.000	1.000
Number with ≥10	0.262±0.480	0.000	0.000	0.000	2.000

Notes: see Table 3.50

Table 3.52: **The important statistics for the location of boundaries within Stage 6, Part 1, Repetition 3 (the third presentation of the melody).**

	Mean & SD	Median	Mode	Min	Max
Minimum	2.967±3.245	1.000	0.000	0.000	14.000
Maximum	8.623±2.230	8.000	8.000	5.000	17.000
Mean	5.933±2.387	5.600	7.250	1.357	15.500
Number with 0	0.721±1.380	0.000	0.000	0.000	9.000
Number with 1	0.508±0.829	0.000	0.000	0.000	3.000
Number with 2	0.262±0.728	0.000	0.000	0.000	4.000
Number with 3	0.230±0.589	0.000	0.000	0.000	3.000
Number with 4	0.295±0.558	0.000	0.000	0.000	2.000
Number with 5	0.262±0.603	0.000	0.000	0.000	3.000
Number with 6	0.869±0.670	1.000	1.000	0.000	2.000
Number with 7	0.869±0.695	1.000	1.000	0.000	3.000
Number with 8	1.131±0.826	1.000	1.000	0.000	3.000
Number with 9	0.115±0.321	0.000	0.000	0.000	1.000
Number with ≥10	0.148±0.441	0.000	0.000	0.000	2.000

Notes: see Table 3.50

Table 3.53: The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 1 (the fourth presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	4.033±6.096	2.000	0.000	0.000	32.000
Maximum	10.000±5.335	8.000	8.000	5.000	32.000
Mean	6.941±5.352	5.600	7.250	2.000	32.000
Number with 0	0.574±0.921	0.000	0.000	0.000	4.000
Number with 1	0.541±0.808	0.000	0.000	0.000	3.000
Number with 2	0.213±0.581	0.000	0.000	0.000	3.000
Number with 3	0.262±0.575	0.000	0.000	0.000	2.000
Number with 4	0.328±0.539	0.000	0.000	0.000	2.000
Number with 5	0.361±0.517	0.000	0.000	0.000	2.000
Number with 6	0.721±0.733	1.000	1.000	0.000	3.000
Number with 7	0.672±0.651	1.000	1.000	0.000	3.000
Number with 8	1.131±0.826	1.000	1.000	0.000	3.000
Number with 9	0.115±0.370	0.000	0.000	0.000	2.000
Number with ≥10	0.213±0.451	0.000	0.000	0.000	2.000

Notes: see Table 3.50

Table 3.54: The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 2 (the fifth presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	2.623±2.776	1.000	0.000	0.000	8.000
Maximum	8.197±2.190	8.000	8.000	5.000	17.000
Mean	5.571±1.930	5.600	7.250	2.000	10.000
Number with 0	0.623±0.916	0.000	0.000	0.000	3.000
Number with 1	0.426±0.741	0.000	0.000	0.000	4.000
Number with 2	0.410±0.901	0.000	0.000	0.000	4.000
Number with 3	0.377±0.687	0.000	0.000	0.000	3.000
Number with 4	0.262±0.575	0.000	0.000	0.000	2.000
Number with 5	0.311±0.593	0.000	0.000	0.000	2.000
Number with 6	1.131±0.826	1.000	1.000	0.000	3.000
Number with 7	0.689±0.672	1.000	1.000	0.000	3.000
Number with 8	1.213±0.798	1.000	1.000	0.000	3.000
Number with 9	0.000±0.000	0.000	0.000	0.000	0.000
Number with ≥10	0.082±0.277	0.000	0.000	0.000	1.000

Notes: see Table 3.50

Table 3.55: The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 3 (the sixth presentation of the melody).

	Mean & SD	Median	Mode	Min	Max
Minimum	3.475±3.486	3.000	6.000	0.000	15.000
Maximum	8.541±2.592	8.000	8.000	4.000	17.000
Mean	6.055±2.653	5.600	7.250	1.750	15.500
Number with 0	0.689±1.104	0.000	0.000	0.000	4.000
Number with 1	0.393±0.690	0.000	0.000	0.000	3.000
Number with 2	0.311±0.765	0.000	0.000	0.000	4.000
Number with 3	0.197±0.542	0.000	0.000	0.000	3.000
Number with 4	0.393±0.842	0.000	0.000	0.000	4.000
Number with 5	0.344±0.602	0.000	0.000	0.000	2.000
Number with 6	0.852±0.573	1.000	1.000	0.000	2.000
Number with 7	0.770±0.616	1.000	1.000	0.000	3.000
Number with 8	1.131±0.922	1.000	2.000	0.000	3.000
Number with 9	0.115±0.370	0.000	0.000	0.000	2.000
Number with ≥10	0.148±0.441	0.000	0.000	0.000	2.000

Notes: see Table 3.50

The correlation matrix for each repetition is included as additional information (Tables 3.56 through 3.61). In contrast to the melody of Stage 4, those subjects having the highest minimum span between boundaries were generally the same subjects who had the highest maximum span between boundaries. That is, some subjects seemed to used larger units overall. Note that the correlations do not change as a function of repetition number.

Table 3.56: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 1 (the first presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.799**	1.000	
Mean	0.966**	0.908**	1.000
Number with 0	-0.404**	-0.272	-0.392**
Number with 1	-0.340*	-0.237	-0.369*
Number with 2	-0.231	-0.271	-0.280
Number with 3	-0.245	-0.323*	-0.327*
Number with 4	-0.223	-0.203	-0.276
Number with 5	-0.358*	-0.422**	-0.384*
Number with 6	-0.208	-0.306*	-0.232
Number with 7	-0.240	-0.278	-0.239
Number with 8	-0.256	-0.493**	-0.340*
Number with 9	-0.082	-0.145	-0.099
Number with ≥10	0.388*	0.671**	0.493**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.57: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 2 (the second presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.482**	1.000	
Mean	0.838**	0.816**	1.000
Number with 0	-0.585**	-0.333*	-0.597**
Number with 1	-0.445**	-0.244	-0.489**
Number with 2	-0.268	-0.251	-0.384*
Number with 3	-0.306*	-0.264	-0.395**
Number with 4	-0.338*	-0.164	-0.385*
Number with 5	-0.367*	-0.165	-0.379*
Number with 6	-0.338*	-0.363*	-0.296
Number with 7	-0.115	-0.336*	-0.117
Number with 8	0.437**	-0.181	0.205
Number with 9	-0.042	0.148	0.061
Number with ≥10	0.489**	0.837**	0.664**

Notes: * Significant (one tailed) at p<0.05
** Significant (one tailed) at p<0.01

Table 3.58: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 3 (the third presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.542**	1.000	
Mean	0.887**	0.730**	1.000
Number with 0	-0.486**	-0.208	-0.597**
Number with 1	-0.520**	-0.309*	-0.638**
Number with 2	-0.314*	-0.215	-0.449**
Number with 3	-0.345*	-0.276	-0.463**
Number with 4	-0.492**	-0.271	-0.620**
Number with 5	-0.200	-0.086	-0.202
Number with 6	-0.171	-0.324*	-0.118
Number with 7	-0.017	-0.248	0.041
Number with 8	0.400**	0.000	0.363*
Number with 9	0.020	0.061	0.045
Number with ≥10	0.411**	0.820**	0.536**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.59: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 1 (the fourth presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.785**	1.000	
Mean	0.967**	0.889**	1.000
Number with 0	-0.419**	-0.193	-0.375*
Number with 1	-0.396**	-0.174	-0.383*
Number with 2	-0.228	-0.198	-0.278
Number with 3	-0.250	-0.234	-0.299*
Number with 4	-0.297*	-0.151	-0.310*
Number with 5	-0.310*	-0.272	-0.304*
Number with 6	-0.278	-0.379*	-0.283
Number with 7	-0.119	-0.341*	-0.173
Number with 8	-0.067	-0.299*	-0.138
Number with 9	-0.017	0.017	0.014
Number with ≥10	0.512**	0.761**	0.599**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.60: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 2 (the fifth presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.289	1.000	
Mean	0.844**	0.621**	1.000
Number with 0	-0.653**	-0.145	-0.674**
Number with 1	-0.480**	-0.186	-0.543**
Number with 2	-0.357*	-0.438**	-0.595**
Number with 3	-0.317*	-0.383*	-0.572**
Number with 4	-0.240	-0.307*	-0.449**
Number with 5	-0.292	-0.382*	-0.449**
Number with 6	-0.291	0.105	-0.087
Number with 7	0.133	-0.048	0.289
Number with 8	0.691**	0.090	0.606**
Number with 9	0.000	0.000	0.000
Number with ≥10	0.193	0.854**	0.468**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 3.61: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 3 (the sixth presentation of the melody).

	Correlations between Measures		
	Minimum	Maximum	Mean
Maximum	0.556**	1.000	
Mean	0.927**	0.764**	1.000
Number with 0	-0.632**	-0.319*	-0.645**
Number with 1	-0.509**	-0.270	0.529**
Number with 2	-0.375*	-0.297	-0.501**
Number with 3	-0.297	-0.314*	-0.365*
Number with 4	-0.383*	-0.359*	-0.466**
Number with 5	-0.373*	-0.185	-0.338*
Number with 6	-0.232	-0.091	-0.161
Number with 7	0.021	-0.109	0.014
Number with 8	0.416**	-0.156	0.286
Number with 9	-0.250	0.056	-0.124
Number with ≥10	0.572**	0.862**	0.717**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

The analysis of the change in boundary statistics is presented in Table 3.62; the correlations between the repetitions are shown in the next table (Table 3.63). The first thing to note is that all these measures of boundary location are changing across repetitions and given the ω^2 or η^2, these changes are not very large (again, this is an experimental within-subjects variable) ranging from 5 to 17% of the variance: Comparably, in Stage 4, the changes were less than 1%. This difference can be attributed to the fact that the melody of Stage 4 was familiar to all subjects, while the melody of Stages 6 and 8 was unfamiliar to all subjects. Much of the change is within Stage 6, again implying that subjects are learning their parsing of the melody. The most dramatic change occurs in the average maximum distance between boundaries. The maximum distance decreases from Stage 6, Repetition 1 to Stage 8 Repetition 3 (the sixth presentation of the melody), which one might think is contrary to expectations: One might expect that the units would get larger as the subject learns the melody, but the reverse seems to manifest itself.

Table 3.62: The analysis of the changes in the boundary statistics across repetitions of the same melody, for Stages 6 and 8, Part 1.

WS-ANALYSIS		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Minimum Distance Between Boundaries	Overall	401.697	5	5500.470	300	4.382	0.001	0.068	0.052
	Ψ_1	202.041	1	2143.459	60	5.656	0.021	0.086	
	Ψ_2	131.041	1	794.126	60	9.901	0.003	0.142	
	Ψ_3	9.475	1	1099.525	60	.517	0.475	0.009	
	Ψ_4	52.033	1	602.967	60	5.187	0.026	0.079	
	Ψ_5	7.107	1	860.393	60	.496	0.484	0.008	
Maximum Distance Between Boundaries	Overall	960.112	5	4216.388	300	13.663	0.000	0.185	0.177
	Ψ_1	584.336	1	1580.164	60	22.188	0.000	0.270	
	Ψ_2	51.281	1	909.552	60	3.383	0.071	0.053	
	Ψ_3	64.926	1	885.574	60	4.399	0.040	0.068	
	Ψ_4	46.888	1	233.279	60	12.060	0.001	0.167	
	Ψ_5	212.680	1	607.820	60	20.994	0.000	0.259	
Mean Distance Between Boundaries	Overall	518.933	5	3782.619	300	8.231	0.000	0.121	0.106
	Ψ_1	304.849	1	1538.393	60	11.890	0.001	0.165	
	Ψ_2	94.066	1	639.391	60	8.827	0.004	0.128	
	Ψ_3	23.918	1	796.642	60	1.801	0.185	0.029	
	Ψ_4	34.951	1	280.244	60	7.483	0.008	0.111	
	Ψ_5	61.148	1	527.949	60	6.949	0.011	0.104	

Notes Ψ_1 compared Repetition 1 to Repetition 3 (within Stage 6)
Ψ_2 compared Repetitions 1 and 3 to Repetition 2 (within Stage 6)
Ψ_3 compared Repetition 4 to Repetition 6 (within Stage 8)
Ψ_4 compared Repetitions 4 and 6 to Repetition 5 (within Stage 8)
Ψ_5 compared Repetitions 1, 2 and 3 to Repetitions 4, 5 and 6 (Stage 6 to Stage 8)

Table 3.63: The correlation of the boundary statistics across repetitions of the same melody, for Stages 6 and 8, Part 1.

Correlation Within Measures			Repetition Number				
			2	3	4	5	6
Minimum Distance Between Boundaries	Repetition Number	1	0.324*	0.222	0.426**	0.072	0.293*
		2	1.000	0.553**	0.366**	0.528**	0.507**
		3		1.000	0.386**	0.519**	0.492**
		4			1.000	0.191	0.298*
		5				1.000	0.506**
Maximum Distance Between Boundaries	Repetition Number	1	0.149	0.220	0.423**	0.195	0.180
		2	1.000	0.571**	0.422**	0.632**	0.357**
		3		1.000	0.409**	0.661**	0.497**
		4			1.000	0.518**	0.205
		5				1.000	0.551**
Mean Distance Between Boundaries	Repetition Number	1	0.288*	0.316*	0.484**	0.240	0.316*
		2	1.000	0.705**	0.544**	0.610**	0.501**
		3		1.000	0.461**	0.731**	0.635**
		4			1.000	0.423**	0.322*
		5				1.000	0.694**

Notes: Repetitions 1, 2 and 3 are Stage 6 and Repetitions 4, 5 and 6 are Stage 8
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Despite some changes in boundary formation, one can generally conclude, as in Stage 4, that there is an average boundary profile that is a fair representation of the boundary profiles of individual subjects (i.e., changes over repetitions is not dramatic). Given this, one can then attempt to use each of Lerdahl and Jackendoff's (1983) rules for parsing (cf., Chapter 2) to predict those empirically derived boundary locations. Thus, the empirically derived boundary profiles were compared to each of the theoretically derived boundary profiles representing each of the rules of Lerdahl and Jackendoff (see Figure 3.24). The correlations between the average boundary profile at each repetition, and the correlations between each repetition of the melody and the rules are presented in Table 3.64. As with Stage 4, the correlation was calculated for three values of relative shift (i.e.,

-1, 0, and 1) between the empirically derived boundary profile and the theoretical boundary profiles based on each rule. In this aligning process, the theoretically derived boundary profiles were shifted to the left, and right and it is the correlations that were determined from each shift that are presented in Table 3.64. From Table 3.64, it can be seen that, as in Stage 4, all the rules are optimally aligned at the same shift, but it is not quite so unequivocal (see Rule 3a). Rule 2b (Attack-point proximity), Rule 3a (Register change) and Rule 3d (Length change) imply that subjects have placed boundaries at the end of units. Only Rule 2b (Attack point) is truly successful at predicting boundaries (average r=0.73; r^2=0.54), while Rule 3a (Register change) is marginally successful at predicting boundaries (average r=0.33; r^2=0.11) and Rule 3d (Length change) is much less effective (average r=0.25, r^2=.06). None of the rules is as successful as its application in Stage 4. The simple correlations between the different rules are also presented in Table 3.64 as a reminder.

Table 3.64 The correlations between the rules of Lerdahl and Jackendoff (1983) and the empirically derived boundary profiles for each repetition of the melody in Stage 6 and 8, Part 1, as well as, the correlations between the rules. The rules were shifted to aligned with the empirical boundaries.

	Shift	Repetition Number					
		1	2	3	4	5	6
Rep 1		1.000	0.947**	0.932**	0.974**	0.905**	0.924**
Rep 2			1.000	0.990**	0.985**	0.982**	0.980**
Rep 3				1.000	0.984**	0.993**	0.993**
Rep 4					1.000	0.972**	0.971**
Rep 5						1.000	0.980**
Rule 2b	+1	0.724**	0.731**	0.748**	0.719**	0.743**	0.744**
	0	0.017	-0.105	-0.116	-0.029	-0.103	-0.139
	-1	-0.161	-0.173	-0.166	-0.172	-0.175	-0.145
Rule 3a	+1	0.396*	0.310	0.307	0.319	0.292	0.326
	0	-0.104	-0.112	-0.118	0.131	-0.177	-0.123
	-1	0.270	0.238	0.258	0.256	0.271	0.254
Rule 3d	+1	0.383*	0.277	0.274	0.313	0.225	0.283
	0	0.144	0.150	0.176	0.152	0.185	0.149
	-1	-0.082	-0.060	-0.107	-0.105	-0.116	-0.039

Correlations Between Rules				
	Rule 2a	Rule 2b	Rule 3a	Rule 3d
Rule 2a	1.000	---	---	---
Rule 2b		1.000	0.078	-0.125
Rule 3a			1.000	-0.221

Notes: Rules 3b, 3c, and 2a concerned attributes that did not change within this particular melody.

Correlations between the rules and the empirically observed boundary profiles are provided for each of three relative shifts

* $p < 0.05$

** $p < 0.01$

Interestingly, with the exception of Rule 2b, the correlations of the rules with the empirical boundaries decreased with the repetition number. It is as if subjects used the rules initially and then, as the melody became more familiar, they digressed from the rules. Note that the digression from Rules 3a and 3d was not compensated for by increased use of Rule 2b (i.e., the correlation did not increase). Again, it is the profiles themselves that are much more informative (see Figure 3.21; the profile is divided for ease of presentation). In Figure 3.21, it is the boundary profile based on the rules of Lerdahl and Jackendoff that is shifted relative to the empirically determined boundary profile because the empirically determined positions are considered the more indicative of "true" parsing than the notions of theory. In the first part of the melody, Rule 2b (Attack point) aligned with the two strong empirically determined boundary positions. In the latter part of the melody, one occurrence of Rule 2b aligned with the empirically determined position, but there were secondary boundaries (Notes 26 and 27) that were not aligned with Rule 2b. In the first part of the melody, neither rule 3a (Register Change) nor 3d (Length Change) align very well. Both Rule 3a and 3d create many false positives (5 and 3 depending what one wants to called a false positive). Rule 2b did not create many false alarms (only 1 at Note 22). Also note that Rule 2a (Slur/Rest) did not apply within this particular piece of music.

Finally, it should be pointed out that the correlations between the theoretically derived boundary profiles and the empirically derived boundary profiles never get close to the intercorrelations between the average empirically-derived boundary profiles per repetition. This implies that there is stable responding that the rules are simply not capturing. The lack of fit between the rules and the empirical data does not reflect noise on the part of subjects; the lack of fit is an lack of understanding of the mechanisms that listeners use to parse music.

The next analysis used multiple regression to determine the best mix of rules for each empirically derived boundary profile at each repetition of the melody (Table 3.65). In Table 3.65, the complete equation for each stage represents the forced entry of all rules, providing the maximal prediction possible.

Table 3.65: The regression equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) in the average boundary profile for each repetition of the melody, in Stage 6 and 8, Part 1.

Regression Equations							
Rep	Constant	Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
1	B = 0.05±0.03	+ (0.60±0.13)*Rule2b	0.523	0.000	0.523	35.185	0.000
		+ (0.13±0.11)*Rule3d	0.021	0.245	0.544	18.517	0.000
		+ (0.11±0.09)*Rule3a	0.022	0.228	0.567	13.080	0.000
2	B = 0.08±0.04	+ (0.84±0.17)*Rule2b	0.534	0.000	0.534	36.678	0.000
		+ (0.04±0.11)*Rule3a	0.002	0.700	0.536	17.928	0.000
		+ (0.03±0.13)*Rule3d			0.537	11.608	0.000
3	B = 0.08±0.04	+ (0.87±0.16)*Rule2b	0.560	0.000	0.560	40.656	0.000
		+ (0.03±0.11)*Rule3a	0.001	0.755	0.561	19.806	0.000
		+ (0.02±0.13)*Rule3d			0.561	12.799	0.000
4	B = 0.07±0.04	+ (0.75±0.16)*Rule2b	0.517	0.000	0.517	34.203	0.000
		+ (0.07±0.13)*Rule3d	0.005	0.576	0.522	16.897	0.000
		+ (0.05±0.11)*Rule3a	0.004	0.608	0.526	11.089	0.000
5	B = 0.09±0.04	+ (0.96±0.02)*Rule2b	0.551	0.000	0.551	39.393	0.000
		- (0.04±0.14)*Rule3d	0.001	0.793	0.552	19.120	0.000
		+ (0.02±0.12)*Rule3a			0.553	12.362	0.000
6	B = 0.08±0.04	+ (0.85±0.17)*Rule2b	0.533	0.000	0.533	39.607	0.000
		+ (0.05±0.11)*Rule3a	0.004	0.619	0.557	19.467	0.000
		+ (0.03±0.13)*Rule3d	0.001	0.791	0.558	12.613	0.000

Notes: Rules 3b, 3c, and 2a concerned attributes that did not change within this particular melody.
Each equation represents the maximal equation when all rules are used to predict boundary location. Within each repetition, the rules that are above the dotted line are those rules that were included when a forward stepwise regression analysis was used. Rules below the dotted line were not included when the stepwise approach was used. The change in R^2 is provided for those occasions when adding a predicted was indicated by the stepwise solution.

Table 3.65 also provides the change R^2 for those terms that were included in the stepwise equation. Note that the statistics for the change in R^2 for the first term entered (always

Rule 2b) are the same as the statistics for R^2 for the entire equation at that point (it is the entire equation at that point). For each repetition the stepwise solution consisted of at least two rules, but only Rule 2b (Attack-Point) was consistently used. The second rule used was Rule 3a (Register Change) in Repetitions 2 and 3, but Rule 3d (Length Change) in Repetition 5. In Repetitions 1, 4 and 6 all the rules were incorporated within the stepwise solution.

As in Stage 4, because there was a high similarity between different repetitions, the analysis of the rules was repeated using the average of the boundary profiles from Repetitions 2 and 3 of Stage 6 and 8 (individually; see Table 3.66). To the extent that the results based on the average replicate the results based on the individual repetitions, the average can be used in future analyses. Generally, the results are similar to the previous analysis (Table 3.65), but note that only Rules 2b and 3a are important enough to be included in the stepwise equations.

Table 3.66: The regression equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) in the average boundary profile for the second and third repetitions, in each of Stages 6 and 8, Part 1.

Correlations Between The Average Profiles and The Rules			
	Shift	Stage 6	Stage 8
Rule 2b	+1	0.741**	0.747**
	0	-0.111	-0.121
	-1	-0.174	-0.161
Rule 3a	+1	0.309	0.310
	0	-0.115	-0.152
	-1	0.284	0.264
Rule 3d	+1	0.276	0.254
	0	0.163	0.168
	-1	-0.084	-0.079

Regression Equation: Stage 6						
Constant Rule Added at Step	ΔR^2	$p(R^2)$	R^2	F	$p(R^2)$
B = 0.17±0.08 + (1.70±0.33)*Rule2b	0.549	0.000	0.549	39.017	0.000
+ (0.08±0.22)*Rule3a	0.002	0.725	0.551	19.038	0.000
+ (0.05±0.26)*Rule3d					0.000
Regression Equation: Stage 8					
B = 0.17±0.08 + (1.81±0.34)*Rule2b	0.558	0.000	0.558	40.337	0.000
+ (0.08±0.22)*Rule3a	0.002	0.728	0.560	19.697	0.000
- (0.00±0.27)*Rule3d			0.560	12.707	0.000

Notes: Rules 2a, 3b and 3c concerned attributes that did not change within this melody.
See Table 3.65.

The previous similarity analyses implied that subjects are similar in their boundary profiles, but they are not identical. As in Stage 4, it is important to check if the similarity actually applies in the comparison of the theoretical and empirical boundary profiles. That is, given that the subjects are somewhat less cohesive in Stages 6 and 8 than in Stage 4, there is more chance that these subjects will produce differing relations with the rules in Stages 6 and 8. The simplest manner in which to accomplish this is to examine the relationship between the theoretical boundary profile and the individual boundary profile of each subject at each repetition.

As in Stage 4, because the boundary profile of any one subject at any one repetition is low in total information content, the individual boundary profile for Stage 6 was created for each subject by averaging that subjects second and third repetition. Similarly, the individual boundary profile for Stage 8 was created for each subject by averaging the second and third repetitions of that stage (the fifth and sixth repetitions overall). Given the previous consistency analysis, these averages are valid. The individual boundary profile for each subject was then compared to each of the rules.

The average correlations between subjects' individual boundary profiles and the theoretical boundaries were $r=0.63\pm0.13$ for Rule 2b, $r=0.26\pm0.10$ for Rule 3a and $r=0.22\pm0.14$ for Rule 3d (see Table 3.67).

Table 3.67: Statistics concerning the correlations between the individual boundary profiles and the prediction of boundaries given the rules of Lerdahl and Jackendoff in Stages 6 and 8, Part 1.

Descriptive Statistics Across Individual Subjects: Stage 6					
	Mean±SD	Median	Mode	Min	Max
Rule 2b	0.621±0.136	0.644	0.744	0.220	0.814
Rule 3a	0.263±0.110	0.261	0.285	0.024	0.536
Rule 3d	0.231±0.085	0.223	0.182	-0.011	0.402

Descriptive Statistics Across Individual Subjects: Stage 8					
	Mean±SD	Median	Mode	Min	Max
Rule 2b	0.629±0.128	0.666	0.744	0.331	0.867
Rule 3a	0.259±0.080	0.273	0.283	0.101	0.476
Rule 3d	0.211±0.175	0.182	0.182	0.068	0.504

Notes: Rules 2a, 3b, and 3c concerned attributes that did not change within this particular melody.
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

To finish this analysis, the correlation between the individual boundary profile and each rule was assumed to be a measure of utility (i.e., utilities for Rules 2b, 3a and 3d). That is, if a subject "used" the particular rule, then the correlation would be high. Conversely, if a subject did not "use" the rule, then the correlation would be low. If different subjects used different rules, then there would be no correlations between the utilities of the different rules. On the other hand, if subjects tended to use all rules in proportion (i.e., those who used Rule 2b also used Rule 3a and/or Rule 3d), then there would be a correlation between the utilities. The results indicated that those who used Rule 2b the most, also used Rule 3a the most (r=0.66 or 0.56 $p<0.01$). The same cannot be said for the Rules 2b and 3d or Rules 3a and 3d (see Table 3.68). Note that the correlation between Rules 2b and 3a is essentially the same as in Stage 4.

Table 3.68: The correlation across subjects between the utility of the different rules for predicting boundaries in the average boundary profile.

Correlations Between Measures: Stage 6			
	Rule 2b	Rule 3a	Rule 3d
Rule 2b	1.000	0.662**	-0.204
Rule 3a		1.000	0.108
Rule 3d			1.000

Correlations Between Measures: Stage 8			
	Rule 2b	Rule 3a	Rule 3d
Rule 2b	1.000	0.562**	0.066
Rule 3a		1.000	0.025
Rule 3d			1.000

Notes: Rules 2a, 3b, and 3c concerned attributes that did not change within this particular melody.
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

To complement this analysis, all subjects were ranked in the degree to which they used the different rules (i.e., the aforementioned utilities). The utilities were then plotted in order of rank (one plot for Rule 2b, one plot for Rule 3a and one plot for Rule 3d, for each stage). Figure 3.22, Panel A plots the value for Rule 2b (in order of rank) and overlays the corresponding values for Rule 3a and 3b. From this, one can see the correlation between the two rules. Figure 3.22, Panel B presents the value for Rule 3a (in order of rank) with the corresponding values for Rule 2b and 3a overlaid while Figure 3.22 Panel C presents the value for Rule 3c with Rules 2b and 3a overlaid. The correlations between the utilities can be seen in these plots (Stage 6 is in Figure 3.22a & Stage 8 is in Figure 3.22b). Note that the utilities for different rules do show some correlation (compare with Figure 3.18).

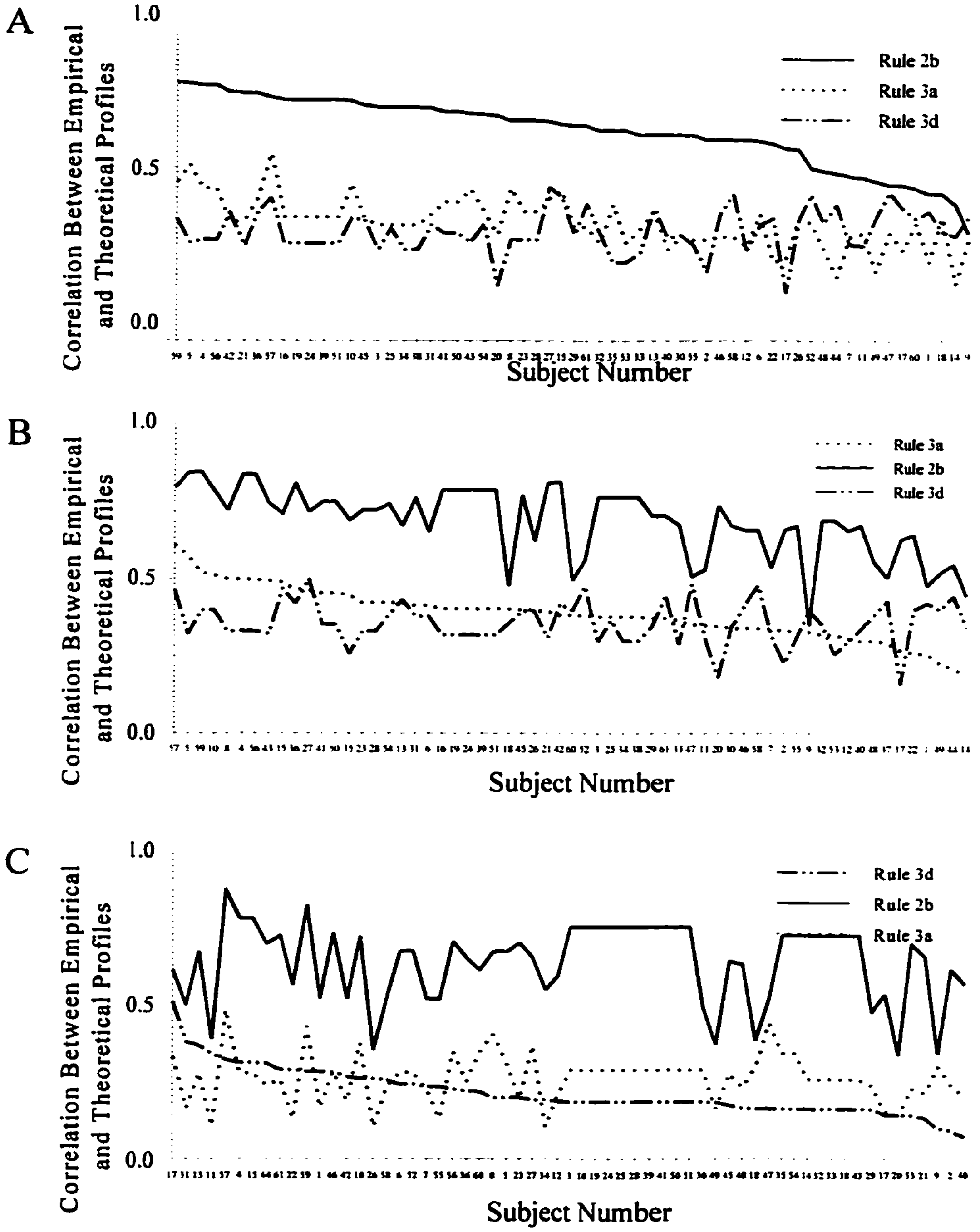


Figure 3.22a: The relationship between individual boundary profiles for Stage 6 and the theoretical rules of Lerdahl and Jackendoff (1983). In A, subjects are ranked by Rule 2b, in B, subjects are ranked by Rule 3a, and in C, subjects are ranked by Rule 3d. Note that until the last few subects, there is a gradual decrease in the use each rule.

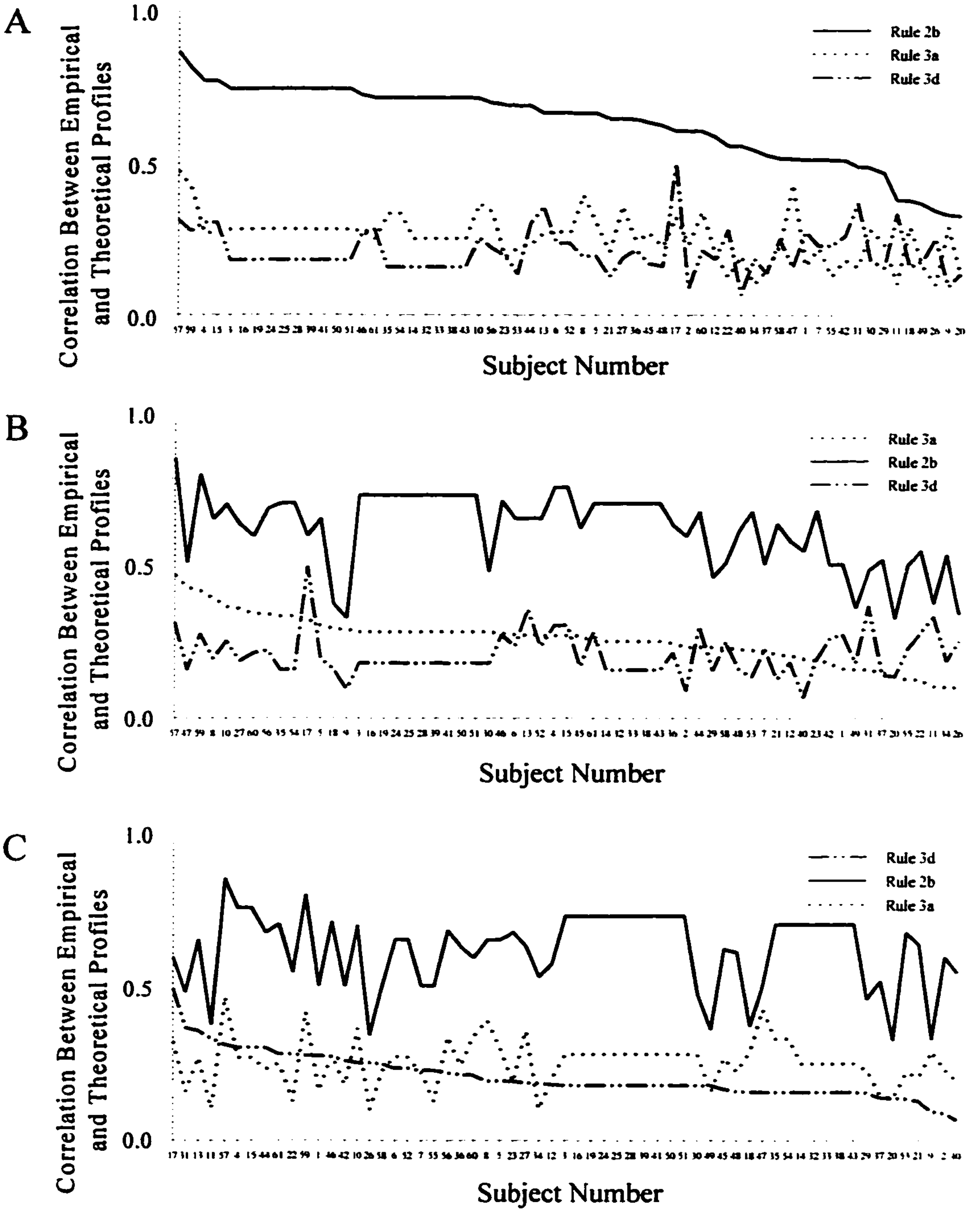


Figure 3.22b: The relationship between individual boundary profiles for Stage 8 and the theoretical rules of Lerdahl and Jackendoff (1983). In A, subjects are ranked by Rule 2b, in B, subjects are ranked by Rule 3a, and in C, subjects are ranked by Rule 3d. Note that until the last few subects, there is a gradual decrease in the use each rule.

In addition, these plots also demonstrate that there is no sharp discontinuity between the use of rules by different subjects. That is, these plots show a gradual decrease in the amount to which the rules correspond to individual subjects, except perhaps in the last few subjects.

Boundary Efficacy: Stages 6 & 8, Part 2

As with Stage 4, the fact that all subjects placed boundaries at, more or less, the same locations, simplifies the analysis of test sequences. Even if the meaning of a subjectively located boundary is uncertain (i.e., is it the end of a unit, or the beginning of the next unit?), it seems that all subjects position boundaries using similar subjective criteria. Given the rules, as in Stage 4, it would seem that subject indicate the boundaries at the end of units.

To recapitulate, to assess boundary efficacy using a recognition memory task, each subject was presented with 16 test sequences that varied on a number of dimensions: primarily seven Conditions (-1, 0, 1, 2, 3, 4 and 5) and two Types (Literal and Foil), but Foil Type was further divided into Location, Size and Tonality. For Foil sequences, Location varied over 1, 2, 3 and 4 while Size varied over 2, 1, -1 and -2 and Tonality varied over -2, -1, 0 and 1. Both Size and Location equal to 0 represented the Literal type. Tonality varied over -2 (a drop by two levels of tonality), -1 (a drop by one level of tonality), 0 (no change in the level of tonality), 1 (an increase by one level of tonality) and 2 (an increase by two levels of tonality). As in Stage 4, Part 2, Tonality was created by using the Krumhansl key-finding algorithm to assess the key of the entire melody (as presented: the key was c-major, as can be seen in Figure 3.21). To create the Foil, one note of a Literal sequences was changed. This change affected the role of that single note within the tonal hierarchy, and it was this change that was categorized (see Table 3.9). Note that Tonality changes of 2 were not possible for this melody since the melody did not contain any accidentals.

The analysis was based on 16 test sequences, in each of two stages, from 61 subjects: essentially 976 test sequences in each stage (1952 in both stages). Ideally, for each subject there were to have been four test sequences in Conditions 0 and 5, and two in each of Conditions 1, 2, 3 and 4, implying 244 (488), 122 (244), 122 (244), 122 (244), 122 (244) and 244 (488) test sequences in each of Conditions 0, 1, 2, 3, 5 and 5 respectively. From this, 33 (in Stage 6) and 11 (in Stage 8) test sequences were lost due to invalid responses (anticipatory responses or invalid key presses). In addition, of the 943 valid sequences that remained in Stage 6, 99 fell into Condition -1 (undesirable but useable), and 42 fell into Condition -2 (undesirable and unusable), leaving 198, 112, 114, 111, 112 and 155 in Conditions 0, 1, 2, 3, 4 and 5 respectively. Of the 965 valid sequences that remained in Stage 8, 92 fell into Condition -1 (undesirable but useable), and 47 fell into Condition -2 (undesirable and unusable), leaving 199, 115, 113, 114, 114 and 171 in Conditions 0, 1, 2, 3, 4 and 5 respectively. In addition, by design one half the sequences were to be Literals, and one half were to be Foils. Of the 943 valid sequences in Stage 6, 469 were Literals (475 when including Condition -2), and 432 were Foils (468 when including Condition -2). Of the 965 valid sequences in Stage 8, 475 were Literals (486 when including dropping Condition -2), and 443 were Foils (479 when including Condition -2). Again the slight imbalance in the Types is due to the act of creating Foil test sequences which resulted in test sequences more properly treated as Literal sequences from a different part of the melody. In general, the intended balancing worked.

As in Stage 4, the best insights are obtained from the raw data, but those insights are supported by a somewhat more complex analysis due to the added consideration of session number (Stage 6 verse Stage 8). As before, Condition, Size and Location, despite their apparent interval scaling, were analyzed as a categorical variables. The same six vectors as in Stage 4 were used to code Condition, one vector was used to code Type, the same three vectors were used to code Location, the same three vectors were used to code Size and same eight vectors were used to code Tonality. In addition, one further vector was used to code Session. The analysis followed the general layout used to analyze Stage 4, with the addition of consideration of Session.

The Analysis of Session, Condition, Type, and Their Interactions

One vector was used for the effect of Session. Six vectors (contrasts) for the effect of Condition in both the initial and secondary analyses, were created to compare (the contrasts are presented later with their analyses in Tables 3.72 and 3.73):

Boundary on Note 2 versus Boundary on Note 3
Boundary on Note 2 or 3 versus Boundary on Note 4
Boundary on Note 2 or 3 versus Boundary on Note 1
Boundary on Note 0 or 5 versus Boundary on Note 1, 2, 3 or 4
Boundary before Sequence versus Boundary After Sequence
Boundary beside Sequence versus Boundary Far from Sequence

Type was analyzed by a single vector. The interaction of Session and Condition required six vectors, Session and Type required one vector, Condition and Type required six vectors and the three-way interaction required six vectors. In addition to these effects, there were 60 subject vectors (using dummy coding) and four covariate vectors (first and second order for the serial position of the test sequence within the melody [with 0 as the centre]: first and second order for the serial position of the test sequence within the test phase [with 0 as the middle]). Using a hierarchical analysis, subject effects were entered first, followed by the covariates, the main effects of Session, Condition and Type, and finally the interactions. As in Stage 4, four analyses were performed, each representing a slightly different approach to the problem of the unbalanced design. The first used the interaction of the subject vectors with the effects vectors as the error term. The second analysis is the more common, simplified form of the first analysis, in which the residual is used as the error term for all effects (effectively, all subjectXeffects interactions are considered equivalent). The third analysis examined the incremental change in R^2 as the different terms were added to the equation. Finally, the fourth analysis tested the part correlations (i.e., the semipartials) for each effect. The fourth is often considered the most analogous to the balanced design (cf., Cohen & Cohen, 1994, Howell, 1995, Keppel & Zeddick, 1992).

As in Part 4, to check the integrity of the analysis, raw residuals, studentized residuals, Leverage, Cook's distance and Mahalanonbis' distance were examined. No outliers were detected. It should be noted that the covariate for melody position was significant while the covariate for sequence position was not.

Table 3.69 provides the data for the main effects of Session, Condition, and Type as well as their various interactions.

Table 3.69: The means, standard deviations and cell counts for Session, Condition, Type, SessionXCondition, SessionXType and ConditionXType, in Stages 6 & 8, Part 2.

Session, *Session X Condition*, **Session X Type**, Session X Condition X Type

Ses 1	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.83±.38	.84±.37	.86±.35	.73±.45	.92±.28	.83±.38	.76±.43	**.82±.39**
	64	70	58	59	60	58	100	**469**
Foil	.43±.50	.45±.50	.54±.50	.56±.50	.55±.50	.24±.43	.85±.36	**.51±.50**
	35	128	54	55	51	54	55	**432**
Cond	*.69±.47*	*.59±.49*	*.71±.46*	*.65±.48*	*.75±.44*	*.54±.50*	*.79±.41*	.67±.47
	99	*198*	*112*	*114*	*111*	*112*	*155*	901

Ses 2	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.75±.43	.87±.34	.83±.38	.73±.45	.92±.28	.85±.36	.74±.43	**.81±.40**
	61	62	60	59	60	60	113	**475**
Foil	.52±.51	.54±.50	.65±.48	.65±.48	.61±.49	.31±.47	.90±.31	**.59±.49**
	31	137	55	54	54	54	58	**443**
Cond	*.67±.47*	*.64±.48*	*.75±.44*	*.69±.46*	*.77±.42*	*.60±.49*	*.80±.40*	.70±.46
	92	*199*	*115*	*113*	*114*	*114*	*171*	918

Condition, **Type**, Condition X Type

Both	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.79±.41	.86±.35	.85±.36	.73±.45	.92±.28	.84±.37	.75±.43	**.81±.39**
	125	132	118	118	120	118	213	**944**
Foil	.47±.50	.50±.50	.60±.49	.61±.49	.58±.50	.28±.45	.88±.33	**.55±.50**
	66	265	109	109	105	108	113	**875**
Cond	*.68±.47*	*.62±.41*	*.73±.45*	*.67±.47*	*.76±.43*	*.57±.50*	*.79±.40*	.69±.46
	191	*397*	*227*	*225*	*225*	*226*	*326*	1819

The analyses are presented in Tables 3.70, 3.71, 3.72 with 3.73 (the last two for the contrast analysis). Session was not significant with a non-existent effect size (ΔR^2=0.00). Condition was significant, having an effect size that was slightly lower than in Stage 4

(ΔR^2=0.02). The major difference from Stage 4 was in the effect of Type (ΔR^2=0.09) with performance on Literals being 26% better than performance on Foils.

Table 3.70: The regression analysis of Session, Condition, Type, SessionXCondition, SessionXType, ConditionXType and SessionXConditionXType, in Stages 6 & 8, Part 2, using both the specific error term, F_{sp} and the pooled error term, F_{po}.

Hierarchical Regression Analysis: Anova Model

Effect	SS	df	MS	F_{sp}	$p(F_{sp})$	F_{po}	$p(F_{po})$
Subjects	23.792	60					
Melody Pos.	5.387	2	2.689	3.477	<0.050	15.383	<0.001
Sequence Pos.	3.798	2	1.899	2.456	<0.100	10.864	<0.001
Session	0.620	1	0.620	4.687	<0.050	3.547	<0.100
Condition	8.479	6	1.413	7.563	<0.001	8.084	<0.001
Type	33.188	1	33.188	109.441	<0.001	189.863	<0.001
Sess X Cond	0.311	6	0.052	0.456	>0.100	0.297	>0.100
Sess X Type	0.797	1	0.797	7.311	<0.010	4.559	<0.050
Cond X Type	12.056	6	2.009	14.683	<0.001	11.493	<0.001
SessXCondXType	0.307	6	0.051	0.595	>0.100	0.292	>0.100
Pooled Residual	301.901	1727	0.175				
SessXSubj	7.937	60	0.132				
CondXSubj	65.084	360	0.181				
TypeXSubj	18.534	60	0.309				
SessXCondXSubj	40.841	360	0.113				
SessXTypeXSubj	6.396	60	0.107				
CondXTypeXSubj	49.657	360	0.138				
SesXConXTypXSub	30.717	360	0.085				
Residual	82.735	107	0.773				
Total	390.637	1818					

Notes: Effects are listed in their order of entry into the analysis.
For F_{sp}, the error term is the interaction of the effect with subjects.
For F_{po}, the error term is the residual.

Table 3.71: The regression analysis of Condition, Type and ConditionXType in Stages 6 and 8, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model							
Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.061	60					
Melody Pos.	0.014	2	13.085	0.000			
Sequence Pos.	0.010	2	9.312	0.000			
Session	0.002	1	3.043	0.081	0.001	1.339	0.247
Condition	0.022	6	7.083	0.000	0.020	7.404	0.000
Type	0.085	1	183.737	0.000	0.015	34.082	0.000
Sess X Cond	0.001	6	0.286	0.944	0.000	0.125	0.993
Sess X Type	0.002	1	4.411	0.036	0.000	0.065	0.798
Cond X Type	0.031	6	11.523	0.000	0.031	11.487	0.000
SessXCondXType	0.001	6	0.293	0.940	0.001	0.293	0.940
Pooled Residual	0.773	1727					
SessXSubj	0.026	60					
CondXSubj	0.166	360					
TypeXSubj	0.047	60					
SessXCondXSubj	0.104	360					
SessXTypeXSubj	0.016	60					
CondXTypeXSubj	0.127	360					
SesXConXTypXSub	0.079	360					
Residual	0.212	107					
Total	1.000	1818					

Notes: Effects are listed in the order in which they enter the equation.
For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.
For R^2_{part}, the error term includes all effects in the full model.

Of the two-way interactions (Tables 3.70 & 3.71), the SessionXCondition was not significant, but SessionXType and ConditionX Type were significant. For SessionXType (collapsed over Condition), performance on Foils increased from Stage 6 to Stage 8 (0.51±0.50 to 0.59±0.49) while performance on Literals did not change (0.82±0.39 to 0.81±0.40). For the ConditionXType (collapsed over Sessions), performance on Literals was fairly consistent dropping off when the boundary followed the sequence. However, performance for Foils was best when the boundary followed the sequence. For Foils, performance fell below change for test sequences that were far from boundaries (Condition -1) and for test sequences that ended on boundaries (Condition 4). The three-way interaction of Session, Condition and Type was not significant. In the remaining discussion, Session is not detailed because it had minimal main effect and only one interaction.

Contrast analysis is present in Tables 3.72 and 3.73. The only contrasts for condition that were not significant were Contrasts 3 and 6. Condition Contrast 3 implied that boundaries on the first, second, or third notes of the extract were of equal difficulty. The interaction of Condition Contrast 3 with Type indicated that this equivalence was the same for both Literals and Foils. Contrast 6 for Condition implied that boundaries outside the test sequence were all equally difficult; This contrast did not interact with Type either. The remaining contrasts and their interactions demonstrated that Session was not important. Beyond that, the Condition Contrast 2 implied that boundaries on Note 4 produced different performance from boundaries on Notes 2 and 3 (and hence, Note 1, by implication). This pattern changed somewhat for Literals and Foils, but not dramatically (see Table 3.69). However, boundaries on Notes 2 and 3 produced different performance (Condition Contrast 1), which changed for Literals and Foil (again, means in Table 3.75 provide the clearer picture). The effect of Contrast 2 was due to the fact that Note 4 produced much lower performance than all other conditions. Boundaries on notes within the test sequence produced different performance from boundaries on notes adjacent to the sequence (Condition Contrast 4). However, boundaries on notes before the sequence produced different performance from boundaries on notes after the test sequence (Condition Contrast 5). Hence, the effect of Condition Contrast 4 was likely due to the inclusion of boundaries on Note 4. In sum, no distinct pattern emerges, except that the various conditions and combinations were different.

Table 3.72: The vector (contrast) analysis on the effect of Session, Condition, and Type in Stages 6 & 8, Part 2. Both the vectors (contrasts) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are listed. The *F* values represent tests of the part correlations.

Contrast	Sess 1	Sess 2
Means	.67±.47	.70±.46
Sess Ψ_1	1	-1

Contrast	Far	Before	Note 1	Note 2	Note 3	Note 4	After
Means	.68±.47	.62±.49	.73±.45	.67±.47	.76±.43	.57±.50	.79±.40
Cond Ψ_1	0	0	0	1	-1	0	0
Cond Ψ_2	0	0	0	1	1	-2	0
Cond Ψ_3	0	0	-2	1	1	0	0
Cond Ψ_4	0	-1	1	1	1	1	-1
Cond Ψ_5	0	1	0	0	0	0	-1
Cond Ψ_6	-2	1	0	0	0	0	1

Contrast	Literal	Foil
Means	.81±.39	.55±.50
Type Ψ_1	1	-1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Sess Ψ_1	-0.035	-0.024	-0.028	1.339	0.247
Cond Ψ_1	-0.048	-0.051	-0.058	5.840	0.016
Cond Ψ_2	0.089	0.057	0.065	7.336	0.007
Cond Ψ_3	-0.008	-0.002	-0.002	0.008	0.927
Cond Ψ_4	0.015	0.056	0.064	7.012	0.008
Cond Ψ_5	-0.118	-0.047	-0.054	5.003	0.025
Cond Ψ_6	0.007	-0.035	-0.040	2.746	0.098
Type Ψ_1	0.280	0.123	0.130	34.082	0.000

Notes: The interaction contrasts are presented in the next table; their contributions are included in the calculations of the part and partial correlations.

The fourth contrast on Condition is actually balanced (approximately) when n is considered.

Table 3.73: The vector (contrast) analysis on the interactions of Session, Condition and Type Stages 6 & 8, Part 2. The vectors (contrasts) used are the products of the vectors presented in Table 3.72. The simple correlations, part (semipartial) correlations, and partial correlations are listed. The *F* values represent tests of the part correlations.

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
SessΨ_1 X CondΨ_1	-0.004	-0.000	-0.000	0.000	0.992
SessΨ_1 X CondΨ_2	0.005	0.006	0.007	0.086	0.770
SessΨ_1 X CondΨ_3	0.003	0.002	0.002	0.007	0.933
SessΨ_1 X CondΨ_4	-0.007	0.004	0.005	0.040	0.842
SessΨ_1 X CondΨ_5	-0.016	-0.014	-0.016	0.418	0.518
SessΨ_1 X CondΨ_6	-0.013	0.003	0.004	0.022	0.882
SessΨ_1 X TypeΨ_1	0.047	0.005	0.006	0.066	0.798
CondΨ_1 X TypeΨ_1	-0.059	-0.086	-0.097	16.425	0.000
CondΨ_2 X TypeΨ_1	-0.098	-0.080	-0.091	14.318	0.000
CondΨ_3 X TypeΨ_1	-0.007	-0.001	0.002	0.004	0.949
CondΨ_4 X TypeΨ_1	-0.099	-0.056	-0.064	7.050	0.008
CondΨ_5 X TypeΨ_1	0.154	0.106	0.121	25.826	0.000
CondΨ_6 X TypeΨ_1	0.157	0.019	0.022	0.848	0.357
SessΨ_1XCondΨ_1XTypeΨ_1	0.004	0.002	0.002	0.005	0.942
SessΨ_1XCondΨ_2XTypeΨ_1	0.004	0.009	0.010	0.165	0.685
SessΨ_1XCondΨ_3XTypeΨ_1	-0.011	-0.015	-0.016	0.489	0.484
SessΨ_1XCondΨ_4XTypeΨ_1	0.006	-0.022	-0.025	1.059	0.304
SessΨ_1XCondΨ_5XTypeΨ_1	0.010	0.001	0.001	0.001	0.979
SessΨ_1XCondΨ_6XTypeΨ_1	0.040	0.022	0.025	1.086	0.297

Notes: The SessionXCondition contrasts are the Session contrast multipled by each Condition contrast. The SessionXType contrast is the product of the Session and Type contrasts. The ConditionXType contrasts are each Condition contrast multiplied by the Type contrast. The SessionXConditionXType are the three-way products.

The main effect contrasts are presented in the previous table; their contributions are included in the computations of the part and partial correlations.

Generally, the results implied that there is some bias to say "yes" to all test sequences, but this is not the whole story. The correlation between corresponding conditions of the Literal and Foil Types was only r=-0.40. Although this implies that those conditions that subjects tended to say "yes" to as Literals (hence, correct) were also the conditions that subjects tended to say "yes" to as Foils (hence, incorrect), the correlation implies that the Literals only explain 2% of the variance in Foils. The difference between Literals and Foils as a function of condition is presented in Table 3.74. Note that Condition 4 (boundaries on Note 4) demonstrates the highest differential value indicating a strong response bias. That is, when there is a boundary on the last note of the sequence, subjects seem to say "yes" (it was in the melody) most of the time. Conversely, when the boundary is in the middle of the sequence (Condition 2: Note 2) or after the sequence (Condition 5: After), subjects are likely to be more accurate (i.e., their responses to Literals and Foils are equally accurate). This, of course, complicates the interpretation of the Condition effect (e.g., if it is response bias, one must explain why there more response bias in a particular condition). Furthermore, it seems to be different from that of Stage 4, although the correponding effects in Stage 4 were quite small.

Table 3.74: The differences between Conditions within Literal and Foil Types (collapsed over sessions).

Both	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.79±.41	.86±.35	.85±.36	.73±.45	.92±.28	.84±.37	.75±.43	.81±.39
	125	132	118	118	120	118	213	944
Foil	.47±.50	.50±.50	.60±.49	.61±.49	.58±.50	.28±.45	.88±.33	.55±.50
	66	265	109	109	105	108	113	875
Differ	.32±.46	.36±.43	.35±.45	.12±.47	.34±.41	.56±.41	-.13±.38	.26±.45

The Analyses of Location, Size and Tonality and Their Interactions

The analysis of Location, Size and Tonality was essentially the same as the analysis of these variables in Stage 4. That is, each was analyzed in isolation (as an addition to the previous Session, Condition and Type), and then all were analyzed in combination. Only the regression approach is provided for reasons cited previously; the SStot is provided for each analysis for those who might wish to convert to the more usual ANOVA table.

Table 3.75 provide the effects of Location (with Literals coded as a Size of 0) and the interaction between Condition and Location.

Table 3.75: The means, standard deviations and cell counts for Condition, Location (Literals correspond to Location equal 0) and ConditionXLocation.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Loc
Liter	.79±.41	.86±.35	.85±.36	.73±.45	.92±.28	.84±.37	.75±.43	.81±.39
	125	132	118	118	120	118	213	944
Loc 1	.25±.44	.37±.48	.41±.51	—	.63±.52	.19±.39	.67±.52	.32±.47
	20	149	17	0	8	82	6	282
Loc 2	.60±.50	.50±.55	.33±.58	.00±.00	—	.25±.50	.88±.33	.78±.42
	20	6	3	2	0	4	100	135
Loc 3	.50±.51	.78±.43	.64±.48	.62±.49	—	.40±.55	—	.62±.49
	18	18	89	107	0	5	0	237
Loc 4	.63±.52	.65±.48	—	—	.58±.50	.71±.47	1.0±.00	.63±.48
	8	92	0	0	97	17	7	221
Cond	.68±.47	.62±.49	.73±.45	.67±.47	.76±.43	.57±.50	.79±.40	.69±.46
	191	397	227	227	225	226	326	1819

Location (considering the three vectors that coded for differences between Locations 1, 2, 3 and 4; i.e., not including Literals) did produce significant differences (see Table 3.76) with an effect size comparable to that of Condition (ΔR^2=0.02).

Table 3.76: The regression analysis of Session, Condition, Type and their interactions with the addition of Location (of change), SessionXLocation and ConditionXLocation, in Stages 6 & 8, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.061	60					
Melody Pos.	0.014	2	13.085	0.000			
Sequence Pos.	0.010	2	9.312	0.000			
Session	0.002	1	3.043	0.081	0.001	0.911	0.340
Condition	0.022	6	7.083	0.000	0.004	1.493	0.177
Type	0.085	1	183.737	0.000	0.006	14.275	0.000
Sess X Cond	0.001	6	0.286	0.944	0.001	0.518	0.795
Sess X Type	0.002	1	4.411	0.036	0.000	0.006	0.936
Cond X Type	0.031	6	11.523	0.000	0.010	3.748	0.001
SessXCondXType	0.001	6	0.293	0.940	0.002	0.691	0.657
Location	0.018	3	13.987	0.000	0.003	2.196	0.087
Sess X Loc	0.002	3	1.843	0.137	0.001	1.094	0.350
Cond X Loc	0.007	12	1.312	0.205	0.007	1.311	0.205
Pooled Residual	0.745	1709					
Total	1.000	1818					

$SS_{tot} = 390.637$

Notes: Effects are listed in the order in which they enter the equation.
For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.
For R^2_{part}, the error term includes all effects in the full model.

The implication from the contrast analyses (Table 3.77) is that changes on the first note of the sequence seriously impaired performance, while other changes in other positions were not so detrimental. This is the same pattern as observed in Stage 4. Neither interaction was significant. As noted in Table 3.75, it is important to realize that since Location was not explicitly controlled, some cells in the interaction had no test sequences. Hence the mean preformance was not necessarily the 0 that appears in the table.

Table 3.77: The vector (contrast) analysis on the effect of Location in Stages 6 & 8. Part 2. Both the vectors (contrasts; Condition vectors have been presented previously) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are listed. The *F* values represent tests of the part correlations.

Contrast	Note 1	Note 2	Note 3	Note 4
Means	.32±.47	.78±.42	.62±.49	.63±.48
Loc Ψ_1	1	-1	0	0
Loc Ψ_2	0	0	1	-1
Loc Ψ_3	1	1	-1	-1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Loc Ψ_1	-0.289	0.032	0.035	2.061	0.071
Loc Ψ_2	-0.007	-0.029	-0.033	1.849	0.174
Loc Ψ_3	-0.109	-0.049	-0.056	5.460	0.013
SessΨ_1 X LocΨ_1	-0.013	0.018	0.021	0.757	0.385
SessΨ_1 X LocΨ_2	-0.037	-0.034	-0.039	2.615	0.106
SessΨ_1 X LocΨ_3	0.026	-0.005	-0.006	0.057	0.812
LocΨ_1 X CondΨ_1	0.030	-0.005	-0.006	0.057	0.812
LocΨ_1 X CondΨ_2	0.219	0.037	0.043	3.161	0.076
LocΨ_1 X CondΨ_3	0.043	0.025	0.029	1.438	0.231
LocΨ_1 X CondΨ_4	-0.061	-0.038	-0.044	3.341	0.068
LocΨ_1 X CondΨ_5	-0.090	0.006	0.007	0.079	0.778
LocΨ_1 X CondΨ_6	-0.197	-0.051	-0.059	5.869	0.016
LocΨ_2 X CondΨ_5	0.033	0.044	0.051	4.416	0.036
LocΨ_2 X CondΨ_6	-0.010	0.008	0.010	0.161	0.688
LocΨ_3 X CondΨ_2	-0.209	-0.007	-0.008	0.108	0.742
LocΨ_3 X CondΨ_4	-0.005	0.016	0.018	0.569	0.451
LocΨ_3 X CondΨ_5	-0.167	-0.011	-0.012	0.253	0.615
LocΨ_3 X CondΨ_6	-0.088	0.020	0.024	0.957	0.328

Notes: The interaction contrasts are each Condition contrast multiplied by each Location contrast.

Table 3.78 provide the effects of Size (with Literals coded as a Size of 0).

Table 3.78: The means, standard deviations and cell counts for Condition, Size (Literals correpond to Size equal 0) and ConditionXSize, in Stages 6 & 8, Part 2.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Size
Liter	.79±.41	.86±.35	.85±.36	.73±.45	.92±.28	.84±.37	.75±.43	.81±.39
	125	132	118	118	120	118	213	944
Size-2	.44±.51	.73±.47	.64±.48	—	.75±.46	—	.43±.53	.61±.49
	25	15	89	0	8	0	7	144
Size-1	.55±.51	.47±.50	.20±.42	—	—	.50±.52	.91±.28	.67±.47
	20	70	10	0	0	12	93	205
Size 1	.00±.00	.49±.50	.60±.52	.61±.49	.57±.50	.33±.52	1.0±.00	.55±.50
	2	177	10	109	97	6	7	408
Size 2	.47±.51	.33±.58	—	—	—	.24±.43	.67±.52	.31±.46
	19	3	0	0	0	90	6	118
Cond	.68±.47	.62±.49	.73±.45	.67±.47	.76±.43	.57±.50	.79±.40	.69±.46
	191	397	227	227	225	226	326	1819

Size, considering the three vectors that coded for differences between Sizes -2, -1, 1 and 2 (i.e., not including Literals), did not produce significant differences (see Table 3.79): neither did the interaction of Size with Session or Condition. Note that Sizes -2 and 2 (i.e., the greatest change) did not generally produce the best performance. The effects of Size are essentially the same as those that occurred in Stage 4 (in Stage 4, the ConditionXSize was significant; here it is merely the largest effect).

Table 3.79: The regression analysis of Session, Condition, Type and their interactions with the addition of Size (of change), SessionXSize and ConditionXSize, in Stages 6 & 8, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.061	60					
Melody Pos.	0.014	2	13.085	0.000			
Sequence Pos.	0.010	2	9.312	0.000			
Session	0.002	1	3.043	0.081	0.002	3.571	0.059
Condition	0.022	6	7.083	0.000	0.006	2.412	0.025
Type	0.085	1	183.737	0.000	0.007	15.449	0.000
Sess X Cond	0.001	6	0.286	0.944	0.001	0.525	0.790
Sess X Type	0.002	1	4.411	0.036	0.001	1.313	0.252
Cond X Type	0.031	6	11.523	0.000	0.015	5.648	0.000
SessXCondXType	0.001	6	0.293	0.940	0.001	0.405	0.876
Size	0.004	3	2.619	0.049	0.004	2.731	0.043
Sess X Size	0.001	3	0.968	0.407	0.001	1.215	0.303
Cond X Size	0.008	11	1.713	0.065	0.008	1.713	0.065
Pooled Residual	0.760	1710					
Total	1.000	1818					

$SS_{tot} = 390.637$

Notes: Effects are listed in the order in which they enter the equation.
For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.
For R^2_{part}, the error term includes all effects in the full model.

Contrasts (Table 3.80) indicated that decreases in size (i.e., a change that lowered the pitch height) were different from increases in size, but small changes were not different from large changes. The interactions of Session by Size and Condition by Size were not significant. It is important to realize that since Size was not explicitly controlled, some cells in the interaction had no test sequences. Their means are not necessarily the 0 that appears in the table.

Table 3.80: The vector (contrast) analysis on the effect of Size, in Stages 6 & 8, Part 2. Both the vectors (contrasts; Condition vectors have been presented previously) used and their corresponding simple correlations, part (semipartial) correlations, and partial correlations are provided. The *F* values represent tests of the part correlations.

Contrast	2	1	-1	-2
Means	.61±.49	.67±.47	.55±.50	.31±.46
Size Ψ_1	1	0	0	-1
Size Ψ_2	0	1	-1	0
Size Ψ_3	1	-1	-1	1

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Size Ψ_1	0.107	0.043	0.049	4.104	0.043
Size Ψ_2	0.112	0.043	-0.050	4.224	0.040
Size Ψ_3	0.010	-0.028	-0.032	1.766	0.184
SessΨ_1 X SizeΨ_1	-0.019	0.007	0.008	0.103	0.749
SessΨ_1 X SizeΨ_2	0.017	-0.031	-0.036	2.198	0.138
SessΨ_1 X SizeΨ_3	-0.003	-0.029	-0.033	1.871	0.172
SizeΨ_1 X CondΨ_1	-0.009	0.036	-0.042	2.976	0.084
SizeΨ_1 X CondΨ_4	-0.124	0.007	0.008	0.099	0.753
SizeΨ_1 X CondΨ_5	0.031	0.016	0.018	0.572	0.450
SizeΨ_1 X CondΨ_6	-0.015	-0.032	-0.037	2.356	0.125
SizeΨ_2 X CondΨ_2	0.067	-0.047	-0.054	4.983	0.026
SizeΨ_2 X CondΨ_3	0.090	-0.025	-0.028	1.357	0.244
SizeΨ_2 X CondΨ_5	0.002	0.025	0.028	1.383	0.240
SizeΨ_2 X CondΨ_6	0.087	0.046	0.053	4.762	0.029
SizeΨ_3 X CondΨ_4	0.068	-0.043	-0.050	4.222	0.040
SizeΨ_3 X CondΨ_5	0.197	0.043	0.049	4.085	0.043
SizeΨ_3 X CondΨ_6	0.025	0.029	0.034	1.930	0.165

Notes: The interaction contrasts are each Condition contrast multiplied by each Size contrast.

Table 3.81 provide the effects of Tonality (see Tables 3.82, and Table 3.83 for the individual contrasts).

Table 3.81: The means, standard deviations and cell counts for Condition, Tonality (Literals correspond, arbitrarily, to Tonality equals 0) and ConditionXTonality.

	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Tonal
Liter	.79±.41	.86±.35	.85±.36	.73±.45	.92±.28	.84±.37	.75±.43	.81±.39
	125	132	118	118	120	118	213	944
Ton 1	.50±.52	.49±.50	.44±.53	.60±.55	1.0±.00	.22±.42	.78±.43	.42±.50
	14	71	9	5	6	82	18	205
Ton 2	.57±.51	.49±.50	.72±.46	.61±.49	.53±.50	.83±.41	.93±.26	.62±.49
	21	162	18	102	85	6	83	477
Ton 3	.23±.44	.56±.51	.54±.51	50.±.71	.83±.41	.33±.53	1.0±.00	.51±.50
	13	25	26	2	6	9	3	84
Ton 5	.00±.00	—	.61±.49	—	1.0±.00	—	.25±.50	.57±.50
	2	0	54	0	1	0	4	61
Ton 7	.56±.51	.43±.53	.50±.71	—	.57±.53	.36±.50	.80±.45	.52±.50
	16	7	2	0	7	11	5	48
Cond	.68±.47	.62±.49	.73±.45	.67±.47	.76±.43	.57±.50	.79±.40	.69±.46
	191	397	227	227	225	226	326	1819

Notes: Ton 1 = Triad to non-diatonic
Ton 2 = Diatonic, non-triad to non-diatonic
Ton 3 = Triad to diatonic, non-triad
Ton 5 = Diatonic, non-triad to diatonic, non-triad
Ton 7 = Diatonic, non-triad to triad

Considering the four vectors (of the eight that represented the full set) that coded for changes (i.e., not including Literals), Tonality did not produce significant differences.

Table 3.82: The regression analysis of Session, Condition, Type and their interactions with the addition of Tonality (of change), SessionXTonality and ConditionXTonality, in Stages 6 & 8, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.061	60					
Melody Pos.	0.014	2	13.085	0.000			
Sequence Pos.	0.010	2	9.312	0.000			
Session	0.002	1	3.043	0.081	0.001	1.611	0.205
Condition	0.022	6	7.083	0.000	0.009	3.205	0.004
Type	0.085	1	183.737	0.000	0.002	4.862	0.028
Sess X Cond	0.001	6	0.286	0.944	0.000	0.148	0.989
Sess X Type	0.002	1	4.411	0.036	0.000	0.311	0.577
Cond X Type	0.031	6	11.523	0.000	0.014	5.404	0.000
SessXCondXType	0.001	6	0.293	0.940	0.001	0.261	0.955
Tonality	0.001	4	0.666	0.616	0.002	1.254	0.281
Sess X Tonality	0.001	4	0.719	0.579	0.001	0.520	0.721
Cond X Tonality	0.015	20	1.730	0.023	0.015	1.730	0.023
Pooled Residual	0.755	1699					
Total	1.000	1818					

$SS_{tot} = 390.637$

Notes: Effects are listed in the order in which they enter the equation.
For ΔR^2, the error term includes only those effects that have been included in the equation up to that point.
For R^2_{part}, the error term includes all effects in the full model.

Unlike Stage 4, Part 2, Tonalities of 1 or 2 (i.e., movement that terminated on a non-diatonic note) did not generally produce the best performance. The interaction of SessionXTonality was not significant although the interaction of ConditionXTonality was significant (see Table 3.82), but this was due to only one contrast of questionable utility (Table 3.83: sequences extracted before and after boundaries had different responses to Tonality 1 and 2).

Table 3.83: The vector (contrast) analysis on the effect of Tonality, in Stages 6 & 8. Part 2.

Contrast	1	2	3	4	5	6	7	8	9
Means	.42±.50	.62±.49	.51±.50	—	.57±.50	—	.52±.50	—	—
Ton Ψ_1	1	-1	0	0	0	0	0	0	0
Ton Ψ_2	1	0	-1	0	0	0	0	0	0
Ton Ψ_4	1	0	0	0	-1	0	0	0	0
Ton Ψ_6	1	0	0	0	0	0	-1	0	0

Contrast	simple	part	partial	*F*(part)	*p*(*F*)
Ton Ψ_1	-0.040	0.001	0.001	0.001	0.981
Ton Ψ_2	-0.118	-0.021	-0.024	1.000	0.317
Ton Ψ_4	-0.149	0.028	0.033	1.827	0.177
Ton Ψ_6	-0.150	-0.032	-0.037	2.345	0.126
SessΨ_1 X TonΨ_1	0.013	-0.021	-0.024	0.991	0.320
SessΨ_1 X TonΨ_2	0.012	0.017	0.019	0.630	0.428
SessΨ_1 X TonΨ_4	0.001	-0.006	-0.007	0.093	0.761
SessΨ_1 X TonΨ_6	-0.011	0.002	0.002	0.009	0.924
TonΨ_1 X CondΨ_2	0.232	0.060	0.068	7.991	0.005
TonΨ_1 X CondΨ_3	0.085	0.019	0.022	0.800	0.371
TonΨ_1 X CondΨ_4	0.036	0.009	0.011	0.188	0.664
TonΨ_1 X CondΨ_5	0.099	0.015	0.018	0.536	0.464
TonΨ_1 X CondΨ_6	-0.003	-0.031	-0.036	2.205	0.138
TonΨ_2 X CondΨ_1	-0.012	0.016	-0.019	0.597	0.440
TonΨ_2 X CondΨ_2	0.187	0.006	0.007	0.073	0.787
TonΨ_2 X CondΨ_3	-0.010	0.014	0.016	0.438	0.508
TonΨ_2 X CondΨ_4	0.071	-0.043	-0.049	4.126	0.042
TonΨ_2 X CondΨ_5	-0.053	0.014	0.016	0.461	0.497
TonΨ_2 X CondΨ_6	-0.011	0.030	-0.035	2.071	0.150
TonΨ_4 X CondΨ_1	-0.029	0.033	0.037	2.389	0.122
TonΨ_4 X CondΨ_3	-0.009	0.027	0.031	1.605	0.205
TonΨ_4 X CondΨ_5	-0.091	-0.008	-0.009	0.139	0.709
TonΨ_6 X CondΨ_1	-0.037	-0.043	-0.050	4.174	0.041
TonΨ_6 X CondΨ_2	0.187	-0.035	-0.040	2.705	0.100
TonΨ_6 X CondΨ_3	0.038	-0.028	-0.032	1.790	0.181
TonΨ_6 X CondΨ_4	0.079	0.031	0.036	2.196	0.139
TonΨ_6 X CondΨ_5	-0.066	0.009	0.010	0.186	0.666
TonΨ_6 X CondΨ_6	-0.042	0.016	0.019	0.590	0.442

Notes: Only those contrast that are relevant are shown.

As noted for both Location and Size, it is important to realize that since Tonality was not controlled (implicitly or explicitly), some cells in the interaction had no test sequences. Their means are not necessarily the 0 that appears in the table. A

The last analysis (Tables 3.84) was the complete analysis that included Session, Condition, Type, their second-order and third-order interactions, Location, Size and Tonality, and their second-order interactions with Session and Condition, in that order, analogous to that which was done for Stage 4, Part 2. Tonality was entered last because it was not controlled to the same extent as Location or Size. Size was entered just before Tonality (rather than first) because Size and Tonality represented a competition of sorts (Size and Tonality both contain aspects of frequency or pitch, while Location is separate from the two). In this final analysis, the subject vectors were entered first, the covariates second, Session third, Condition fourth, Type fifth, TypeXCondition sixth and then Location, Size and Tonality (since they represent a finer breakdown of one level of Type, and as such are logically orthogonal to Type and the ConditionXType interaction), and their interactions with Session and Condition. Other two-way interactions (e.g., Location by Size, Location by Tonality & Size by Tonality) and the three-way interactions (or higher) were not analyzed, nor presented, due to an overabundance of empty cells. The same contrasts (vectors) for Session, Condition, Type, ConditionXType, Location, Size, Tonality, SessionXLocation, ConditionXLocation, SessionXSize, ConditionXSize, SessionXTonality and ConditionXTonality were used.

In this overall analysis, Condition, ConditionXType and Size emerge as the significant factors, with some consideration of ConditionXSize. This is most clearly seen in the analysis of the part correlation for the effects (i.e., R^2_{part}): This analysis is thought to be the most comparable to that of the balanced ANOVA.

Table 3.84: The complete regression analysis of Session, Condition, Type, Location, Size, Tonality and their interactions, in Stages 6 & 8, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{\text{part}})$	$p(R^2_{\text{part}})$
Subjects	0.061	60					
Melody Pos.	0.014	2	13.085	0.000			
Sequence Pos.	0.010	2	9.312	0.000			
Session	0.002	1	3.043	0.081	0.000	0.415	0.519
Condition	0.022	6	7.083	0.000	0.006	2.231	0.038
Type	0.085	1	183.737	0.000	0.001	2.176	0.140
Sess X Cond	0.001	6	0.286	0.944	0.001	0.521	0.793
Sess X Type	0.002	1	4.411	0.036	0.000	0.001	0.970
Cond X Type	0.031	6	11.523	0.000	0.006	2.147	0.046
SessXCondXType	0.001	6	0.293	0.940	0.001	0.553	0.768
Location	0.018	3	13.987	0.000	0.001	0.843	0.470
Size	0.003	3	2.623	0.049	0.003	2.539	0.055
Tonality	0.003	4	1.860	0.115	0.002	1.160	0.327
Sess X Loc	0.003	3	2.041	0.106	0.002	1.358	0.254
Cond X Loc	0.007	12	1.267	0.232	0.008	1.611	0.082
Sess X Size	0.001	3	0.612	0.608	0.000	0.155	0.926
Cond X Size	0.007	8	1.900	0.056	0.006	1.816	0.070
Sess X Ton	0.001	4	0.490	0.723	0.001	0.330	0.858
Cond X Ton	0.010	20	1.150	0.290	0.010	1.150	0.290
Pooled Residual	0.720	1667					
Total	1.000	1818					

Notes: The effects are listed in the order in which they entered the equation.

When significances were tested hierarchically, Condition, Type, SessionXType, ConditionXType, Location, Size, and ConditionXSize all increased the predicability of the equation. In comparison with Stage 4, Part 2, the main difference has been the loss of the significance of Tonality (recall that this was an unfamiliar melody with a weaker tonal center, as defined by the Krumhansl key-finding algorithm [Frankland & Cohen, 1996; Krumhansl, 1990]).

As was done with the previous melody, Figure 3.23 presents the melody, the empirically determined boundaries and the results for test sequences as a function of the extraction point of the sequence from the melody in order to provide an alternative perspective on the performance of subjects. Again, before averaging different subjects. the mean of each subject was subtracted from each score of that subject (equivalent in function to the within-subjects analysis; the Subjects effect was large enough to require consideration). This presentation collapses over Session and includes Condition -2. Although this depiction captures elements of the aforementioned Condition, it must be remembered that it is not the same as Condition, particularly when subjects do not show a high degree of cohesion with respect to boundary location.

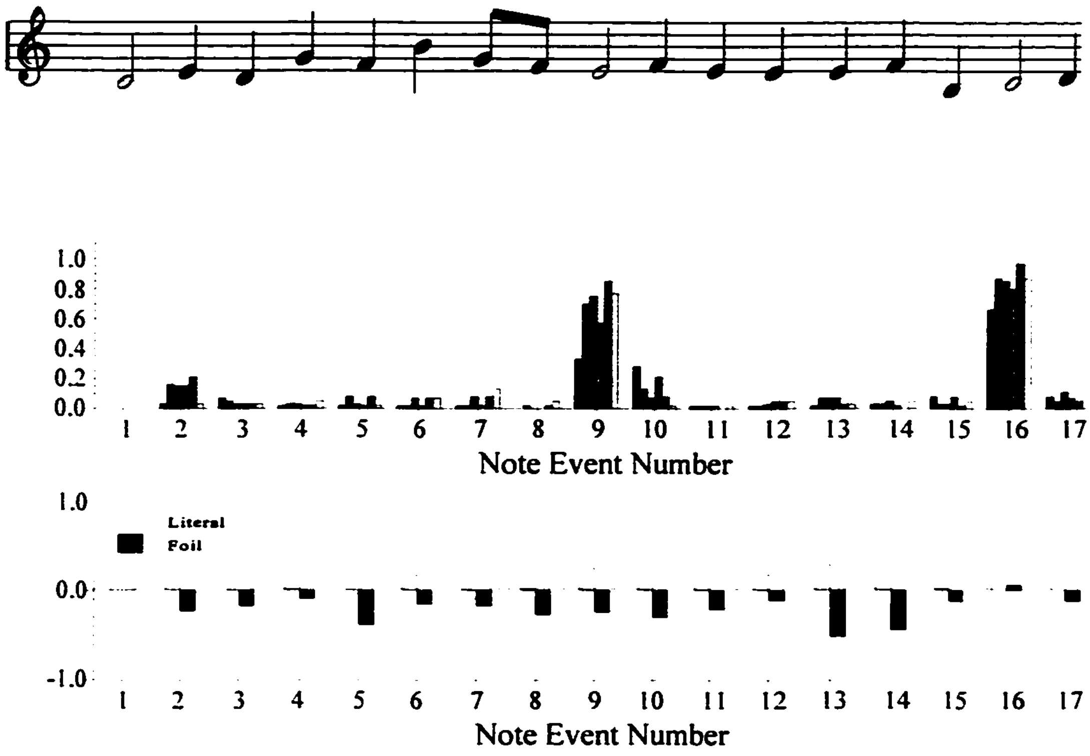


Figure 3.23a: The first part of the melody used in Stages 6 and 8, Part 1, and the location of subjective boundaries based on the average of all subjects (first, second, third, fourth, fifth and sixth repetitions of the song). The boundaries placed by subjects can be compared with performance on sequences extracted from positions near the boundaries. Sequence performance is denoted as a funcition of the point of the extraction of the first note. Hence, a sequence extracted at note event 7 is likely to have a boundary located on its second position.

■ **First Presentation**
■ **Second Presentation**
☐ **Third Presentation**
■ **Fourth Presentation**
■ **Fifth Presentation**
☐ **Sixth Presentation**

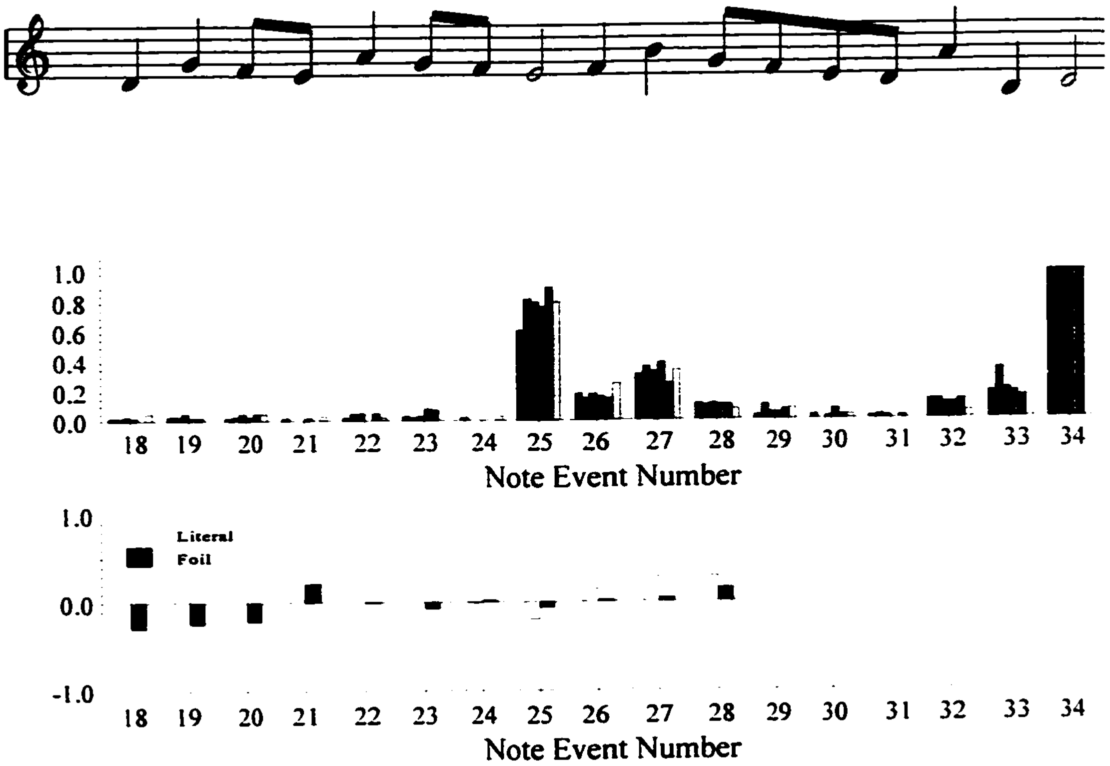


Figure 3.23b: The second part of the melody used in Stages 6 and 8, Part 1, and the location of subjective boundaries based on the average of all subjects (first, second, third, fourth, fifth and sixth repetitions of the song). The boundaries placed by subjects can be compared with performance on sequences extracted from positions near the boundaries. Sequence performance is denoted as a function of the point of the extraction of the first note. Hence, a sequence extracted at note event 22 is likely to have a boundary located on its last position.

▨ Rule 2a
▩ Rule 2b
■ Rule 3a
▧ Rule 3d
■ First Presentation
■ Second Presentation
☐ Third Presentation
■ Fourth Presentation
■ Fifth Presentation
☐ Sixth Presentation

Synopsis

At this point, it is prudent to pause and take stock of all the preceding analyses in order to simplify any and all subsequent analysis (i.e., analyses within groups; the relationship between tonality and boundary formation). Specifically, with respect to boundary location, this study would best be described as exploratory. Because it is exploratory, a large number of analysis were completed: Some of these can now be dropped.

Firstly, five different measures of binary similarity (Simple, Jaccard, Kulczynski 2, Sokal & Sneath 4, and phi) and their associated analyses were employed: There was no good, a priori, theoretical reason to choose one over the others. In fact, a subgoal of this analysis was to choose the best measure(s). However, results indicate that the use of all measures is redundant. That is, although the five measures are different, particular in the weight given to the co-occurrence of a non-boundary (see Table 3.8 and associated discussion), in this context, they all seemed to tap a common structure. All measures show similar results for comparisons across repetitions (consistency analysis) and for comparisons between subjects (similarity analyses). As a check on redundancy, the different measures were correlated against each other using both the Pearson correlation (linear relationships) and the Spearman rank correlation (this would capture more general monotonic, not necessarily linear, relationships) for Stage 4 (Table 3.85) and Stages 6 and 8 (Table 3.86). In this analysis, the consistency between Repetitions 1 and 2 using the simple similarity measure was correlated against the consistency between Repetitions 1 and 2 for each the other measures (Jaccard, Kulczynski 2, Sokal & Sneath 4 and phi) over the 61 subjects, and so on.

All resulting correlations were very high: In Stage 4, only one was below r=0.90 and even in Stages 6 and 8, although generally lower, none were below r=0.80. All were significant with a Type 1 Error Rate of α=0.01. The Spearman rank correlations were higher (only one was r_s=1.00), implying that there are real, if subtle, differences between the ranking by each measure (i.e., the differences are not a question of relative linearity).

Table 3.85: The Pearson correlations (r) and the Spearman rank correlations (r_s) between the various measures of profile consistency, within Stage 4. Part 1.

Stage 4		Jaccard		Kulczynski 2		Sokal & Sneath 4		Phi	
		r	r_s	r	r_s	r	r_s	r	r_s
Simple	1 to 2	0.956	0.967	0.967	0.968	0.980	0.980	0.980	0.980
	1 to 3	0.929	0.968	0.950	0.973	0.972	0.983	0.969	0.982
	2 to 3	0.896	0.985	0.922	0.987	0.958	0.988	0.969	0.988
Jaccard	1 to 2			0.986	0.997	0.986	0.995	0.986	0.997
	1 to 3			0.985	0.995	0.980	0.990	0.983	0.994
	2 to 3			0.984	0.997	0.975	0.993	0.978	0.998
Kulczynski 2	1 to 2					0.998	0.998	0.998	0.998
	1 to 3					0.997	0.997	0.997	0.997
	2 to 3					0.994	0.999	0.995	0.999
Sokal & Sneath 4	1 to 2							1.000	0.999
	1 to 3							1.000	0.998
	2 to 3							1.000	0.997

Notes: Correlations were computed between the same repetitions only (e.g., Phi of Repetition 2 was correlated against Jaccard of Repetition 2 only)
All values were significant, $p<0.01$

Table 3.86: The Pearson correlations (r) and the Spearman rank correlations (r_s) between the various measures of profile consistency, within Stages 6 and 8, Part 1.

Stage 6 & 8		Jaccard		Kulczynski 2		Sokal & Sneath 4		Phi	
		r	r_s	r	r_s	r	r_s	r	r_s
Simple	1 to 2	0.835	0.909	0.881	0.933	0.941	0.960	0.939	0.960
	1 to 3	0.808	0.858	0.812	0.887	0.900	0.943	0.908	0.939
	2 to 3	0.842	0.934	0.868	0.931	0.938	0.964	0.937	0.962
	4 to 5	0.876	0.959	0.930	0.965	0.969	0.984	0.966	0.982
	4 to 6	0.880	0.925	0.908	0.935	0.953	0.965	0.950	0.961
	5 to 6	0.918	0.970	0.936	0.972	0.970	0.984	0.969	0.983
Jaccard	1 to 2			0.960	0.960	0.941	0.951	0.959	0.972
	1 to 3			0.952	0.959	0.938	0.946	0.956	0.964
	2 to 3			0.990	0.998	0.970	0.992	0.971	0.993
	4 to 5			0.977	0.988	0.957	0.983	0.965	0.989
	4 to 6			0.975	0.990	0.967	0.977	0.972	0.987
	5 to 6			0.986	0.996	0.977	0.993	0.979	0.995
Kulczynski 2	1 to 2					0.988	0.993	0.987	0.990
	1 to 3					0.984	0.987	0.980	0.983
	2 to 3					0.986	0.991	0.987	0.992
	4 to 5					0.992	0.993	0.992	0.994
	4 to 6					0.992	0.992	0.992	0.993
	5 to 6					0.993	0.997	0.997	0.997
Sokal & Sneath 4	1 to 2							0.997	0.994
	1 to 3							0.997	0.996
	2 to 3							1.000	1.000
	4 to 5							0.999	0.998
	4 to 6							0.999	0.997
	5 to 6							1.000	0.999

Notes: Correlations were computed between the same repetitions only (e.g., Phi of Repetition 2 was correlated against Jaccard of Repetition 2 only)
All values were significant, $p<0.01$

Some readers might complain that the presentation of all measures was not necessary, but in fact, until this work had been completed in detail, one could not say for certain that the measures are redundant. In fact, despite the high correlations between measures, the measures are not identical: They have different central tendencies and, most importantly, different dispersions. The question is then, which measure is the best? Of all the measures, it is the Jaccard and the phi that show the greatest dispersion (relative to the range of 0 to 1; although the phi can actually take on negative values, such would not be theoretically meaningful in this context) for both consistency and similarity analyses. In the analysis of consistency changes in both Stage 4 and in Stages 6 and 8, the Jaccard measure shows the largest ω^2 (they are all about the same on η^2). In the analysis of the change in similarity measures (Stages 6 and 8) none of the measures shows any effects, though all are significant. The implication is that the Jaccard and phi measures have the greatest ability to delineate subjects. These also have the lowest chance and critical values (see Table 3.8) implying the largest headroom for subsequent studies. Between these two measures, the choice is much more ambiguous. However, as can be seen by the correlation matrices of Tables 3.85 and 3.86, only one should be necessary. The decision to retain the phi measure instead of the Jaccard measure was based on two reasons. Firstly, the Jaccard measure ignores the no-boundary match condition. I think it is premature to drop consideration of this condition (there are only two melodies in these analyses). Secondly, the phi measure is the binary analogue of the standard Pearson correlation, and as such, most people are familiar with its properties. Therefore, subsequent analysis of boundary location will only use the phi measure (this is also the reason only the phi cluster analyses were presented previously).

Finally, with respect to the analysis of boundary efficacy, it is clear that although there are differences in the results of the various analyses (i.e., within-subjects using pooled or specific error terms, hierarchical regression examining the change at each step, and the analyses of part correlations), the differences did not affect interpretation enough to warrant the retention of all. In addition, the individual analyses of Location, Size and Tonality do not add enough to the interpretation to support their retention (the Location and Size effects seem to be fairly independent, and Size and Tonality cannot be and Size effects seem to be fairly independent, and Size and Tonality cannot be disentangled anyway, particularly given the lack of balancing within Tonality). Hence, in subsequent analyses, only the regression approaches (hierarchical and part) to the overall analysis (all effects: [Session], Condition, Type, Location, Size, and Tonality) will be retained. This is motivated by the interest in the part analysis, and by the need for efficiency (statistical packages provide the regression approach: the ANOVA approach has to be reconstructed by hand from the regression approach provided by the statistical package).

It would also be convenient to limit analyses to the average profiles of subjects based on the last two repetitions within each Stage. Overall consistencies imply that this can be done. However, because this work is designed, in part, to explore this point (i.e., the consistency/learning of subjects between repetitions), this reduction cannot be made at this time. It is possible that different groups of subjects do demonstrate different consistencies.

Tonality Assessment

The field of music cognition grants considerable significance to the general concept of tonality, however defined or instantiated. As such, one would predict that tonality should have some impact on the formation of units within a melody. In particular, units are thought to begin and/or end on tonal centres (i.e., the tonic, the dominant, perhaps the subdominant or mediant). The higher order rules of Lerdahl and Jackendoff (1983) address, in part, the role of tonality in the structure of a melody. To explore this notion, the internal representation of tonality for all subjects was assessed using a modified probe-tone task (cf., Frankland & Cohen, 1990; Krumhansl, 1990). For each subject, this resulted in a tonality profile which quantified the sensitivity of the subject to tonality. For each subject, the boundary profiles (for each stage) and the tonality profile were then compared. Similarly, the boundary efficacy tasks (for each stage) and the tonality profile were also compared.

Although tonality and its relationship to boundary assessment was analysed in detail, there were, as this section will demonstrate, only minor effects of differential sensitivity to tonality within both the boundary location and the boundary efficacy tasks. If the reader can accept that general statement, then the reader may (temporarily) bypass this section, proceeding to the Discussion of this chapter (p. 412 of this work), without a large loss of continuity. Results pertaining to tonality have been retained at this point of the current work so that one may explore the details of design and analysis that lead to the previous conclusion. Issues concerning the relationship between the sensitivity to tonality and boundary assessment are best considered at this point because the analysis of the relationship between sensitivity to tonality and boundary assessment utilized the same techniques as the previous analysis of boundary assessment.

Information pertaining to the assessment of the internal representation of pitch (i.e., tonality) was collected in Stages 1, 3, 5, and 7. Since Stage 1 was practice, it was not analysed. In each of Stages 3, 5, and 7 subjects were presented with two blocks of trials, each block containing 26 trials, each trial consisting of a context and a probe tone (see Figures 3.7 and 3.8). Within a block, there were two types of contexts, ascending and descending major scales confined to the same octave, and there were 13 probe tones consisting of the 13 chromatic tones bounded by the octave defining the ascending and descending scales. Subjects assessed the relationship between the probe and the preceding context using a *good-*, *medium-* or *poor-fit* criteria. For analyses, good was coded as a 1, medium was coded as a 0.5 and poor was coded as a 0 (i.e., a range from 0 to 1). In addition to the valance of the response (hereafter called the response rating, or more simply, the rating), the reaction time was measured. However, the focus of this analysis is the response rating since speed was not emphasized during the task (the reaction-time data will not be presented; note that the results for the reaction time data mimicked those of Frankland & Cohen, 1990).

In each stage, response ratings over the two blocks were averaged to create probe-tone rating profiles (hereafter, more simply, the rating profiles) for the ascending and descending contexts. Rating profiles for the ascending and descending contexts were not averaged, but rather, placed end-to-end to create a single rating profile for the 26 probe tones. Hence, for each subject, there was one rating profile for each of Stages 3, 5, and 7.

The average rating profile for each subject was created by taking the mean of the rating profiles created within each of Stages 3, 5, and 7. To create this rating profile, the rating profiles obtained in Stages 3, 5, and 7 were adjusted to the common key of C. In actuality, only the profile of Stage 3 had to be adjusted, since the response tasks of both Stages 5 and 7 were presented in the key of C (the melodies of Stages 6 and 8 were the same, and both were presented in the key of C, and the probe-tone tasks were adjusted to match). One subject did not use the designated keys during Stage 3, so this subject's rating profile was based only on Stages 5 and 7.

The first analysis examined the similarity of rating profiles across repetitions on a subject-by-subject basis. One may view this as a check on the validity of the rating profile. If subjects did not produce the same rating profile on each repetition of the task, then the profile had little validity. This is analogous to the consistency analysis of boundary location. The average Pearson correlation between the different stages are presented in Table 3.87.

Table 3.87: Statistics for correlations between the Rating Profiles of Stages 3, 5, and 7, over all subjects (Consistency analysis for tonality).

		n	Stage 3 to 5	Stage 3 to 7	Stage 5 to 7
Overall	mean & sd	61	0.615±0.237	0.601±0.271	0.641±0.261
	min\max		-0.062\0.964	-0.359\0.964	-0.096\0.964
	median		0.686	0.673	0.712

Note: For Stage 3 to 5 & Stage 3 to 7, there were values from only 60 subjects.

In addition to high similarity for the pattern of the profile, the absolute mean rating (collapsed over the 26 probe tones) did not change as a function of repetition, as can be seen in Table 3.88. Generally, the results indicate that individuals perform consistently across repetitions of the task. Hence, the rating profiles can be safely considered representative of the performance of individual subjects at each point in the experiment.

Table 3.88: Statistics for the mean of the Ratings Profiles of individual subjects (collapsed over the 26 probe tones) of Stages 3, 5, and 7, over all subjects.

		n	Stage 3	Stage 5	Stage 7
Overall	mean & sd min\max median	61	0.469±0.127 0.029\0.708 0.468	0.468±0.124 0.038\0.740 0.471	0.496±0.126 0.154\0.769 0.490

Note: For Stages 3 and 5, there were values from only 60 subjects.

Given the high similarities between repetitions, the Experimentwise Ratings Profile was created for each subject by averaging the Ratings Profiles of Stages 3, 5 and 7 (the term is borrowed from its use in statistics for analysis of contrasts across experiments).

The correlation between subjects for the Experimentwise Ratings Profile is shown in Table 3.89. This is analogous to the similarity analysis for boundary location. Note that on average there is a high correlation between subjects. To complete this section, the average Experimentwise Rating Profile is shown in Figure 3.24. Note that this profile approximates what has been labeled as the triadic profile by a number of researchers (cf., Frankland & Cohen, 1990; Krumhansl, 1979, 1990[36];).

Table 3.89: Mean, minimum and maximum correlations between individual subjects Experimentwise Rating Profiles over all subjects (Similarity analysis for tonality).

		n	Between Subjects Correlations
Overall	mean & sd min\max median	61	0.472±0.226 -0.374\0.958 0.499

---

[36] Krumhansl typically collapses the ascending and descending profiles, but this can only be done if one is certain that they are the same. See later analyses within groups.

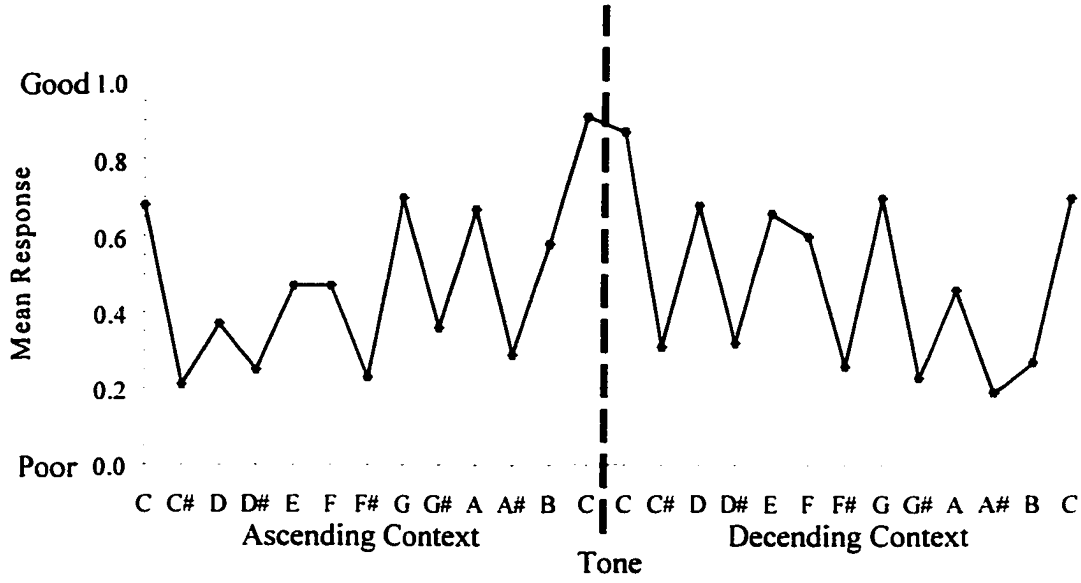


Figure 3.24: Mean response, adjusted to the key of C major collapsed over all subjects, collapsed over all repetitions. Scaling is designed to facilitate subsequent comparisons.

The main point of the analysis of tonality was creation of groups of subjects equivalent in their internal representation of tonality (as assessed by the probe-tone task). It was important that this homogeneity be high so that a lack of homogeneity in tonality (within each tonality group) would not be a confound in subsequent analyses of boundary location within tonality groups. However, for reasons of statistical power, a limited number of large groups was desired, which is often not easy to achieve in conjunction with high homogeneities.

A cluster analysis (cf., Frankland & Cohen, 1990; Krumhansl, 1979) was used to create groups who produced similar Experimentwise Ratings Profiles (with 26 probe-tones). The cluster analysis used the Pearson correlation coefficient as the measure of similarity. Four different clustering methods were explored -- between-groups linkage, within-groups linkage, single linkage and complete linkage -- because each method has some advantages. To balance the need for only a few groups with the desire for a high level of similarity within each group, the group similarity matrices for both the within-groups and the between-groups methods were examined at all the clustering stages between 1 and 25 groups. The criterion was set such that between groups cut line for the correlation was between r=0.50 and r=0.71 (r^2=0.25 & r^2=0.50), while no single correlation within a group fell below r=0.40 (r^2=0.16). In addition, the number of groups had to be minimized. In effect, this analysis was a manual form of the complete-linkage in combination with the stage-by-stage output of both the within-linkage and the between-linkage analyses (performed separately). In the final analysis (see Figure 3.25), it was the within-groups linkage with 12 groups that most closely approximated these criteria (a small number of groups; an average correlation above r=0.50 [r^2=0.25]; a minimum correlation above r=0.40 [r^2=0.16]). The result was three relatively large groups (n=23, n=18 and n=8) having average within-group correlations of r=0.71±0.13, r=0.71±0.14 and r=0.71±0.13, respectively. There were two small groups (n=2 and n=3) having average correlations of r=0.71±0.00 and r=0.71±0.13. The remaining seven "groups" were individual subjects.

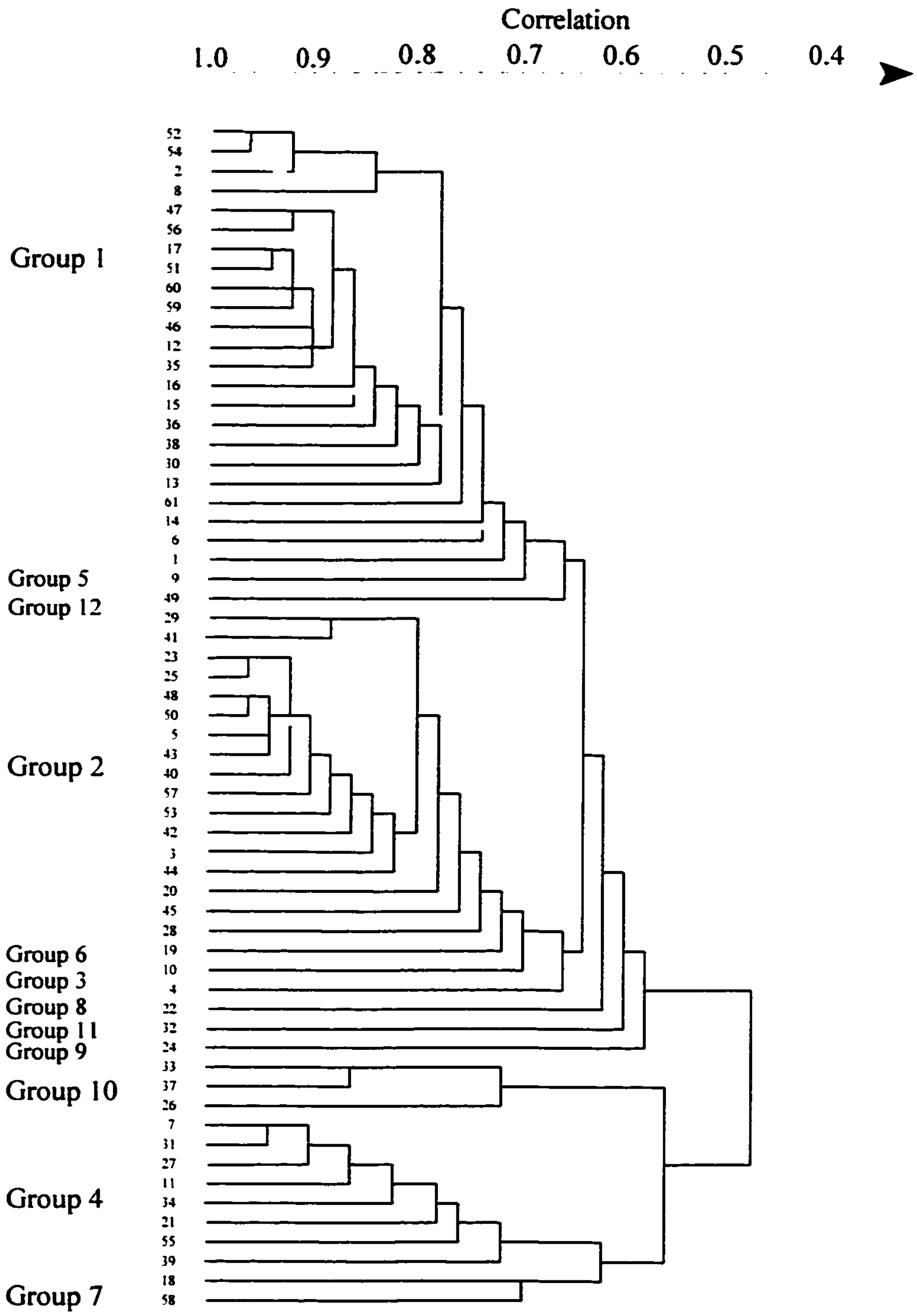


Figure 3.25 Clustering and group formation on the basis of responses in the probe-tone task (tonality). "Groups" in smaller fonts consist of single subjects. Compare with Figures 3.15 and 3.20.

Note that in comparison to the clustering for boundary locations (Figure 3.15 and 3.20), the analysis of tonality demonstrated a lower minimum, a lower maximum and generally lower consistencies (this is why it is reasonable to define subgroups for tonality but not for boundary location). The tonality (Figure 3.25) and boundary cluster analyses (Figure 3.15 and 3.20) are all presented using the average within-subjects clustering method, and the same scale to facilitate comparisons.

The mean correlations between the Experimentwise Rating Profiles for the different subjects within each group are presented in Table 3.90 (the similarity over all subjects is also provided for comparison). This is analogous to the similarity analysis for boundary location.

Having defined groups, the average profiles for each group were computed. These are presented in Figure 3.26 (the numerical labels from the groups are taken directly from the cluster analysis to facilitate comparison with Figure 3.25): The "groups" consisting of single subjects are also presented to highlight the reasons for their lack of inclusion in the other groups. From these profiles, one could label the two largest, Groups 1 and 2, as triadic: Group 1 exhibits some effects of recency in its profile (compare with Group 4) while Group 2 exhibits the more prototypical triadic profile. Groups 4, 7, and 10 exhibit what has been called a recency or proximity profile, although Group 4 is the only good example.

In addition to the Experimentwise Rating Profiles (Figure 3.26) and the between-subject correlations between Experimentwise Ratings-Profiles within each group (Table 3.90), the mean subject-by-subject correlations across repetitions (i.e., Stages 3, 5 & 7) were computed within each group. This is analogous to the consistency analysis of boundary location. These data are presented in Table 3.91: These can be compared with value collapsed over all subjects. Note that the main groups exhibit high degrees of similarity across repetitions (i.e., the main group is more consistent), while it is the groups that consist of an individual subject that do not. In fact, within Group 1, only two subjects demonstrated correlations below r=0.50 in the Stage 5 to Stage 7 comparison, while within Group 2, this number was 3 and within Group 4 this number was 2.

Table 3.90: Mean, minimum and maximum correlations between individual subjects Experimentwise Rating Profiles (i.e, the average of Stages 3, 5 & 7) within each group (Similarity analysis for tonality within each group).

		n	Between Subjects Correlations
Overall	mean & sd min\max median	61	0.472±0.226 -0.374\0.958 0.499
Group 1	mean & sd min\max median	23	0.712±0.216 0.396\0.958 0.721
Group 2	mean & sd min\max median	18	0.706±0.138 0.358\0.952 0.714
Group 3	mean	1	---
Group 4	mean & sd min\max median	8	0.706±0.125 0.338\0.924 0.734
Group 5	mean	1	---
Group 6	mean	1	---
Group 7	mean & sd min\max median	2	0.705±0.000 0.705\0.705 0.705
Group 8	mean	1	---
Group 9	mean	1	---
Group 10	mean & sd min\max median	3	0.711±0.126 0.611\0.852 0.670
Group 11	mean	1	---
Group 12	mean	1	---

Note: For groups having only 1 subject, there are no correlations to compute.

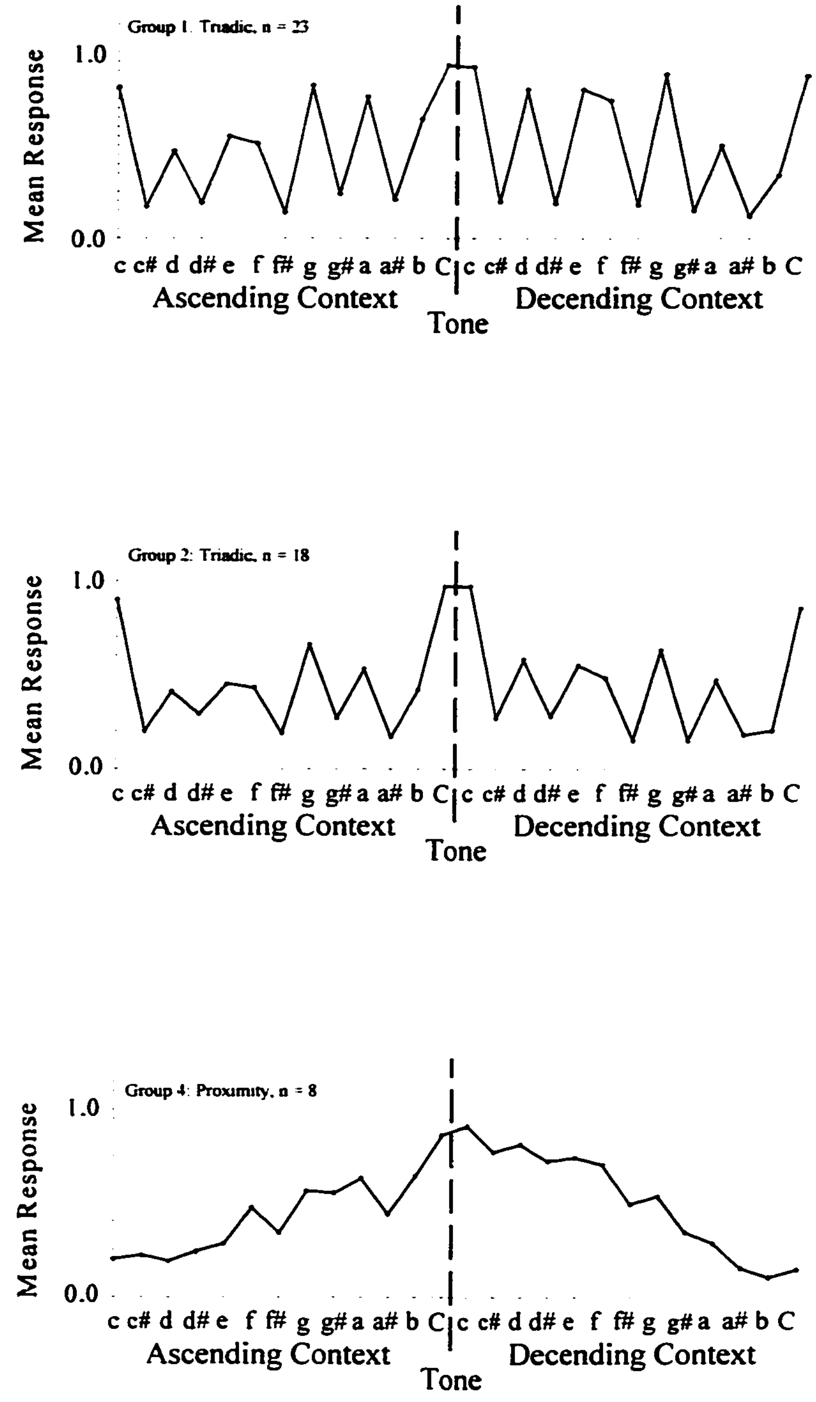


Figure 3.26a: Three main tonality groups determined by use of the probe-tone task. Groups 1 & 2 can be labelled as triadic, while Group 4 can be labelled as proximity or recency.

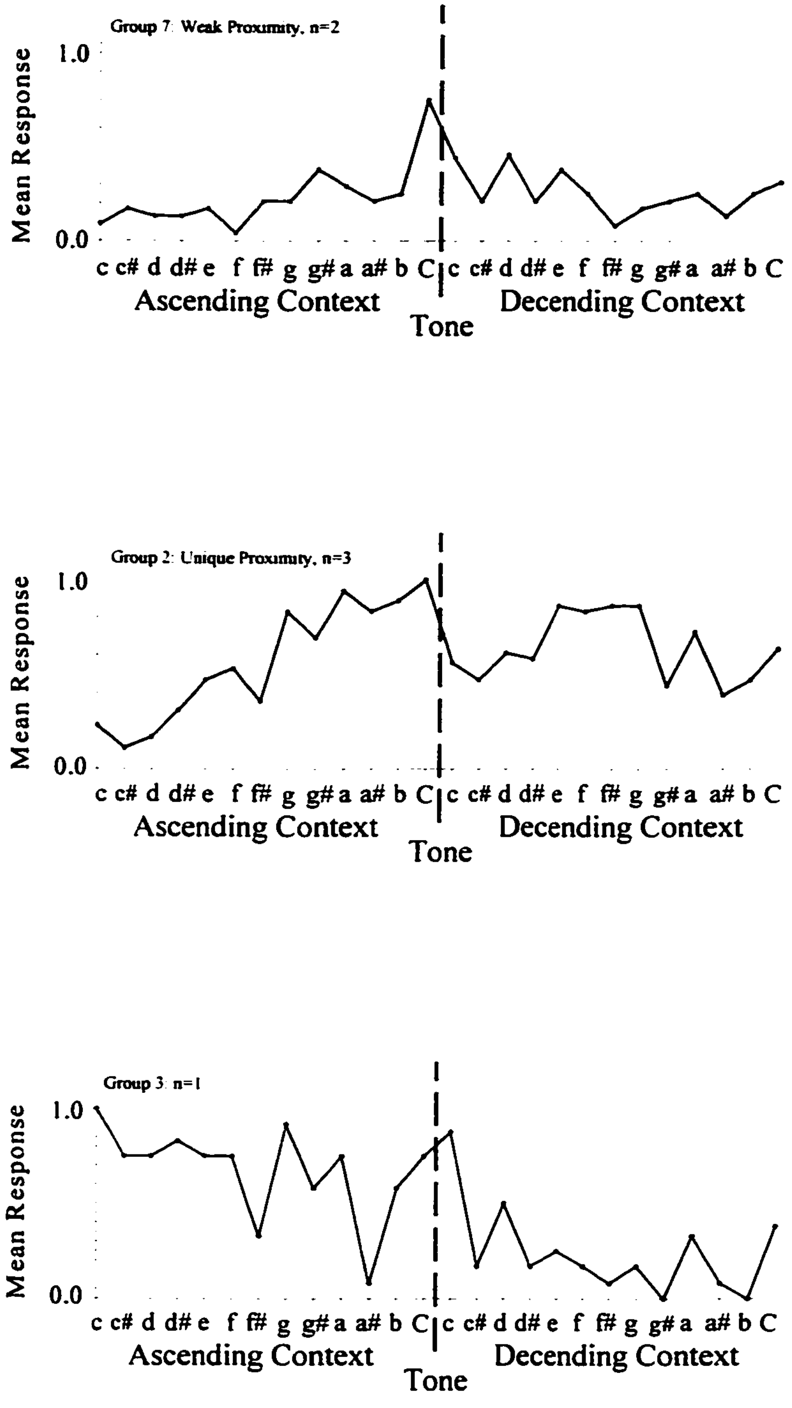


Figure 3.26b: Two secondary tonality groups determined by use of the probe-tone task. Groups 7 and 10 can be tentatively labelled as a variation of proximity. Group 3 consists of a single subject and exhibits a profile that seems to combine elements of the triadic and proximity profiles.

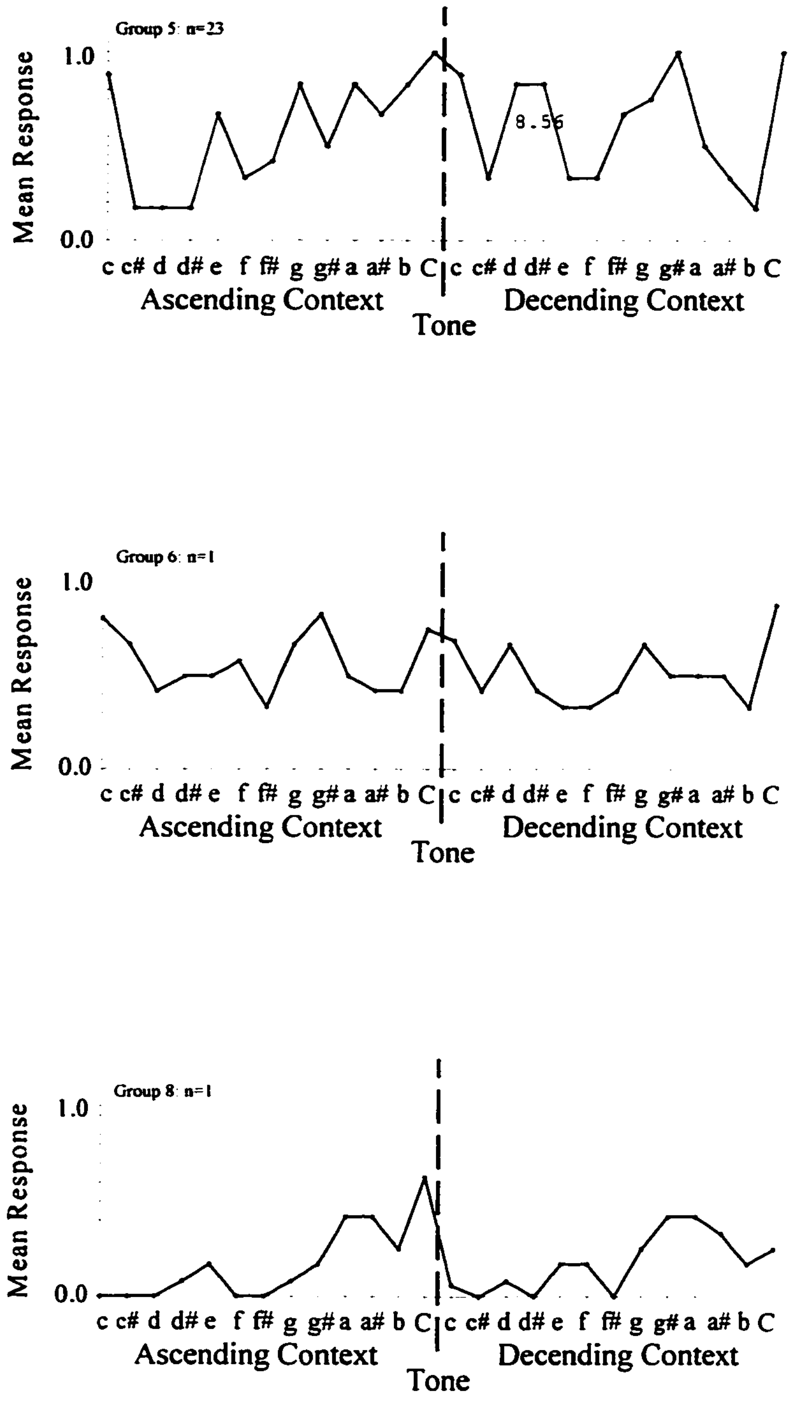


Figure 3.26c: Three tonality groups composed of individual subjects. Each of these groups is difficult to classify.

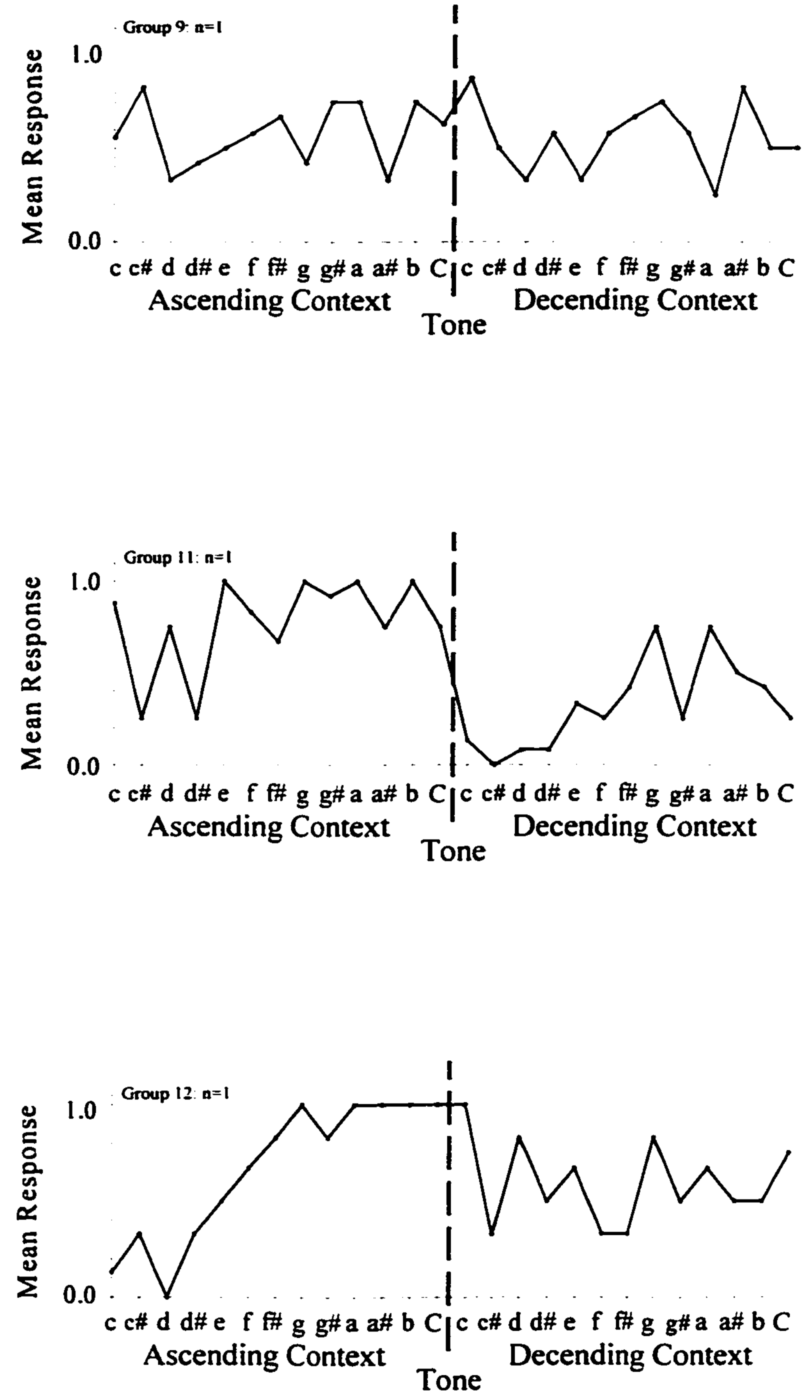


Figure 3.26d: Three tonality groups composed of individual subjects determined by use of the probe-tone task. Each of these groups is difficult to classify.

Table 3.91: Statistics for correlations between the Rating Profiles of Stages 3, 5, and 7, over all subjects and within each group Consistency analysis for tonality within each group).

		n	Stage 3 to 5	Stage 3 to 7	Stage 5 to 7
Overall	mean & sd min\max median	61	0.615±0.237 -0.062\0.964 0.686	0.601±0.271 -0.359\0.964 0.673	0.641±0.261 -0.096\0.964 0.712
Group 1	mean & sd min\max median	23	0.696±0.176 0.373\0.964 0.759	0.707±0.180 0.358\0.964 0.760	0.730±0.215 0.039\0.964 0.811
Group 2	mean & sd min\max median	18	0.683±0.209 0.133\0.905 0.773	0.674±0.231 0.081\0.885 0.716	0.708±0.223 0.192\0.957 0.765
Group 3	mean	1	0.650	0.576	0.686
Group 4	mean & sd min\max median	8	0.581±0.240 0.228\0.865 0.522	0.604±0.216 0.295\0.863 0.644	0.606±0.186 0.371\0.934 0.573
Group 5	mean	1	0.067	0.067	0.518
Group 6	mean	1	0.253	0.037	-0.096
Group 7	mean & sd min\max median	2	0.371±0.201 0.229\0.513 0.371	0.229±0.114 0.148\0.310 0.229	0.359±0.311 0.139\0.579 0.359
Group 8	mean	1	0.408	0.533	0.403
Group 9	mean	1	-0.062	-0.359	-0.010
Group 10	mean & sd min\max median	3	0.618±0.165 0.428\0.717 0.710	0.471±0.173 0.272\0.583 0.558	0.599±0.144 0.434\0.583 0.657
Group 11	mean	1	0.407	0.575	0.678
Group 12	mean	1	0.306	0.234	0.037

Note: For Stage 3 to 5 & Stage 3 to 7, there were 60 subjects overall and 22 subjects in Group 1.
For groups having only 1 subject, only the mean is reported since this is the median, minimum and maximum.

Table 3.92 provides the mean ratings (collapsed over the 26 probe tones) as a function of group and repetition (also presented is the data over all subjects). Note that all groups have mean ratings near the middle (i.e., 0.50) except Groups 7 and 8 which have very low mean ratings. The reason for these low ratings can be seen in the profiles of Figure 3.26. In Groups 7 and 8, only the upper tonic in the ascending context received high ratings.

Table 3.92: Statistics for the mean of the Ratings Profiles of individual subjects (collapsed over the 26 probe tones) over all subjects and within each group.

		n	Stage 3	Stage 5	Stage 7
Overall	mean & sd min\max median	61	0.469±0.127 0.029\0.708 0.468	0.468±0.124 0.038\0.740 0.471	0.496±0.126 0.154\0.769 0.490
Group 1	mean & sd min\max median	23	0.481±0.112 0.288\0.708 0.481	0.502±0.100 0.260\0.740 0.510	0.75230.106 0.375\0.712 0.471
Group 2	mean & sd min\max median	18	0.445±0.088 0.282\0.596 0.446	0.441±0.107 0.260\0.615 0.404	0.456±0.102 0.327\0.654 0.442
Group 3	mean	1	0.471	0.452	0.471
Group 4	mean & sd min\max median	8	0.485±0.110 0.317\0.622 0.458	0.434±0.090 0.269\0.558 0.433	0.489±0.152 0.260\0.769 0.510
Group 5	mean	1	0.680	0.558	0.558
Group 6	mean	1	0.417	0.587	0.635
Group 7	mean & sd min\max median	2	0.204±0.248 0.029\0.380 0.204	0.197±0.224 0.038\0.356 0.197	0.313±0.224 0.154\0.471 0.313
Group 8	mean	1	0.136	0.183	0.192
Group 9	mean	1	0.596	0.558	0.558
Group 10	mean & sd min\max median	3	0.611±0.051 0.553\0.646 0.635	0.577±0.098 0.471\0.663 0.596	0.612±0.083 0.538\0.702 0.596
Group 11	mean	1	0.574	0.481	0.596
Group 12	mean	1	0.577	0.673	0.635

Note: For Stages 3 and 5, there were 60 subjects overall and 22 subjects in Group 1. For groups having only 1 subject, only the mean is reported since this is the median, minimum and maximum.

All of these analysis were intended to insure that the different groups defined on the basis of tonality were internally consistent and distinct. That is, when exploring the relationship between tonality and boundary formation, it is important to be certain that the tonality was well defined. These analyses demonstrate that the tonality groups can be considered such. In each of these groups, the rating profiles of individual subjects demonstrated a large degree of stability across repetitions (i.e., different stages and hence, over the duration of the experiment) and a large degree of similarity between subjects. However, in addition to these groups, there were a few subjects whose performance is best described as unique. Note that of the three main groups, two are very similar (Groups 1 and 2), while the third is quite distinct (Group 4). This result mirrors other work in this area (Frankland & Cohen, 1990; Krumhansl, 1979). Given that there were distinct groups of subjects, the analysis of boundary location and boundary efficacy proceeded within each group. From this point on, analyses are only provided for Groups 1, 2 and 4 since only those groups had enough subjects for a meaningful analysis.

The Assessment of Music Study within Tonality Groups

Tonality is associated with involvement in music and training in music. Previously, several measures of subjects' involvement were discussed: Training, Practicing, Playing, Listening, as well as Age, Sex and Handedness. Those measures were also examined as a function of group. More detailed analysis were carried out within Groups 1, 2 and 4.

Table 3.93 presents the background statistics for all the groups. Basically, the groups had similar representations of Age, Sex and Handedness.

Table 3.93: Demographic statistics for each group.

Group	Sex		Handedness		Age				
	Male	Female	Left	Right	Mean±SD	Median	Mode	Min	Max
1	7	16	1	22	23.783±8.268	20	19	16	51
2	8	10	3	15	24.611±8.493	21	20	18	52
4[1]	3	5	1	6	19.375±1.188	19	19	18	22

Notes: [1] The one ambidextrous subject fell into this group.

The descriptive statistics concerning number of instruments studied and played, as well as, the hours per week spent listening to music (by choice and default) are presented in Table 3.94. Note that there were substantial differences between Groups 1, 2 and 4 on number of instruments studied and played. In particular the median number, the modal number and the maximum number per group are different.

Table 3.94: Number of instruments studied and played and hours/week listening to music by choice and by default. "Groups" consisting of a single subject are listed separately.

Group	Background Measure	Mean ± SD	Median	Mode	Min	Max
1	Trained	1.565 ± 1.037	2.000	2	0	3
	Played	1.826 ± 1.302	2.000	2	0	5
	Listen$_1$ (hrs/wk)	12.000 ± 9.826	10.000	3	1	42
	Listen$_2$ (hrs/wk)	4.043 ± 5.121	3.000	0	0	24
2	Trained	2.111 ± 1.811	2.000	2	0	7
	Played	2.389 ± 1.685	2.000	2	0	7
	Listen$_1$ (hrs/wk)	14.333 ± 9.604	15.000	20	2	30
	Listen$_2$ (hrs/wk)	9.278 ± 6.594	10.000	10	0.5	20
4	Trained	1.250 ± 1.488	1.000	0	0	4
	Played	1.750 ± 1.669	1.000	1	0	4
	Listen$_1$ (hrs/wk)	8.375 ±18.220	12.000	10	2	60
	Listen$_2$ (hrs/wk)	5.875 ± 4.581	4.500	4	0	15

Notes: Listen$_1$ refers to listening by choice; Listen$_2$ refers to listening by default.

The differences in the mean Age for Groups 1, 2 and 4 were not significant ($F(2,46)=1.337$, $p<.273$; see Table 3.95) but this was likely due to the low numbers of subjects per group (Group 4, which is the most distinct, had only 8 subjects). Chi-square analysis of Sex and Handedness revealed no significant GroupXSex ($\chi^2(2)=0.855$, $p<.652$) or GroupXHandedness ($\chi^2(2)=1.774$, $p<.412$) interactions. Listening by Default was actually significant ($F(2,46)=4.37$, $p<0.02$; see Table 3.95) and this difference is due to the difference between Groups 1 and 2. The remaining variables did not demonstrate any differences. However, on the basis of the contrasts, note that Age and Listening by

Choice differences were concentrated in the comparison of Groups 1 and 2 with 4 (i.e., all of the SS when into the second contrast), while Played and Listening by Default differences were concentrated in the comparison of Groups 1 and 2 (all of the SS went into the first contrast). The implication is that different factors lie behind the previous clustering.

Table 3.95: The analysis of differences in the demographic variables for Groups 1, 2 and 4.

Measure		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Age	Overall	159.32	2	2740.07	46	1.34	0.273	0.055	0.014
	Ψ_1	4.06	1			0.07	0.775	0.001	
	Ψ_2	155.25	1			1.27	0.113	0.054	
Studied	Overall	5.07	2	94.93	46	1.23	0.302	0.051	0.009
	Ψ_1	2.06	1			1.43	0.253	0.021	
	Ψ_2	2.31	1			1.12	0.296	0.030	
Played	Overall	3.90	2	105.08	46	0.85	0.433	0.036	0.000
	Ψ_1	3.04	1			1.33	0.254	0.028	
	Ψ_2	0.85	1			0.37	0.544	0.008	
Listening by Choice	Overall	241.61	2	6015.87	46	0.94	0.397	0.039	0.000
	Ψ_1	65.48	1			0.50	0.483	0.010	
	Ψ_2	181.14	1			1.39	0.245	0.029	
Listening by Default	Overall	278.11	2	1462.94	46	4.37	0.018	0.160	0.121
	Ψ_1	273.99	1			8.62	0.005	0.160	
	Ψ_2	4.12	1			0.13	0.721	0.000	

With respect to training, practicing and playing of instruments, the descriptive statistics presented in Tables 3.96, 3.97, and 3.98 do show substantial differences between groups. In particular, the two triadic groups (Groups 1 and 2) demonstrate the highest total hours of training both in the mean values and more impressively, the median values. (For equivalencies, divide the total hours by n*52 obtain the total years of training, practicing or playing at n hr/wk.). Note that any subject who was currently studying an instrument would have no associated playing time for that instrument. It is for this reason that Groups 1 and 2 have the counterintuitive median and mode playing times of 0.

Table 3.96: Total hours devoted to training, practicing and playing, and any RCM equivalent grade achieved, for each group.

Group	Total Hours & Grade	Mean ± SD	Median	Mode	Min	Max
1	Training	396.095 ± 456.870	225.770	0.000	0.000	1764.336
	Practicing	964.884 ±1921.449	208.404	0.000	0.000	7966.745
	Playing	189.294 ± 557.274	0.000	0.000	0.000	2595.584
	Grade	2.217 ± 3.741	0.000	0.000	0.000	12.000
2	Training	549.014 ± 646.041	424.705	0.000	0.000	2569.535
	Practicing	487.287 ± 621.798	190.051	0.000	0.000	1768.283
	Playing	318.466 ± 923.543	0.000	0.000	0.000	3754.442
	Grade	1.389 ± 2.933	0.000	0.000	0.000	10.000
4	Training	180.429 ± 282.018	85.254	0.000	0.000	830.064
	Practicing	162.520 ± 262.534	14.208	0.000	0.000	708.101
	Playing	15.097 ± 35.044	0.000	0.000	0.000	101.045
	Grade	0.375 ± 1.061	0.000	0.000	0.000	3.000

Subjects in Groups 1 and 2 did not appear to differ from those in Group 4 on intensity of training, practicing or playing (Table 3.97).

Table 3.97: Maximum intensity (hours/week) of training, practicing and playing for each group.

Group	Intensity (hrs/wk)	Mean ± SD	Median	Mode	Min	Max
1	Training	1.587 ± 1.320	2.000	2.000	0.000	5.000
	Practicing	2.696 ± 3.595	2.000	0.000	0.000	15.000
	Playing	0.805 ± 1.695	0.000	0.000	0.000	6.000
2	Training	2.444 ± 2.431	2.000	1.000	0.000	10.000
	Practicing	1.944 ± 1.947	1.250	0.000	0.000	6.000
	Playing	1.278 ± 2.567	0.000	0.000	0.000	10.000
4	Training	2.125 ± 2.167	2.000	0.000	0.000	5.000
	Practicing	1.875 ± 2.357	0.500	0.000	0.000	5.000
	Playing	0.813 ± 1.132	0.000	0.000	0.000	2.500

practisingn Groups 1 and 2 did not appear to differ from those in Group 4 on recency of training (and practicing) or playing, but it must be remembPractisingthis experiment was based in a university population.

Table 3.98: Recency of Training (and Practicing) and Playing (years in the past) for each groups.

Group	Recency (yrs)	Number	Mean ± SD	Median	Mode	Min	Max
1	Training	18	6.749 ± 9.047	3.625	1.667	0.000	39.000
	Playing	9	1.222 ± 3.667	0.000	0.000	0.000	11.000
2	Training	15	5.231 ± 5.751	3.667	0.000	0.000	18.667
	Playing	8	2.719 ± 5.933	0.000	0.000	0.000	16.750
4	Training	5	7.534 ± 4.463	9.000	0.667	0.667	11.667
	Playing	3	2.306 ± 2.008	3.250	0.000	0.000	3.667

Notes: Number is the number of individuals within a group to whom the concept of recency applied (i.e., one had to have played to have a recency value for playing).

Despite some apparent differences in total training, practicing and playing no signiPractisingults were noted in the corresponding ANOVAs (Table 3.99). With respect to the maximum intensity of training, practicing or playing, groups demonstrated little difference, which was echoed in the ANOVAs (see Table 3.99). The lack of significance is likely due to the fact that Groups 1 and 2 were generally similar and although Group 4 was different, it had only 8 subjects.

However, note again (as in the Table 3.95) that, in general, the analysis does support the notion that Group 4 was distinct from Groups 1 and 2 on Total Hours of Instruction, Total Hours of Playing, and Grade (i.e., most of the SS for the effect went into the second contrast). Conversely, Groups 1 and 2 were distinct on Maximum Intensity of Training, Practising, and Playing (all of the Maximum Intensities) as well as Recency of Playing (i.e., most of the SS went into the first contrast). The implication is that there are differences between these groups that can be explored in the future.

All the subjects with some formal grade (i.e., RCM equivalent) were to be found in Groups 1 and 2, with only two exceptions. In Group 1, sixteen had not been assessed, one each had achieved Grades 4 and 5, two had achieved Grade 6, and one each had achieved Grades 8, 10 and 12. Of the 23 subjects in the group, only 4 (17%) had not studied (trained or played) an instrument. In Group 2, fourteen had not been assessed (RCM equivalent), two had achieved Grade 4, and one each had achieved Grades 7 and 10. Of the 18 subjects in this group, only 2 (11%) had not studied (trained or played) an instrument. By contrast, in Group 4, only one subject had been assessed (RCM equivalent) reaching Grade 3. Of the 8 subjects, 2 (25%) had not studied (trained or played) an instrument. (The only other subject to have formally graded was in Group 6: Grade 9.)

Table 3.99: The analysis of differences in the training on, practising with, and playing of, instruments for Groups 1, 2 and 4.

Measure		SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Total Hours Training	Overall	771476	2	$1.224*10^7$	46	1.45	0.245	0.059	0.018
	Ψ_1	201644	1			0.76	0.389	0.015	
	Ψ_2	569832	1			2.14	0.150	0.044	
Total Hours Practising	Overall	4654661	2	$8.828*10^7$	46	1.21	0.307	0.050	0.009
	Ψ_1	2533872	1			1.32	0.256	0.027	
	Ψ_2	2120789	1			1.11	0.299	0.023	
Total Hours Playing	Overall	525384	2	$2.134*10^7$	46	0.57	0.572	0.024	0.000
	Ψ_1	144657	1			0.31	0.579	0.007	
	Ψ_2	380728	1			0.82	0.370	0.016	
Grade	Overall	21.57	2	462.07	46	1.07	0.350	0.045	0.003
	Ψ_1	7.95	1			0.79	0.378	0.016	
	Ψ_2	13.62	1			1.36	0.250	0.029	
Max Intensity of Training	Overall	7.60	2	171.65	46	1.02	0.369	0.042	0.001
	Ψ_1	7.52	1			2.02	0.162	0.042	
	Ψ_2	.08	1			0.02	0.884	0.000	
Max Intensity of Practising	Overall	7.31	2	387.69	46	0.43	0.651	0.019	0.000
	Ψ_1	5.99	1			0.71	0.404	0.016	
	Ψ_2	1.32	1			0.16	0.694	0.003	
Max Intensity of Playing	Overall	2.52	2	184.17	46	0.32	0.731	0.014	0.000
	Ψ_1	2.17	1			0.54	0.465	0.012	
	Ψ_2	.35	1			0.09	0.769	0.002	
Recency of Training (18,15,5)	Overall	28.30	2	1934.16	46	0.26	0.776	0.014	0.000
	Ψ_1	17.96	1			0.32	0.572	0.009	
	Ψ_2	10.34	1			0.19	0.668	0.005	
Recency of Playing (18,15,5)	Overall	9.85	2	362.05	46	0.23	0.796	0.026	0.000
	Ψ_1	9.57	1			0.45	0.512	0.026	
	Ψ_2	0.29	1			0.01	0.909	0.000	

Notes: Ψ_1 compares Group 1 with Group 2 (i.e., the two triadic groups)
Ψ_2 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group)

When instruments studied (trained or played) was considered as a binary condition (had studied one or more versus had studied none) there was no GroupXStudied interaction ($\chi^2(2)=0.818$, $p<.664$). When grading was considered as a binary condition (graded or not graded) there was no GroupXGrade interaction ($\chi^2(2)=1.112$, $p<.573$).

In general, although the groups were not significantly different on a large number of measures, the groups do seem to be delineated with respect to their involvement in music.

Boundary Assessment within Groups

The first question addresses the issue of boundary location (within each of Part 4 and Parts 6 & 8). Essentially all the previous analysis pertaining to consistency and similarity were conducted within the different groups. Statistical comparisons were conducted between Groups 1, 2 and 4. To summarize the following results succinctly, there were no dramatic differences between groups. More precisely, there were differences, but those differences were minor. Groups 1 and 2 were very similar, and Group 4 demonstrated some distinction with respect to those two groups. The discussion that follows should make this apparent.

The relationship between the boundary profiles and the Group Preference Rules of Lerdahl and Jackendoff (1983) were examined within each group. Statistical comparisons were also conducted between Groups 1, 2 and 4.

The second set of analyses examined the efficacy of boundary location on retention for the test sequences. Essentially all the previous analysis of boundary efficacy were repeated within groups (Groups 1, 2 and 4 only). Again, there were only minor differences. Retention was examined within Groups 1, 2 and 4 as a function of Session (Stages 6 & 8 only), Condition, Type, their interactions, Location, Size, Tonality. It should be remembered that this will bias the results against finding effects due to Tonality, since Tonality was the last variable entered, and since Tonality must, by definition, overlap with the variable Size.

Boundary Location Within Groups: Stage 4, Part 1

For all groups, the mean consistency values for subject-by-subject repetitions of the same melody within each of the groups, using the phi similarity measure, are presented in Table 3.100. Clearly all values indicated a high degree of consistency within individual subjects within groups. Interestingly, it was only in Groups 1 and 2 (both labeled as Triadic) that the modal values were 1.00. Hence, those subjects who showed the greatest degree of melody-to-melody consistency were confined to the triadic groups. In general, Group 4 (labeled as Proximity) had lower consistencies. However, in spite of any differences between groups, by both the standard deviations and the minimums, subjects rarely fell to the level of chance indicated in Table 3.8 (recall that these chance levels assumed equal numbers of boundaries per boundary profile and a particular probability for boundaries within a boundary profile).

Table 3.100: The average consistency ratings for individual subjects within each group for the different repetitions of the same melody within Stage 4, Part 1, given the Phi similarity measure of Table 3.8.

		Rep	Mean±SD	Median	Mode	Min	Max
Phi Similarity	Group 1 n=23	1-2	0.809±0.179	0.833	1.000	0.379	1.000
		1-3	0.806±0.174	0.863	1.000	0.415	1.000
		2-3	0.831±0.183	0.863	1.000	0.379	1.000
	Group 2 n=18	1-2	0.872±0.115	0.872	1.000	0.645	1.000
		1-3	0.819±0.143	0.862	0.863	0.475	1.000
		2-3	0.905±0.137	0.931	1.000	0.423	1.000
	Group 4 n=8	1-3	0.705±0.230	0.756	0.174	0.174	0.915
		1-3	0.725±0.236	0.771	0.788	0.174	0.928
		2-3	0.709±0.212	0.737	0.709	0.222	0.941

The analysis of the change in consistency ratings (Table 3.101) demonstrated that the groups had minor differences in their change in consistency ratings. The overall effect was marginal and the significant interaction was significant. When the interaction was analyzed as simple effects within each group, only Group 2 showed a change in consistency ($F(2,34)=3.54, p<.04$). The associated effect size ($\omega^2=12\%$, $\eta^2=17\%$) was not large, however, given that this variable can be considered as a within-subjects experimental variable with 61 subjects.

Table 3.101: The overall analysis of the change in consistency ratings (Phi measure) for the different repetitions of the same melody within Stage 4, Part 1, as a function of group.

Overall Analysis		SS	df	MS	F	p	η^2	ω^2
Between	Groups	0.24	2	0.12	2.94	0.063	0.113	0.026
Subjects	Ψ_3	0.05	1	0.05	1.10	0.300	0.023	
	Ψ_4	0.20	1	0.20	4.79	0.034	0.094	
	$\text{Error}_{B\text{-}S}$	1.89	46	0.04				
Within	Repetition	0.03	2	0.02	2.43	0.062	0.059	0.034
Subjects	Ψ_1	0.01	1	5.40	11128	0.264	0.027	
	Ψ_2	0.02	1	0.98	4.45	0.040	0.088	
	RepetitionXGroups	0.03	4	0.01	1.11	0.358	0.046	0.004
	$\text{Error}_{W\text{-}S}$	0.59	92	0.01				
	$\text{Error}_{\Psi 1}$	0.38	46	0.01				
	$\text{Error}_{\Psi 2}$	0.22	46	0.00				
Total		2.78	146					

Notes: Ψ_1 compares Consistency from 1 to 2 with Consistency from 2 to 3 (i.e., nearest neighbours in time)

Ψ_2 compares the average of Consistency from 1 to 2 and Consistency from 2 to 3 with the Consistency from 1 to 3 (close in time to far in time)

Ψ_3 compares Group 1 with Group 2 (i.e., the two triadic groups)

Ψ_4 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group)

All analyses were performed on log-transformed scores and no violations of the ws-anova were noted

When comparing across subjects (Table 3.102), average similarity ratings within all groups were lower (as expected), with minima that dip below the aforementioned critical values. However, these values were still above the chance levels specified in Table 3.8. Note that these within group values were generally higher than the previous overall analysis (e.g., mean *r*=0.69, min *r*=0.03; see Table 3.13) implying that the previous lower values were produced by comparisons between groups.

Table 3.102: The average similarity ratings using the Phi similarity measure for different subjects for the same melody, for each group, within Stage 4, Part 1, Repetition 3 only.

		Mean±SD	Median	Mode	Min	Max
Phi Similarity	Group 1	0.699±0.201	0.787	0.787	0.216	1.000
	Group 2	0.748±0.175	0.787	0.787	0.269	1.000
	Group 4	0.645±0.151	0.667	0.660	0.296	0.860

Statistically, the three groups differed on their mean similarity ratings (Table 3.103), but given the ω^2 for these analysis, the result is not too important.

Table 3.103: The analysis of differences in mean similarity rating (Groups 1, 2 and 4 only) using the Phi similarity measure within Stage 4, Part 1.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between Subjects	Groups	0.21	2	0.11	5.42	0.005	0.025	0.020
	Ψ_3	0.11	1	0.11	5.43	0.020	0.013	
	Ψ_4	0.11	1	0.11	5.41	0.020	0.013	
	$\text{Error}_{B\text{-}S}$	8.48	431	0.02				
Total		8.70	433					

Notes: The analysis was performed on log-transformed scores.
Ψ_3 compares Group 1 with Group 2 (i.e., the two triadic groups)
Ψ_4 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group)

To further assess the homogeneity of groups, Table 3.104 presents the average correlation between subjects in Groups 1, 2 and 4 (i.e., the average similarity of subjects in Group 1 with subjects in Group 2, etc.). In this analysis, one must remember that the groups were defined on the basis of tonality and then compared on the basis of boundary similarity. Note that subjects in Groups 1 and 2 were similar, but subjects in Groups 1 and 2 were less similar to subjects in Group 4. These values can be compared with those in Table 3.103. The average similarity between Groups 1 and 2 was less than the average similarity within Group 1, but slightly higher than the average similarity within Group 2. Interestingly, the average similarity between Groups 1 and 4 or 2 and 4 was higher than the average similarity within Group 4. As such, these results indicate that grouping by tonality is not capturing differences on boundary formation.

Table 3.104: The average similarity rating between subjects in different groups (Groups 1, 2 and 4 only), using the Phi measure.

	Mean±SD	Median	Mode	Min	Max
Group 1 with Group 2 n=414	0.719±0.199	0.787	0.787	0.143	1.000
Group 1 with Group 4 n=184	0.681±0.172	0.709	0.709	0.222	1.000
Group 2 with Group 4 n=144	0.680±0.196	0.709	0.709	0.249	0.936

Notes: These averages consist of subjects in different groups (within groups comparisons are not included).
The n's represent the number of pairs involved in each calculation, given that Groups 1, 2 and 3 had 23, 18 and 8 subjects, respectively.

Because the differences between groups are minor, boundary profiles for the groups are not provided. That is, in comparing the boundary profiles between the groups and the overall profile (for all subjects), I could not detect any difference visually. However, such an inspection does not alleviate the need for further analysis.

The mean, minimum and maximum distances between boundaries were analyzed within each group. However, there were only minor differences within groups in comparison to the overall analysis. That is, in the overall mixed ANOVA, there were effects of Repetition collapsed across groups, but no differences between groups and only one interaction. The effect of Repetition collapsed over groups is the same as the effect of repetition cited previously because these three groups account for most of the subjects (49 of 61). Differences between groups would be demonstrated by the effect of Group and more importantly by the interaction of Group and Repetition. To explore differences between groups, each analysis was also conducted as a simple effects analysis within each group.

To summarize briefly, there was a tendency for Groups 1 and 2 to be different from Group 4 on minimum distance between boundaries (Tables 3.105 & 3.106). This pattern did not change with repetition (Group 1 had a mean minimum distance between that of Groups 2 and 4). Effect sizes ranged from 0 to 12% which is indicative of a small but consistent effect. Simple effects demonstrated that there were no differences in any groups (Group 1: $F(2,44)=0.77$, $p<.469$; Group 2: $F(2,34)=0.22$, $p<.802$; Group 4: $F(2,14)=0.79$, $p<.478$). Effect sizes were equally unimpressive ($\omega^2=0.00$, 0.00 and 0.00, respectively).

Table 3.105: The important statistics for the minimum distance between boundaries, for each group, within Stage 4, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Repetition 1	Group 1	0.783±0.671	1.000	1.000	0.000	3.000
	Group 2	1.222±0.878	1.000	1.000	0.000	3.000
	Group 4	0.875±0.354	1.000	1.000	0.000	1.000
Repetition 2	Group 1	0.870±0.458	1.000	1.000	0.000	2.000
	Group 2	1.111±0.471	1.000	1.000	1.000	3.000
	Group 4	0.625±0.518	1.000	1.000	0.000	1.000
Repetition 3	Group 1	0.739±0.449	1.000	1.000	0.000	1.000
	Group 2	1.389±2.173	1.000	1.000	0.000	10.000
	Group 4	0.625±0.518	1.000	1.000	0.000	1.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.106: The overall analysis of the changes in the minimum distance between boundaries, across repetitions of the same melody, for Stage 4, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	7.58	2	3.79	3.01	0.059	0.116	0.025
Subjects	Ψ_3	0.26	1	.26	0.21	0.651	0.004	
	Ψ_4	7.32	1	7.32	5.80	0.020	0.112	
	$\text{Error}_{B\text{-}S}$	57.99	46	1.26				
Within	Repetition	0.05	2	0.03	0.04	0.961	0.001	0.000
Subjects	Ψ_1	0.00	1	0.00	0.00	1.000	0.000	
	Ψ_2	0.05	1	0.05	0.09	0.767	0.002	
	RepetitionXGroups	1.19	4	0.30	0.43	0.783	0.019	0.000
	$\text{Error}_{W\text{-}S}$	62.76	92	0.68				
	$\text{Error}_{\Psi 1}$	34.48	46	0.75				
	$\text{Error}_{\Psi 2}$	28.28	46	0.61				
Total		129.57	146					

Notes: Ψ_1 compares Boundaries in Repetition 1 with Boundaries in Repetition 3 (furthest apart in time)

Ψ_2 compares Boundaries in Repetitions 1 and 3 with Boundaries in Repetition 2 (together, the contrasts are a trend analysis)

Ψ_3 compares Group 1 with Group 2 (i.e., the two triadic groups)

Ψ_4 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group)

The Mauchly Sphericity Test was rejected, but the multivariate conclusions were the same.

For maximum distance between boundaries, there was an interaction between Groups and Repetition (Tables 3.107 & 3.108): The effect size reached 14% within Group 4. It seems that the patterns of responding in Group 4 changed while the pattern of responding in Groups 1 and 2 was fairly stable. Simple effects demonstrated that there were no differences in any groups (Group 1: $F(2,44)=1.46, p<.242$; Group 2: $F(2,34)=0.89, p<.419$; Group 4: $F(2,14)=2.38, p<.129$). Effect sizes for Groups 1 and 2 were low (ω^2=0.02, and 0.00) but Group 4 did demonstrate some effects (ω^2=0.14) implying that the lack of significance reflected the limited sample size. Contrast analysis indicated that, in Group 4, Repetition 2 had a significantly lower maximum.

Table 3.107: The important statistics for the maximum distance between boundaries, for each group, within Stage 4, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Repetition 1	Group 1	18.696±9.310	14.000	10.000	9.000	33.000
	Group 2	19.222±11.016	12.000	10.000	11.000	33.000
	Group 4	19.750±8.812	19.000	11.000	11.000	33.000
Repetition 2	Group 1	21.000±10.291	21.000	33.000	7.000	33.000
	Group 2	20.556±11.179	13.500	33.000	10.000	33.000
	Group 4	12.250±5.922	10.000	10.000	5.000	24.000
Repetition 3	Group 1	18.870±9.864	13.000	10.000	8.000	33.000
	Group 2	18.222±10.109	12.500	10.000	10.000	35.000
	Group 4	16.625±9.561	12.500	10.000	7.000	32.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.108: The overall analysis of the changes in the maximum distance between boundaries, across repetitions of the same melody, for Stage 4, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	210.67	2	105.34	0.44	0.646	0.019	0.000
Subjects	Ψ_3	144.46	1	144.46	0.60	0.441	0.013	
	Ψ_4	66.22	1	66.22	0.28	0.601	0.006	
	$\text{Error}_{B\text{-}S}$	10987.18	46	238.85				
Within	Repetition	33.65	2	16.82	0.56	0.571	0.012	0.000
Subjects	Ψ_1	15.52	1	15.52	0.477	0.010		
	Ψ_2	18.13	1	18.13	0.669	0.014		
	RepXGroups	318.51	4	79.63	2.67	0.037	0.102	0.064
	$\text{Error}_{W\text{-}S}$	2743.84	92	29.82				
	$\text{Error}_{\Psi 1}$	1498.09	46	25.57				
	$\text{Error}_{\Psi 2}$	1245.75	46	27.08				
Total		14293.85	146					

Notes: see Table 3.118
No violations of the mixed analysis were noted.

With respect to mean distance between boundaries (Tables 3.109 & 3.110), there were no main effects or interactions. Simple effects Simple effects demonstrated that there were no differences in any groups (Group 1: *F*(2,44)=2.65, p<.082; Group 2: *F*(2.34)=0.09, p<.916; Group 4: *F*(2,14)=1.74, p<.211). Effect sizes were generally small although Group 1 was slightly more impressive (ω^2=0.07, 0.00 and 0.00, respectively).

Table 3.109: The important statistics for the mean distance between boundaries, for each group, within Stage 4, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Repetition 1	Group 1	5.975±1.906	5.714	4.875	1.938	10.750
	Group 2	6.662±2.805	5.295	4.875	3.700	14.667
	Group 4	6.913±3.397	5.714	4.875	4.222	14.667
Repetition 2	Group 1	6.284±1.859	5.714	8.400	1.611	8.400
	Group 2	6.458±1.684	5.714	8.400	4.222	8.400
	Group 4	5.001±1.251	4.875	4.875	3.273	6.833
Repetition 3	Group 1	5.697±1.864	4.875	4.875	1.611	8.400
	Group 2	6.760±4.192	5.295	4.875	3.700	22.500
	Group 4	5.041±1.338	4.875	4.875	3.273	6.833

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.110: The overall analysis of the changes in the mean distance between boundaries, across repetitions of the same melody, for Stage 4, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between Subjects	Groups	20.08	2	10.04	0.85	0.433	0.036	0.000
	Ψ_3	0.00	1	0.00	0.00	0.996	0.000	
	Ψ_4	20.08	1	20.08	1.70	0.198	0.036	
	$\text{Error}_{B\text{-}S}$	542.40	46	11.79				
Within Subjects	Repetition	3.99	2	1.99	0.67	0.516	0.014	0.000
	Ψ_1	3.93	1	3.93	0.98	0.328	0.021	
	Ψ_2	0.06	1	0.06	0.03	0.864	0.001	
	RepetitionXGroups	19.92	4	4.98	1.67	0.164	0.068	0.026
	$\text{Error}_{W\text{-}S}$	274.93	92	2.99				
	$\text{Error}_{\Psi 1}$	185.25	46	4.03				
	$\text{Error}_{\Psi 2}$	89.68	46	1.95				
Total		861.32	146					

Notes: see Table 3.119

The Mauchly Sphericity Test was rejected, but the multivariate conclusions were the same.

Each of the groups was analyzed with respect to the rules of Lerdahl and Jackendoff (1983). As in the overall analysis, the first analysis examined the correlation between each of the rules and the empirical boundary profile (within each group). In this case, only the optimal shift is presented (it is universally the same as in the overall analysis) and for each subject the average of Repetitions 2 and 3 was used to increase power (i.e., the aforementioned individual boundary profiles). These correlations are presented in Table 3.111. Note that although Groups 1 and 2 had the highest correlations, these groups were not dramatically different from Group 4 in their use of the rules.

Table 3.111: The correlations between the average profile and the predicted boundaries, given the rules of Lerdahl and Jackendoff, within each group in Stage 4, Part 1.

	Stage 4		
	Rule 2b	Rule 3a	Grammar
Group 1	0.771**	0.336*	0.568**
Group 2	0.784**	0.348*	0.627**
Group 4	0.716**	0.320*	0.526*

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.

Only the optimal shift is shown, which was the same in all cases as the overall shift.

* Significant (one tailed) at $p<0.05$

** Significant (one tailed) at $p<0.01$

As was done in the overall analysis, the rules were combined to find the best fitting regression equation. The results, presented in Table 3.112 demonstrated that all rules were most effective in Groups 1 and 2 and that generally, only Rule 2b was necessary. That is, Rule 3a did not add to the predictability of the equation (statistically, Rule 3a helped in Group 1, but the effect was minimal). Note that all equations were fairly similar.

Figure 3.112: The regression equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) in the average boundary profile for the second repetitions, within Groups 1, 2 and 4, Part 1.

Regression Equation: Stage 4, Group 1					
Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
B = 0.04±0.03 + (0.88±0.12)*Rule2b	0.594	0.000	0.594	67.379	0.000
- (0.02±0.09)*Rule3a	0.001	0.780	0.595	33.054	0.000

Regression Equation: Stage 4, Group 2					
Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
B = 0.02±0.03 + (0.93±0.12)*Rule2b	0.594	0.000	0.615	73.519	0.000
- (0.02±0.09)*Rule3a			0.616	36.019	0.000

Regression Equation: Stage 4, Group 4					
Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
B = 0.06±0.04 + (0.84±0.14)*Rule2b	0.594	0.000	0.512	48.336	0.000
- (0.01±0.10)*Rule3a			0.513	23.665	0.000

Notes: Rules 3b, 3c, 2a and 3d concerned attributes that did not change within this particular melody.

The equation represents the maximal equation when all rules are used to predict boundary location. The rule that is above the dotted line is the only rule that was included when a forward stepwise regression analysis was used. The rule below the dotted line was not included when the stepwise approach was used. The change in R^2 is the incremental improvement in the fit when using the stepwise solution.

A between-subjects ANOVA (Table 3.113) did not demonstrate significant differences between Groups 1, 2 and 4, although Rule 2b was marginal ($F(2,36)=3.09, p<0.06$).

Table 3.113: The analyses of the difference between the use of rules (the correlations between the boundary profiles of subjects and the predicted boundaries given the rule of Lerdahl and Jackendoff) within Groups 1, 2 and 4, Stage 4, Part 1.

Rule 2b Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	0.020	2	0.010	3.09	0.055	0.118	0.078
Subjects	Ψ_3	0.009	1	0.009	2.84	0.099	0.053	
	Ψ_4	0.011	1	0.011	3.33	0.075	0.065	
	$\text{Error}_{B\text{-}S}$	0.149	46	0.003				
Total		0.169	48					

Rule 3a Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	0.003	2	0.001	0.95	0.396	0.040	0.000
Subjects	Ψ_3	0.002	1	0.001	1.43	0.238	0.026	
	Ψ_4	0.001	1	0.001	0.46	0.500	0.014	
	$\text{Error}_{B\text{-}S}$	0.071	46	0.002				
Total		0.070	48					

Grammar Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	0.008	2	0.004	1.29	0.285	0.053	0.012
Subjects	Ψ_3	0.007	1	0.007	2.33	0.133	0.046	
	Ψ_4	0.001	1	0.001	0.24	0.623	0.007	
	$\text{Error}_{B\text{-}S}$	0.143	46	0.003				
Total		0.151	48					

Notes: The analyses were performed on log-transformed scores.
Ψ_3 compares Group 1 with Group 2 (i.e., the two triadic groups)
Ψ_4 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group).

The final analysis examined the correlation between the use of the rules (Table 3.114) within groups (i.e., the utility of rules analysis; see Table 3.25). Note that all groups demonstrated reasonably high correlations between the use of the rules. That is, in all groups, the individuals who used Rule 2b also tended to used Rule 3a.

Table 3.114: The correlation across subjects (within each group) between the efficacy of the different rules for predicting boundaries in the individual boundary profile of each subject.

		Correlations Between Measures		
		Rule 2b	Rule 3a	Grammar
Group 1 n=23	Rule 2b	1.000	0.606**	0.872**
	Rule 3a		1.000	0.606**
Group 2 n=18	Rule 2b	1.000	0.561*	0.605**
	Rule 3a		1.000	0.276
Group 4 n=8	Rule 2b	1.000	0.466	0.797**
	Rule 3a		1.000	0.366

Notes: * $p<0.05$
** $p<0.01$

Generally, although there were some group differences, boundary profile analysis was not strongly associated with tonality groups. That is, different tonality groups did not tend to produce different boundary profiles. Figure 3.27 demonstrates graphically the reason for the lack of a strong association between tonality grouping and boundary profile similarity by presenting the cluster analysis for boundaries (hence, demonstrating which subjects are most highly associated) along with the tonality group identifications. Note that subjects who were in the same tonality group were not necessarily neighbours in the boundary profile cluster analysis. Conversely, note that those tonality groups that were composed of a single subject are interspersed throughout the boundary profile cluster analysis.

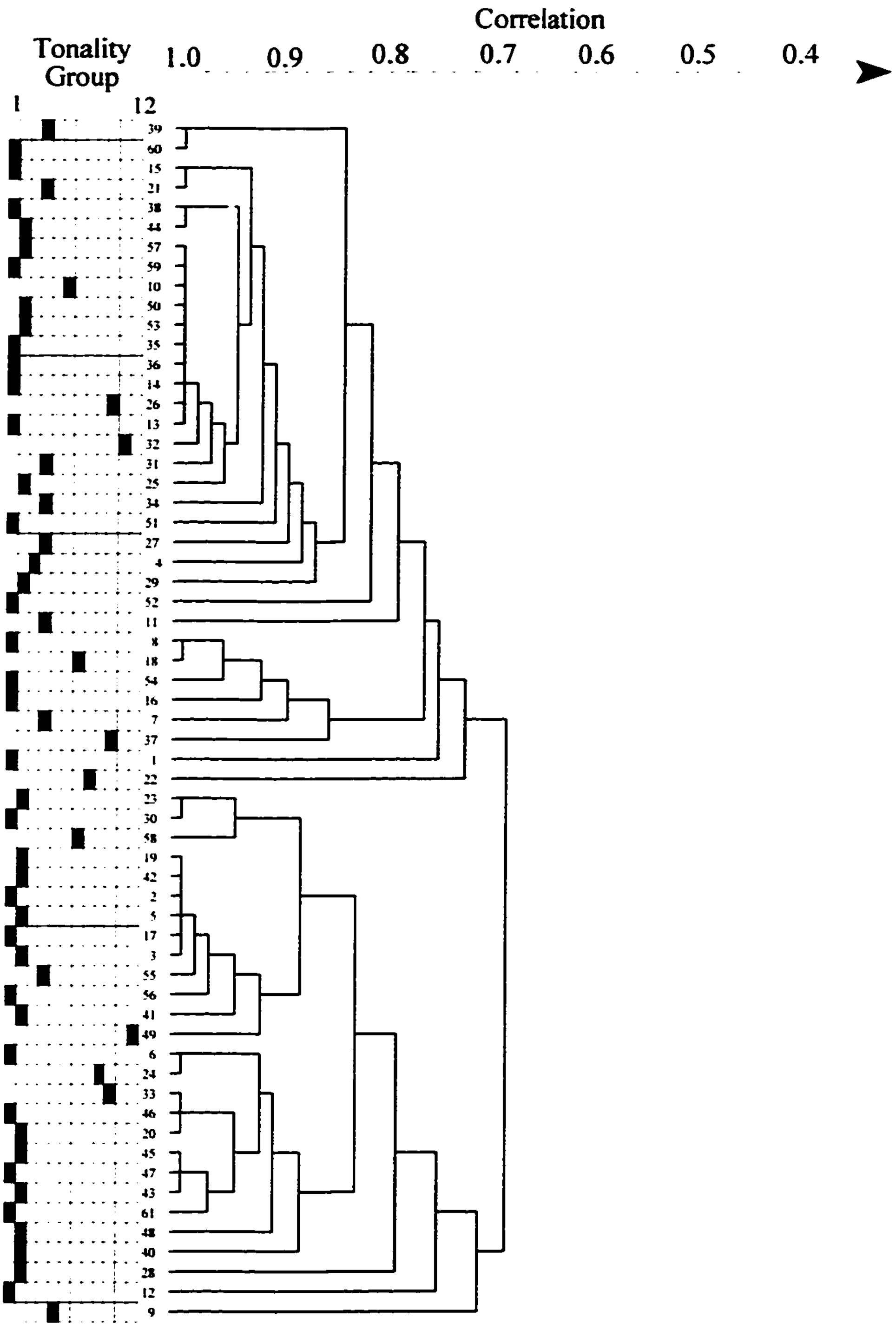


Figure 3.27: Comparison of cluster analysis for Boundary Location (phi measure, Stage 4) and group identification on the basis of Tonality Assessment (Tonality Groups 1 to 12). Note that members of the same tonality group are not necessarily adjacent in the boundary location.

Boundary Efficacy Within Groups: Stage 4, Part 2

Each subject was presented with 16 test sequences that varied on a number of dimensions: primarily seven Conditions (-1, 0, 1, 2, 3, 4 and 5) and two Types (Literal and Foil), but Type: Foil was further divided into Size, Location and Tonality. For Foil sequences, Location varied over 1, 2, 3 and 4 while Size varied over 2, 1, -1 and -2. Both Size and Location equal to 0 represented the Literal type. Tonality actually assessed the degree of change in tonality when the Foil sequence was created from its parent Literal (see Table 3.9). In essence, any changes that resulted in a large change in tonality should be more detectable than any changes that do not result in a large change in tonality (remember that all changes are relatively small in terms of absolute pitch). However, one would expect that this prediction would only be valid for those subjects demonstrating a triadic tonality profile: that is, Tonality Groups 1 and 2. Groups not demonstrating a triadic tonality profile would be expected to be insensitive to such distinctions: That is Tonality Group 4 Conversely, one might expect groups demonstrating a triadic tonality profile to be less sensitive to the magnitude of size changes (since they are focussed on the tonality aspect of note frequency) while groups not demonstrating a triadic tonality profile would be more sensitive to absolute size changes (since it is the only dimension that frequency is processed along).

Initially, Condition, Type, ConditionXType, Location, Size, Tonality, LocationXCondition, SizeXCondition and TonalityXCondition were analyzed within each group separately, using hierarchical multiple regression examining the incremental change in R^2 as the different terms were added to the equation and testing the part correlations (i.e., the semipartials) for each effect (analogous to the balanced design; cf., Cohen & Cohen, 1994, Howell, 1995, Keppel & Zeddick, 1992).

The previous contrasts for each effect were used (see Tables 3.29 for Condition and Type, 3.33 for Location, 3.36 for Size and 3.39 for Tonality). In addition to these effects, there were a changing number of subject vectors (using dummy coding) and four covariate vectors (first and second order for the serial position within the melody -- with 0 defined as the middle of the melody; first and second order for the serial position of the test sequence within the test phase -- with 0 defined as the middle of the set of test sequences). Using a hierarchical analysis, subject effects were entered first, followed by the covariates, the main effects of Condition and Type, their interaction, Location, Size, Tonality and finally their interactions with Condition. Tonality was entered last because it was not controlled to the same extent as Location or Size. Size was entered just before Tonality because Size and Tonality represented a competition of sorts (Size and Tonality are both aspects of pitch height, while Location is separate from the two).

Four the three groups, the means for Condition and Type (and their interaction) are presented in Table 3.115 through 117.

Table 3.115: The means, standard deviations and cell counts for Condition, Type and ConditionXType for Group 1 in Stage 4, Part 2.

				Group 1				
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.83±.39	.93±.27	.65±.49	.72±.46	.75±.44	.87±.35	.71±.47	.80±.40
	23	40	23	18	20	30	14	168
Foil	.70±.48	.71±.46	.80±.41	.40±.51	.68±.48	.90±.30	.83±.39	.73±.44
	10	21	15	15	22	21	23	127
Cond	.79±.42	.85±.36	.71±.46	.58±.50	.71±.46	.88±.33	.78±.42	.77±.41
	33	61	38	33	42	51	37	295

Table 3.116: The means, standard deviations and cell counts for Condition, Type and ConditionXType for Group 2 in Stage 4, Part 2.

				Group 2				
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.78±.43	.92±.28	.72±.46	.59±.51	.31±.48	.75±.44	.80±.45	.73±.44
	18	36	18	17	13	24	5	131
Foil	.62±.51	.81±.40	.92±.29	.56±.53	.75±.44	1.0±.00	.67±.48	.76±.43
	13	16	12	9	20	15	21	106
Cond	.71±.46	.88±.32	.80±.41	.58±.50	.58±.50	.85±.37	.69±.47	.75±.44
	31	52	30	26	33	39	26	237

Table 3.117: The means, standard deviations and cell counts for Condition, Type and ConditionXType for Group 4 in Stage 4, Part 2.

	Group 4							
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.75±.50	1.0±.00	.88±.35	.80±.45	.70±.48	.67±.50	.71±.49	.80±.40
	4	13	8	5	10	9	7	56
Foil	1.0±.00	.50±.52	.63±.52	.67±.58	.63±.52	.88±.35	.89±.33	.69±.47
	1	12	8	3	8	8	9	49
Cond	.80±.45	.76±.43	.75±.45	.75±.46	.67±.49	.76±.44	.81±.40	.75±.43
	5	25	16	8	18	17	16	105

In general, the pattern of responding for Condition and Type within Groups 1 and 2 follows that of the overall analysis (though Group 2 shows much more range in responding). In Group 4, the only discernable pattern is that there was a difference for Type: For Literals, proportions decrease as Condition number increases while for Foils proportions increase as Condition number increases. To show the similarity of Groups 1 and 2 to the overall performance, the correlation between the seven conditions (divided into Literals and Foils) across groups was computed. As can be seen in Table 3.118, Groups 1, 2 and Overall have high correlations (Literals with Literals; Foils with Foils:), while Group 4 is distinctly different. It is not surprising that Groups 1 and 2 are most related to the Overall ratings (since they comprise the bulk of the subjects). What is interesting is that Group 4 is not the same as Groups 1 and 2 (r's=0.21, 0.33, 0.47 & -0.18) and that, in general, Groups 1 and 2 are the same (r's=0.42 & 0.73). Also note that only Group 4 shows a substantial negative correlation between responses on Literals and Foils. Finally, it should be noted that the lack of "significance" is due to the fact that there were only 7 values in each computation.

Table 3.118: Correlation over the pattern of responding on Condition.

		Group 1		Group 2		Group 4		Overall	
		Literal	Foil	Literal	Foil	Literal	Foil	Literal	Foil
1	Literal	1.000	.167	.428	.208	.210	.083	.798*	.062
	Foil		1.000	.361	.731	-.213	.326	.255	.958**
2	Literal			1.000	.169	.467	.180	.837*	.271
	Foil				1.000	.070	-.178	.249	.747
4	Literal					1.000	-708	.589	-.387
	Foil						1.000	-.050	.438
O	Literal							1.000	.109
	Foil								1.000

Notes: * sig at p<.05
* sig at p<.01

Performance as a function of Location is shown in Table 1.119. Again Groups 1 and 2 showed the same pattern as the overall responses (though here it is Group 1 that showed more variation), with changes in the first location being very detrimental to performance: It is likely that the lack of a proper first note makes it difficult to establish the location of the test sequence within the melody. That is, the subject, on the basis of the false first note, probably (possibly without conscious effort) tried to match the test sequence to an inappropriate location in the melody. Changes on note 3 were also detrimental although it is not easy to construct a rationale for this phenomenon. In Group 4, the pattern was different (though the third cell has only a count of 1). Note that the previous correlational analysis (Table 3.118) was not computed on Location due to range of cell counts (Location, Size and Tonality were not explicitly controlled): All means (and there are only 4 of them) could not be assumed to be equally representative of the true population mean.

Table 3.119: The means, standard deviations and cell counts for Location for Groups 1, 2 and 4 in Stage 4, Part 2.

	Location of Change				
	Literal	Note 1	Note 2	Note 3	Note 4
Group 1	.80±.39	.56±.50	.94±.24	.47±.51	.81±.39
	168	32	33	19	43
Group 2	.73±.44	.52±.51	.93±.27	.54±.52	.88±.33
	131	25	27	13	41
Group 4	.80±.40	.59±.50	.83±.41	1.0±.00	.75±.44
	46	22	6	1	20

For Size (Table 3.120), Group 1 showed slightly better responding for larger changes, while Group 2 did not have much differentiation at all. Group 4 appeared to be the opposite of Group 1 (though the first cell has only a count of 1).

Table 3.120: The means, standard deviations and cell counts for Size for Groups 1, 2 and 4 in Stage 4, Part 2.

	Size of Change				
	Literal	Size -2	Size -1	Size 1	Size 2
Group 1	.80±.39	.88±.33	.71±.46	.68±.47	.76±.44
	168	17	41	44	25
Group 2	.73±.44	.85±.38	.78±.42	.76±.44	.72±.46
	131	13	32	29	32
Group 4	.80±.40	.00±.00	.67±.48	.81±.40	.56±.53
	46	1	18	21	9

For Tonality (Table 3.121), Groups 1 and 2 showed a general decrease in performance across conditions, while Group 4 had a much flatter response (though the cell counts make interpretation difficult). Initially, it was predicted the Tonality 1 should be the easiest, with a general decrease associated with condition number (see Table 3.9), and that this should hold for the triadic groups, but not for the proximity groups. The predictions were largely borne out by the pattern, with the exception that in Group 1, Tonality 1 did not produce the highest performance.

Table 3.121: The means, standard deviations and cell counts for Tonality for Groups 1, 2 and 4 in Stage 4, Part 2.

	Tonality of Change					
	Literal	Ton 1	Ton 2	Ton 3	Ton 5	Ton 7
Group 1	.80±.39	.79±.41	.93±.27	.69±.47	.33±.58	.33±.52
	168	53	14	51	3	6
Group 2	.73±.44	.86±.35	.70±.48	.79±.41	.00±.00	.50±.58
	131	36	10	52	4	4
Group 4	.80±.40	.68±.48	1.0±.00	.58±.51	—	1.0±.00
	56	31	4	12	0	2

Initially, this was puzzling, but with the usual 20/20 hindsight the answer is obvious. If one has a short sequence of four notes, and then in that sequence, one changes one of the triad tones to a non-diatonic tone, one has dramatically changed the sequence. The change will be dramatic since a short sequence is not likely to contain a multiple of triad tones, and even if it does, changing one of four triad tones makes a sequence sound very different. In this situation, it is possible that subjects simply fail to recognize the sequence at all, so performance drops. In contrast, when a non-triad tone (either diatonic, or non-diatonic) is changed, the character of the sequence is not so dramatically altered, so that the parent from which the sequence was derived can still be recognized within the melody. Of course, this would only matter to a group of listeners who actively use triad tones as reference points, in either the melody or the sequence. A more detailed analysis of the combinations of Tonality and Location might reveal a more interpretable pattern, but this will have to wait until there are studies using the techniques detailed herein that control Tonality and Location more precisely (i.e., in the same manner as Condition and Type were herein).

The ANOVA tables for each group are presented in 3.122, 3.123 and 3.124 (Tonality Group 1, 2 and 4 respectively). In this section, the interesting analyses are the comparisons between groups, particularly the interactions of groups and other variables. However, when Group was included in an overall analysis, Group was not important (ΔR^2=0.000); note that the Group variable simply implies that the average performance does not change as a function of group. The interactions of Group with the other variables were the last vectors to be entered into that overall analysis (GroupXLocation, GroupXSize and GroupXTonality; these must be entered after their corresponding main effects), and as such they showed nothing.

Since the interesting questions concerned a comparison of what happened within each group, the omnibus analysis within each group is presented (equivalent to a simple effects analysis within each group). Given the greater number of subjects, it is not surprising that Tonality Group 1 shows more "significant effects", but one should note the size of the corresponding correlations to get an idea of the importance of the various effects.

In Tonality Group 1 (Table 122), Condition, Location, Size and ConditionXLocation were all significant when assessed as the incremental change in R^2, but only Condition remained significant in the analysis of the part correlations while ConditionXSize achieves significance. ConditionXType, Location, Tonality and ConditionXLocation were marginal (p<0.10). However, by inspection of the part correlations, Condition and ConditionXLocation clearly had the most to contribute. In the analysis of contrasts, Contrasts 2 and 5 for Condition were significant with Contrast 4 being close. The implications are that Condition 4 is not the same as Conditions 2 and 3 (see Table 3.115), that Condition 0 not the same as Condition 5, and the Conditions 1, 2, 3 and 4 are not the same as Conditions 0 and 5. For the interaction, the two significant contrast implied that the pattern across Conditions 2 and 3 verse 4 changes as a function of Type and that the pattern between Condition 1 and 5 changes as a function of Type. Of the remaining contrasts, only the fourth Tonality contrast was significant, implying that changes from triad to non-diatonic notes were processed more accurately than changes from diatonic (non-triad) to triad notes. Strangely, the same mean values were not significantly different for changes from triad to non-diatonic notes were not processed more accurately than changes from diatonic (non-triad) to diatonic (non-triad) notes -- such are the trials of unequal n analyses.

Table 3.122: The regression analysis of Condition, Type, ConditionXType, Location, Size and Tonality and the interactions of Condition with Location, Size and Tonality for Group 1, Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model							
Effect	ΔR^2	df	$F(\Delta R^2)$	$p(^{\Delta R2})$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.075	22	1.007	0.456			
Melody Pos.	0.032	2	4.858	0.008			
Sequence Pos.	0.003	2	0.467	0.627			
Condition	0.069	6	3.666	0.002	0.057	3.672	0.017
Type	0.007	1	2.087	0.150	0.004	1.645	0.201
ConditionXType	0.033	6	1.778	0.104	0.028	1.829	0.095
Location	0.072	3	8.523	0.000	0.017	2.175	0.092
Size	0.039	3	4.795	0.003	0.015	1.902	0.130
Tonality	0.019	4	1.760	0.138	0.020	1.976	0.099
CondXLocation	0.064	14	1.802	0.039	0.057	1.587	0.084
ConditionXSize	0.012	4	1.193	0.315	0.011	1.030	0.039
CondXTonality	0.008	7	0.489	0.842	0.009	0.489	0.842
Pooled Residual	0.567	220					
Total	1.000	294					

Notes: The effects are listed in the order in which they entered the equation.

In Tonality Group 2 (Table 3.123), Condition, ConditionXType, Location and Tonality were significant when assessed as the incremental change in R^2, but none remain so in the analysis of the part correlations, though Type and Location are close ($p<0.10$). By inspection of the part correlations, ConditionXLocation and ConditionXType clearly had the most to contribute. In the analysis of contrasts, Contrasts 1, 2 and 3 for Condition were significant. The implications are that Conditions 2 and 3 were not equivalent (see

Table 3.129), that Conditions 2 and 3 together were different from Condition 4 and that Condition 1 was not the same as Conditions 2 and 3 together. From the interaction contrasts, it is implied that the pattern across Conditions 2 and 3 changed as a function of Type and that the pattern between Condition 1 and Conditions 2 and 3 changed as a function of Type.

Table 3.123: The regression analysis of Condition, Type, ConditionXType, Location, Size and Tonality and the interactions of Condition with Location, Size and Tonality for Group 2, Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

	Hierarchical Regression Analysis: Regression Model						
Effect	ΔR^2	df	$F(\Delta R^2)$	$p(^{\Delta R2})$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.050	17	0.681	0.820			
Melody Pos.	0.071	2	8.748	0.000			
Sequence Pos.	0.003	2	0.388	0.679			
Condition	0.051	6	2.171	0.047	0.022	1.099	0.365
Type	0.009	1	2.232	0.137	0.012	3.479	0.064
ConditionXType	0.058	6	2.581	0.020	0.031	1.562	0.161
Location	0.049	3	4.630	0.004	0.022	2.208	0.089
Size	0.016	3	1.484	0.220	0.013	1.268	0.287
Tonality	0.048	4	3.550	0.008	0.022	1.696	0.153
CondXLocation	0.051	10	1.569	0.119	0.050	1.523	0.135
ConditionXSize	0.008	4	0.581	0.677	0.017	1.260	0.287
CondXTonality	0.011	4	0.806	0.523	0.011	0.806	0.523
Pooled Residual	0.575	174					
Total	1.000	236					

Notes: The effects are listed in the order in which they entered the equation.

In Tonality Group 4 (Table 3.124), there were no significant effects, although Size did approach significance when assessed as the incremental change in R^2, and Type did approach significance when tested as a part correlation ($p<0.10$). By comparison with Groups 1 and 2, it is likely that this was an effect of the limited number of subjects (8). Of the contrasts, the third contrast for Size was significant implying that small changes produced better performance that large changes. Although an effect of Size was expected for this group, this effect is actually the opposite of that which might be predicted (small changes produced better performance).

Table 3.124: The regression analysis of Condition, Type, ConditionXType, Location, Size and Tonality and the interactions of Condition with Location, Size and Tonality for Group 4, Stage 4, Part 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

	Hierarchical Regression Analysis: Regression Model						
Effect	ΔR^2	df	$F(\Delta R^2)$	$p(^{\Delta R2})$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.070	7	1.039	0.410			
Melody Pos.	0.095	2	5.381	0.006			
Sequence Pos.	0.005	2	0.266	0.767			
Condition	0.013	6	0.238	0.963	0.036	0.662	0.681
Type	0.006	1	0.642	0.425	0.034	3.710	0.058
ConditionXType	0.070	6	1.258	0.286	0.034	0.616	0.717
Location	0.030	3	1.095	0.356	0.024	0.868	0.462
Size	0.067	3	2.593	0.059	0.056	2.052	0.115
Tonality	0.022	3	0.845	0.474	0.010	0.360	0.782
CondXLocation	0.019	3	0.732	0.536	0.018	0.647	0.588
ConditionXSize	0.001	1	0.115	0.736	0.000	0.014	0.908
CondXTonality	0.000	1	0.043	0.836	0.000	0.043	0.836
Pooled Residual	0.600	66					
Total	1.000	104					

Notes: The effects are listed in the order in which they entered the equation.

Boundary Location Within Groups: Stages 6 and 8, Part 1

In Stages 6 and 8, there were three repetitions of the same song within two stages. Hence the analyses of boundary location has six repetitions to deal with, as well as, the effect of Session (Stage 6 to 8).

For all groups, the mean consistency values for subject-by-subject repetitions of the same melody within each of the groups, using the phi similarity measure, are presented in Tables 3.125 (Group 1), 3.126 (Group 2) and 3.127 (Group 4). Clearly all values indicated a high degree of consistency within individual subjects, and as would be expected, the consistencies increase with repetition number.

As in Stage 4, it is in the two triadic groups (Groups 1 and 2) that one found the highest number of modal values equal to 1.00: Group 2 had definitely the highest consistency (particularly in the latter repetitions), but more generally, those subjects who show the greatest degree of melody to melody consistency were confined to the Triadic groups.

Table 3.125: The average similarity ratings for individual subjects for the different repetitions of the same melody within Stage 6 and 8, Part 1, given the Phi similarity measure of Table 3.16, within Group 1 (n=23).

Stage	Rep	Mean±SD	Median	Mode	Min	Max
6	1 to 2	0.642±0.200	0.610	0.549	0.199	1.000
	1 to 3	0.642±0.200	0.610	0.549	0.199	1.000
	2 to 3	0.693±0.174	0.717	0.746	0.404	1.000
6 to 8	1 to 4	0.642±0.200	0.610	0.549	0.199	1.000
	1 to 5	0.642±0.200	0.610	0.549	0.199	1.000
	1 to 6	0.642±0.200	0.610	0.549	0.199	1.000
	2 to 4	0.664±0.176	0.673	0.266	0.266	1.000
	2 to 5	0.726±0.168	0.718	1.000	0.456	1.000
	2 to 6	0.691±0.220	0.746	1.000	0.355	1.000
	3 to 4	0.684±0.214	0.679	0.897	0.139	1.000
	3 to 5	0.732±0.183	0.746	1.000	0.364	1.000
	3 to 6	0.707±0.277	0.789	1.000	0.704	1.000
8	4 to 5	0.705±0.219	0.766	0.789	0.227	1.000
	4 to 6	0.664±0.228	0.640	1.000	0.199	1.000
	5 to 6	0.754±0.194	0.816	0.897	0.393	1.000

Table 3.126: The average similarity ratings for individual subjects for the different repetitions of the same melody within Stage 6 and 8, Part 1, given the Phi similarity measure of Table 3.8, within Group 2 (n=18).

Stage	Rep	Mean±SD	Median	Mode	Min	Max
6	1 to 2	0.646±0.195	0.589	0.879	0.379	1.000
	1 to 3	0.646±0.195	0.589	0.879	0.379	1.000
	2 to 3	0.721±0.210	0.766	0.461	0.364	1.000
6 to 8	1 to 4	0.646±0.195	0.589	0.879	0.379	1.000
	1 to 5	0.646±0.195	0.589	0.879	0.379	1.000
	1 to 6	0.646±0.195	0.589	0.879	0.379	1.000
	2 to 4	0.742±0.220	0.800	0.897	0.244	1.000
	2 to 5	0.752±0.254	0.866	0.8794	0.209	1.000
	2 to 6	0.762±0.212	0.826	0.879	0.244	1.000
	3 to 4	0.745±0.211	0.766	1.000	0.337	1.000
	3 to 5	0.778±0.247	0.825	1.000	0.280	1.000
	3 to 6	0.718±0.269	0.812	1.000	0.337	1.000
8	4 to 5	0.772±0.204	0.782	1.000	0.198	1.000
	4 to 6	0.743±0.267	0.800	1.000	0.146	1.000
	5 to 6	0.787±0.231	0.875	1.000	0.244	1.000

Table 3.127: The average similarity ratings for individual subjects for the different repetitions of the same melody within Stage 6 and 8, Part 1, given the Phi similarity measure of Table 3.8, within Group 4 (n=8).

Stage	Rep	Mean±SD	Median	Mode	Min	Max
6	1 to 2	0.532±0.204	0.600	0.107	0.107	0.707
	1 to 3	0.532±0.204	0.600	0.107	0.107	0.707
	2 to 3	0.624±0.295	0.603	0.106	0.106	1.000
6 to 8	1 to 4	0.532±0.204	0.600	0.107	0.107	0.707
	1 to 5	0.532±0.204	0.600	0.107	0.107	0.707
	1 to 6	0.532±0.204	0.600	0.107	0.107	0.707
	2 to 4	0.564±0.266	0.538	0.106	0.106	1.000
	2 to 5	0.624±0.291	0.646	0.610	0.610	1.000
	2 to 6	0.569±0.357	0.582	1.000	0.126	1.000
	3 to 4	0.612±0.233	0.593	0.191	0.191	1.000
	3 to 5	0.699±0.207	0.693	0.337	0.337	1.000
	3 to 6	0.615±0.214	0.593	0.491	0.289	1.000
8	4 to 5	0.617±0.213	0.611	0.337	0.337	1.000
	4 to 6	0.614±0.264	0.625	0.107	0.107	1.000
	5 to 6	0.516±0.294	0.479	0.404	0.605	1.000

In general, Group 4 (labeled as Proximity) had lower consistencies, but then, all groups had lower consistencies in Stages 6 and 8 than in Stage 4 and all groups had lower consistencies in Stage 6 than in Stage 8. The analysis of the change in consistency ratings (Table 3.128) demonstrated that only Group 2 showed changes in consistency, basically between Stages 6 and 8 (Contrast 5 of Table 3.128): this may seem paradoxical, but the reason is that in Stage 8, subjects in Group 2 became extremely consistent, and hence, even small deviations can be noted. The ω^2 of 11% was similar to the value found in Stage 4, again indicating that this change was not a minor effect. Group 1 did have one significant contrast indicating that the consistency in Stage 8 nearest neighbours in time was higher than those further apart in time.

Table 3.128: The ws-analysis of the change in consistency ratings (Phi measure) for the different repetitions of the same melody within Stage 6 and 8. Part 1.

WS-ANOVA		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Group 1	Overall	0.273	14	3.654	308	1.643	0.067	0.070	0.027
	Ψ_1	0.015	1	0.375	22	0.895	0.354	0.039	
	Ψ_2	0.005	1	0.125	22	0.895	0.354	0.039	
	Ψ_3	0.016	1	0.199	22	1.789	0.195	0.075	
	Ψ_4	0.036	1	0.166	22	4.837	0.039	0.180	
	Ψ_5	0.054	1	0.325	22	3.681	0.068	0.143	
Group 2	Overall	0.507	14	2.673	238	3.223	0.000	0.159	0.110
	Ψ_1	0.031	1	0.180	17	2.971	0.103	0.148	
	Ψ_2	0.010	1	0.060	17	2.971	0.103	0.149	
	Ψ_3	0.002	1	0.164	17	0.230	0.637	0.013	
	Ψ_4	0.006	1	0.146	17	0.725	0.406	0.041	
	Ψ_5	0.173	1	0.283	17	10.389	0.005	0.379	
Group 4	Overall	0.171	14	1.515	98	0.792	0.676	0.102	0.000
	Ψ_1	0.023	1	0.249	7	0.658	0.444	0.086	
	Ψ_2	0.009	1	0.083	7	0.658	0.444	0.086	
	Ψ_3	0.016	1	0.126	7	0.898	0.375	0.114	
	Ψ_4	0.006	1	0.078	7	0.546	0.484	0.072	
	Ψ_5	0.003	1	0.083	7	0.284	0.610	0.039	

Notes: All analyses were performed on log-transformed scores.
Ψ_1 & Ψ_3 compare Consistency 1 to 2 (4 to 5) with 2 to 3 (5 to 6).
Ψ_2 & Ψ_4 compare the average Consistency of 1 to 2 (4 to 5) and 2 to 3 (5 to 6) with Consistency 1 to 3 (4 to 6).
Ψ_5 compares the averages within Stage 6 to those within Stage 8.

The analysis of the differences between groups (Table 3.129) shows no effect of Group and no interaction between Group and Repetition (the significant main effect of Repetition is the aforementioned effect of Group 2).

Table 3.129: The overall analysis of the change in consistency ratings (Phi measure) for the different repetitions of the same melody within Stages 6 and 8, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	0.85	2	0.42	1.82	0.174	0.073	0.006
Subjects	Ψ_7	0.20	1	0.20	0.85	0.362	0.017	
	Ψ_8	0.65	1	0.65	2.79	0.102	0.056	
	$\text{Error}_{B\text{-}S}$	10.74	46	0.23				
Within	Repetition	0.78	14	0.06	4.56	0.000	0.090	0.069
Subjects	Ψ_1	0.06	1	0.06	3.69	0.061	0.074	
	Ψ_2	0.02	1	0.02	3.68	0.061	0.074	
	Ψ_3	0.00	1	0.00	0.39	0.535	0.008	
	Ψ_4	0.02	1	0.02	2.55	0.117	0.052	
	Ψ_5	0.19	1	0.19	12.60	0.001	0.218	
	RepetitionXGroups	0.17	28	0.01	0.51	0.984	0.022	0.000
	$\text{Error}_{W\text{-}S}$	7.84	644	0.01				
	$\text{Error}_{\Psi 1}$	0.80	46	0.80				
	$\text{Error}_{\Psi 2}$	0.27	46	0.27				
	$\text{Error}_{\Psi 3}$	0.49	46	0.49				
	$\text{Error}_{\Psi 4}$	0.39	46	0.39				
	$\text{Error}_{\Psi 5}$	0.69	46	0.69				
Total		20.38	734					

Notes: All analyses were performed on log-transformed scores.

Ψ_1 & Ψ_3 compare Consistency 1 to 2 (4 to 5) with 2 to 3 (5 to 6).

Ψ_2 & Ψ_4 compare the average Consistency of 1 to 2 (4 to 5) and 2 to 3 (5 to 6) with Consistency 1 to 3 (4 to 6).

Ψ_5 compares the averages within Stage 6 to those within Stage 8.

Ψ_7 compares Group 1 with Group 2 (i.e., the two triadic groups)

Ψ_8 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group)

No violations of the ws-anova were noted

When comparing across subjects (Table 3.130), average similarity ratings within all groups were lower in Stages 6 and 8 than in Stage 4 (as expected), with minima that dip below the aforementioned critical values. However, with only one exception, these values are still above the chance levels specified in Table 3.8.

Table 3.130: The average similarity ratings using the Phi similarity measure (see Table 3.8) between different subjects for the same melody, for each group, within Stages 6 and 8, Part 1, Repetition 3 only.

		Mean±SD	Median	Mode	Min	Max
Group 1	Stage 6	0.597±0.218	0.610	0.897	-0.133	1.000
	Stage 8	0.687±0.202	0.718	1.000	0.199	1.000
Group 2	Stage 6	0.674±0.187	0.679	0.987	0.243	1.000
	Stage 8	0.680±0.224	0.679	1.000	0.039	1.000
Group 4	Stage 6	0.609±0.213	0.643	0.278	0.150	1.000
	Stage 8	0.463±0.215	0.494	0.518	0.098	0.897

Because two stages were involved, the change in similarity across stages could be examined. Table 3.131 presents the correlations between stages and the associated analysis of the change in similarity ratings. Group 4 showed a significant decrease in the similarity ratings across these stages (which is a bit odd). In these computations, all possible pairing of subjects (within a group) are compared across Stages: For example, Group 1 contains 23 subjects and therefore 23*22/2=253 possible pairs.

Table 3.131: The within-groups analyses of differences in mean similarity rating (Groups 1, 2 and 4 only) and the associated correlations between Stages 6 and 8, within each group.

Correlations	
Group 1	0.481**
Group 2	0.453**
Group 4	0.184

WS-ANOVA	SS	df	SS	df	*F*	*p*	η^2	ω^2
Group 1	0.55	1	3.07	252	45.42	0.143	0.149	0.006
Group 2	0.01	1	1.95	152	0.49	0.486	0.003	0.000
Group 4	0.14	1	0.49	27	7.93	0.009	0.227	0.193

Notes: The analyses were performed on log-transformed scores.
* Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Statistically, the three main groups all differed on their mean similarity ratings (Table 3.132): The main effect of Group and both its contrasts, the main effect of Session (Stage 6 to 8), and the interaction of Group and Session were all significant, but given the ω^2 for these results, the results are not too important (remember that Session can be considered a within-subjects experimental variable: as such, large effects sizes would be expected)

Table 3.132: The mixed-analysis of the differences in mean similarity rating in Stages 6 and 8, Part 1.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between Subjects	Groups	0.51	2	0.26	7.17	0.001	0.032	0.005
	Ψ_7	0.10	1	0.11	2.81	0.006		
	Ψ_8	0.41	1	0.11	11.53	0.026		
	$\text{Error}_{B\text{-}S}$	15.38	431	0.04				
Within Subjects	Stage	0.27	1	0.27	21.04	0.000	0.047	0.041
	StageXGroup	0.43	2	0.22	17.00	0.000	0.073	0.063
	$\text{Error}_{W\text{-}S}$	5.51	431	0.01				
Total		22.10	867					

Notes: The analysis was performed on log-transformed scores.
Ψ_7 compares Group 1 with Group 2 (i.e., the two triadic groups)
Ψ_8 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group).

As was done in Stage 4, Table 3.133 presents the average similarity ratings between subjects in different groups. Note that these were not truly different from the similarities within groups, and that as in Stage 4, Groups 1 and 2 were more similar to each other than Group 4 was to either Group 1 or 2. It also seems that the similarity between Groups 1 and 2 increased from Stage 6 to 8 while the similarity of Group 4 with either Groups 1 and 2 decreased from Stage 6 to 8. It is possible that more complex melodies or more repetitions might demonstrate greater divergence between these groups. More complex melodies might require greater use of tonality. More repetitions might allow subjects to better encode the melody which, in turn, would be facilitated by an understanding of tonality.

Table 3.133: The average similarity rating using the Phi measure between subjects in different groups.

		Mean±SD	Median	Mode	Min	Max
Group 1 with Group 2 n=414	Stage 6	0.643±0.204	0.622	0.897	0.101	1.000
	Stage 8	0.688±0.216	0.679	1.000	0.040	1.000
Group 1 with Group 4 n=184	Stage 6	0.600±0.233	0.622	0.897	-0.035	1.000
	Stage 8	0.547±0.228	0.553	0.897	0.072	1.000
Group 2 with Group 4 n=144	Stage 6	0.633±0.216	0.658	0.897	0.052	1.000
	Stage 8	0.541±0.222	0.553	0.357	0.072	1.000

Notes: The n's represent the number of pairs involved in each calculation, given that Groups 1, 2 and 3 had 23, 18 and 8 subjects, respectively.

Generally, the conclusion is that although there are minor differences in similarity, all groups are pretty consistent in the degree of similarity shown from Stage 6 to Stage 8. Given the results in Tables 3.130 and 3.131, Group 4 evidenced less similarity over time; hence, it is possible that Group 4 was in the process of partitioning into two or more groups. More data would be need to assess this inference properly (i.e., more subjects in Group 4; remember that the comparison of Groups is essentially a quasi-experimental variable, and as such control over the number of subjects cannot be insured). The plots of boundary profiles within groups were very similar to the overall boundary profiles so they are not included here.

Minimum, maximum and mean distances between boundaries were analyzed within each group. In the mixed ANOVAs, the effects of Repetition collapsed across groups, resemble those of the analysis overall subjects because these three groups represent 49 of the 61 subjects. Differences between groups would be demonstrated by the effect of Group and more importantly by the interaction of Group and Repetition. To explore differences between groups, each analysis was also conducted as a simple effects analysis within each group.

The overall analysis of the change in minimum distance between groups is shown in Tables 3.134 and 3.135. There is only a minor overall difference between groups

(Groups 1 & 2 are different from 4), and the significant effect of Repetition is the aforementioned patterns between groups. There was no interaction between Groups and Repetition. Simple effects analysis indicated that minimum distance changed in Group 1 ($F(5,110)=2.21, p<.058$), but this change was limited to Stage 6. There was also a significant change in Group 2 ($F(5,85)=2.85, p<.020$), due to the fact that Repetition 1 was very different from the rest. There was no change in Group 4 ($F(5,35)= p<..580$). Effect sizes were not large in any group (($\omega^2=0.05$, 0.09 and 0.00, respectively)

Table 3.134: The important statistics for the minimum distance between boundaries, for each group, within Stages 6 and 8, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Stage 6: Repetition 1	Group 1	6.000±8.868	5.000	0.000	0.000	32.000
	Group 2	7.500±9.389	5.500	8.000	0.000	32.000
	Group 4	1.500±2.268	0.000	0.000	0.000	5.000
Stage 6: Repetition 2	Group 1	3.043±3.052	1.000	0.000	0.000	8.000
	Group 2	3.667±3.865	3.500	0.000	0.000	14.000
	Group 4	2.500±2.928	1.000	0.000	0.000	6.000
Stage 6: Repetition 3	Group 1	0.739±0.449	1.000	1.000	0.000	1.000
	Group 2	1.389±2.173	1.000	1.000	0.000	10.000
	Group 4	0.625±0.518	1.000	1.000	0.000	1.000
Stage 8: Repetition 1	Group 1	5.739±8.817	4.000	0.000	0.000	32.000
	Group 2	3.833±3.808	4.000	0.000	0.000	14.000
	Group 4	1.750±2.550	0.000	0.000	0.000	6.000
Stage 8: Repetition 2	Group 1	2.826±3.010	1.000	0.000	0.000	8.000
	Group 2	3.278±2.803	4.000	6.000	0.000	6.000
	Group 4	1.750±2.053	1.500	0.000	0.000	6.000
Stage 8: Repetition 3	Group 1	4.000±3.729	5.000	6.000	0.000	15.000
	Group 2	4.389±3.616	6.000	6.000	0.000	14.000
	Group 4	1.250±2.053	0.500	0.000	0.000	6.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.135: The overall analysis of the changes in the minimum distance between boundaries, across repetitions of the same melody, for Stages 6 and 8, Part 1, as a function of group.

Overall Analysis		SS	df	MS	F	p	η^2	ω^2
Between	Groups	248.76	2	139.23	2.18	0.124	0.087	0.020
Subjects	Ψ_6	7.84	1	7.84	0.12	0.728	0.002	
	Ψ_7	270.61	1	270.61	4.24	0.045	0.085	
	$\text{Error}_{B\text{-}S}$	2935.91	46	63.82				
Within	Repetition	348.90	5	69.78	3.78	0.003	0.076	0.054
Subjects	Ψ_1	169.81	1	169.81	5.20	0.027	0.102	
	Ψ_2	113.91	1	113.91	9.04	0.004	0.164	
	Ψ_3	11.80	1	11.80	0.52	0.474	0.011	
	Ψ_4	48.98	1	48.98	3.99	0.052	0.080	
	Ψ_5	4.41	1	4.41	0.36	0.549	0.008	
	RepetitionXGrp	165.19	10	16.52	0.89	0.539	0.037	0.054
	$\text{Error}_{W\text{-}S}$	4245.91	230	18.46				
	$\text{Error}_{\Psi 1}$	1501.73	46	32.65				
	$\text{Error}_{\Psi 2}$	579.76	46	12.60				
	$\text{Error}_{\Psi 3}$	1042.44	46	22.66				
	$\text{Error}_{\Psi 4}$	564.97	46	12.28				
	$\text{Error}_{\Psi 5}$	557.00	46	12.11				
Total		7944.67	293					

Notes: Ψ_1 & (Ψ_3) compare minimum in Repetitions 1 to 3 (4 to 6) (furthest apart in time).

Ψ_2 & (Ψ_4) compare average minimum in Repetition 1 & 3 to 2 (4 & 6 to 5).

Ψ_5 compares the averages within Stage 6 to those within Stage 8.

Ψ_6 compares Group 1 with Group 2 (i.e., the two triadic groups)

Ψ_7 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group).

The Mauchly sphericity test was rejected and the multivariate results pertaining to repetition were not significant (p<.079) while the conclusion pertaining to the interaction was unchanged.

The change in maximum distance between boundaries is shown in Tables 3.136 and 3.137. There were no differences between groups (Group 4 was marginally different from Groups 1 & 2). There was a significant effect of Repetition and an interaction

RepetitionXGroup. Simple effects analysis revealed that the maximum distance changed in Group 1 (*F*(5,110)=3.79, *p*<.003) and Group 2 (*F*(5,85)=12.97, *p*<.000) but not in Group 4 (*F*(5,35)=0.63, *p*<.661). Effect sizes in Groups 1 and 2 were quite large (ω^2=0.10 and 0.40, respectively) but not in Group 4 (ω^2=0.00), so the difference is not an issue of sample size. The main effect seems to be a dramatic reduction from Repetition 1 to 2, and relative stability thereafter (in both groups). Hence, the maximum size of units gets smaller as these subjects become more familiar with the melody.

Table 3.136: The important statistics for the maximum distance between boundaries, for each group, within Stages 6 and 8, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Stage 6: Repetition 1	Group 1	13.261±7.823	9.000	8.000	5.000	32.000
	Group 2	15.389±8.197	14.000	8.000	7.000	32.000
	Group 4	8.375±0.518	8.000	8.000	8.000	9.000
Stage 6: Repetition 2	Group 1	9.870±4.475	8.000	8.000	5.000	26.000
	Group 2	10.222±3.457	8.000	8.000	7.000	17.000
	Group 4	8.375±2.875	8.000	8.000	5.000	15.000
Stage 6: Repetition 3	Group 1	9.174±2.516	8.000	8.000	8.000	17.000
	Group 2	8.722±2.562	8.000	8.000	7.000	17.000
	Group 4	7.750±1.389	8.000	8.000	5.000	10.000
Stage 8: Repetition 1	Group 1	11.565±7.335	8.000	8.000	7.000	32.000
	Group 2	8.722±2.585	8.000	8.000	6.000	17.000
	Group 4	8.875±3.482	8.000	8.000	5.000	17.000
Stage 8: Repetition 2	Group 1	8.870±2.599	8.000	8.000	5.000	17.000
	Group 2	8.111±2.423	8.000	8.000	5.000	17.000
	Group 4	7.625±0.744	8.000	8.000	6.000	8.000
Stage 8: Repetition 3	Group 1	9.000±2.860	8.000	8.000	4.000	17.000
	Group 2	8.167±2.455	8.000	8.000	5.000	17.000
	Group 4	7.750±1.035	8.000	8.000	6.000	9.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.137: The overall analysis of the changes in the maximum distance between boundaries, across repetitions of the same melody, for Stages 6 and 8, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between Subjects	Groups	168.60	2	84.30	1.70	0.193	0.069	0.002
	Ψ_6	14.00	1	14.00	0.28	0.597	0.006	
	Ψ_7	154.60	1	154.60	3.13	0.084	0.063	
	$\text{Error}_{B\text{-}S}$	2275.66	46	49.47				
Within Subjects	Repetition	818.72	5	163.74	12.13	0.000	0.209	0.179
	Ψ_1	489.40	1	489.40	19.11	0.000	0.294	
	Ψ_2	51.46	1	51.46	3.66	0.062	0.074	
	Ψ_3	62.08	1	62.08	4.17	0.047	0.083	
	Ψ_4	26.34	1	26.34	8.16	0.006	0.151	
	Ψ_5	189.44	1	189.44	19.52	0.000	0.298	
	RepetitionXGrp	254.60	10	25.46	1.89	0.048	0.076	0.029
	$\text{Error}_{W\text{-}S}$	3105.34	230	13.50				
	$\text{Error}_{\Psi 1}$	1177.85	46	25.61				
	$\text{Error}_{\Psi 2}$	647.12	46	14.07				
	$\text{Error}_{\Psi 3}$	685.49	46	14.90				
	$\text{Error}_{\Psi 4}$	148.53	46	8.16				
	$\text{Error}_{\Psi 5}$	446.35	46	19.52				
Total		6622.92	293					

Notes: see Table 3.148

The Mauchly sphericity test was rejected and the multivariate results pertaining to the interaction were not significant ($p<.331$), while the conclusion pertaining to repetition was unchanged.

Finally, the change in the mean distance between boundaries is presented in Tables 3.138 and 3.139. There were no overall differences between groups, a significant effect of Repetition and no interaction. Simple effects analysis within each group indicated that Group 1 changed over repetitions ($F(5,110)=2.76$, $p<.022$) as did Group 2 ($F(5,85)=6.87$, $p<.000$). Both Groups 1 and 2 showed a pattern of higher means in the first repetition of a stage, followed by much lower means in the second repetition of a stage, followed by a bit of an increase in the third repetition. It is almost as if subjects parsed too crudely, then parsed too finely, and then settled on a final optimal parsing. The effect sizes were moderate (ω^2=0.07 and 0.24, respectively). Group 4 did not show any effects (*F*(5,35)=1.36, *p*<.262; ω^2=0.04), although the effect size is close to that of Group 1.

Table 3.138: The important statistics for the mean distance between boundaries, for each group, within Stages 6 and 8, Part 1, Repetitions 1, 2 and 3.

		Mean±SD	Median	Mode	Min	Max
Stage 6: Repetition 1	Group 1	9.523±7.797	7.250	7.250	2.300	32.000
	Group 2	11.181±8.232	8.625	7.250	2.667	32.000
	Group 4	5.158±1.677	5.600	5.600	3.125	7.250
Stage 6: Repetition 2	Group 1	5.943±2.939	5.600	3.714	1.750	15.500
	Group 2	7.012±3.041	7.250	7.250	3.125	15.500
	Group 4	4.958±1.854	5.050	4.500	2.300	7.250
Stage 6: Repetition 3	Group 1	6.176±2.154	5.600	7.250	2.000	10.000
	Group 2	6.639±2.783	6.425	7.250	3.125	15.500
	Group 4	5.039±2.077	5.050	7.250	2.667	7.250
Stage 8: Repetition 1	Group 1	8.418±7.981	5.600	7.250	2.000	32.000
	Group 2	6.639±2.783	6.425	7.250	3.125	15.500
	Group 4	5.381±2.349	5.050	3.714	2.667	10.000
Stage 8: Repetition 2	Group 1	6.127±1.852	5.600	7.250	2.000	10.000
	Group 2	6.033±1.949	6.425	7.250	2.300	10.000
	Group 4	4.588±1.264	4.107	3.714	3.714	7.250
Stage 8: Repetition 3	Group 1	6.714±2.652	7.250	7.250	2.300	15.500
	Group 2	6.531±2.807	7.250	7.250	2.667	15.500
	Group 4	4.212±1.535	3.714	3.125	2.667	7.250

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Table 3.139: The overall analysis of the changes in the mean distance between boundaries, across repetitions of the same melody, for Stages 6 and 8, Part 1, as a function of group.

Overall Analysis		SS	df	MS	*F*	*p*	η^2	ω^2
Between	Groups	222.81	2	111.41	2.20	0.122	0.087	0.007
Subjects	Ψ_6	0.54	1	0.54	0.00	0.374	0.000	
	Ψ_7	222.28	1	222.28	1.25	0.270	0.087	
	Error$_{B\text{-}S}$	2326.65	46	50.58				
Within	Repetition	446.01	5	89.20	7.21	0.000	0.135	0.112
Subjects	Ψ_1	260.27	1	260.27	11.52	0.001	0.200	
	Ψ_2	85.20	1	85.20	9.03	0.004	0.164	
	Ψ_3	26.01	1	26.01	1.59	0.213	0.033	
	Ψ_4	27.18	1	27.18	4.79	0.034	0.094	
	Ψ_5	47.34	1	47.34	6.03	0.018	0.116	
	RepetitionXGrp	139.20	10	13.92	1.12	0.334	0.047	0.004
	Error$_{W\text{-}S}$	2846.06	230	12.37				
	Error$_{\Psi 1}$	1039.07	46	22.59				
	Error$_{\Psi 2}$	434.25	46	9.44				
	Error$_{\Psi 3}$	760.56	46	16.32				
	Error$_{\Psi 4}$	261.04	46	5.67				
	Error$_{\Psi 5}$	361.05	46	7.85				
Total		5980.73	293					

Notes: see Table 3.148

The Mauchly sphericity test was rejected and the multivariate results pertaining to repetition were not significant ($p<.057$) while the conclusion pertaining to the interaction was unchanged.

Each of the groups was analyzed with respect to the rules of Lerdahl and Jackendoff (1983). As in the overall analysis, the first analysis examined the correlation between each of the rules and the average empirical boundary profiles (within each group). As in Stage 4, only the optimal shift is presented (it is universally the same as in the overall analysis). These correlations are presented in Table 3.140.

Table 3.140: The correlations between the average profile and the predicted boundaries, given the rules of Lerdahl and Jackendoff, within each group in Stages 6 and 8, Part 1.

	Stage 6			Stage 8		
	Rule 2b	Rule 3a	Rule 3d	Rule 2b	Rule 3a	Rule 3d
Group 1	0.734**	0.314	0.287	0.748**	0.333	0.256
Group 2	0.771**	0.314	0.240	0.755**	0.295	0.213
Group 4	0.739**	0.303	0.245	0.737**	0.276	0.313

Notes: Rules 2a, 3b, and 3c concerned attributes that did not change within this particular melody.
Only the optimal shift is shown.
significant at $p<.05$
significant at $p<.01$

Note that Groups 1, 2 and 4 had equivalent correlations. In addition, it can be observed that, although these correlations are lower, they are not dramatically different from those of Stage 4.

As was done in the overall analysis, the rules were combined to find the best fitting regression equation. The results, presented in Table 3.141 demonstrate that rules were equally effective in all groups and that, generally, only two rules were necessary.

Table 3.141: The regression equations for the prediction of boundaries (given the rules of Lerdahl and Jackendoff) in the average boundary profile within Groups 1, 2 and 4, in Stages 6 and 8, Part 1.

Group 1

Stg	Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
6	B=0.08±0.04 + (0.83±0.17)*Rule2b	0.539	0.000	0.539	37.406	0.000
	+ (0.05±0.11)*Rule3a	0.003	0.677	0.542	18.311	0.000
	+ (0.04±0.13)*Rule3d	0.002	0.751	0.543	11.888	0.000
8	B=0.06±0.04 + (0.92±0.17)*Rule2b	0.460	0.000	0.460	40.678	0.000
	+ (0.06±0.11)*Rule3a	0.004	0.581	0.564	20.056	0.000
	+ (0.00±0.14)*Rule3d			0.564	12.940	0.000

Group 2

Stg	Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
6	B=0.06±0.04 + (0.92±0.16)*Rule2b	0.594	0.000	0.594	46.739	0.000
	+ (0.03±0.10)*Rule3a	0.001	0.750	0.595	22.766	0.000
	- (0.03±0.13)*Rule3d			0.596	14.724	0.000
8	B=0.08±0.04 + (0.97±0.17)*Rule2b	0.570	0.000	0.570	46.461	0.000
	- (0.05±0.13)*Rule3d	0.002	0.682	0.573	20.767	0.000
	+ (0.02±0.12)*Rule3a			0.573	13.420	0.000

Group 4

Stg	Constant Rule Added at Step	ΔR^2	$p(\Delta R^2)$	R^2	F	$p(R^2)$
6	B=0.10±0.04 + (0.92±0.18)*Rule2b	0.547	0.000	0.547	38.562	0.000
	+ (0.03±0.12)*Rule3a	0.001	0.765	0.548	18.779	0.000
	- (0.01±0.14)*Rule3d			0.548	12.118	0.000
8	B=0.12±0.04 + (0.83±0.16)*Rule2b	0.543	0.000	0.543	37.966	0.000
	+ (0.07±0.13)*Rule3d	0.004	0.600	0.547	18.699	0.000
	+ (0.01±0.11)*Rule3a			0.546	12.069	0.000

Notes: Rules 2a, 3b and 3c concerned attributes that did not change in this melody. The complete equation represents the maximal equation when all rules were used. The rules above the dotted line were included when a forward stepwise regression analysis was used. The ΔR^2 is the improvement at each step in the stepwise solution.

Rule 2b was used by all groups, but for the second rule, there was competition between Rules 3a and 3d. Again, there is the odd revelation that the total use of the rules decreases with exposure to the melody. It is as if subjects used the rules to guide parsing initially, but then diverged as the melody became more familiar.

A mixed ANOVA on Groups 1, 2 and 4 across Sessions, for each of the rules (Table 3.142) did not demonstrate significant main effects (Groups or Session), but did demonstrate a significant interaction for the analysis of Rule 2b and Rule 3a: The relative declines (Group 2) or increases (Group 1) were different.

The final analysis examined the correlation between the use of the rules (Table 3.143) within groups (i.e., the utility of rules analysis; see Table 3.68). Note that in contrast to the previous melody (Stage 4), the use of the rules is not strongly correlated. Within each group, those who used Rule 2b (the dominant rule) did not necessarily use Rule 3a or 3d in proportion. It may be that the availability of two alternative eliminated the need for a strong correlation.

Table 3.142: The analyses of the difference between the use of rules (the correlations between the boundary profiles of subjects and the predicted boundaries given the rule of Lerdahl and Jackendoff) within Groups 1, 2 and 4. Stages 6 and 8, Part 1.

Rule 2b Analysis		SS	df	MS	F	p	η^2	ω^2
Between	Groups	0.029	2	0.015	1.57	0.219	0.064	0.001
Subjects	Ψ_3	0.009	1	0.009	1.01	0.320	0.020	
	Ψ_4	0.020	1	0.011	2.13	0.151	0.044	
	Error$_{B\text{-}S}$	0.425	46	0.009				
Within	Session	0.001	1	0.001	0.30	0.587	0.006	0.000
Subjects	Session X Group	0.030	2	0.004	3.89	0.027	0.145	0.105
	Error$_{W\text{-}S}$	0.175	46	0.004				
Total		0.660	97					

Rule 3a Analysis		SS	df	MS	F	p	η^2	ω^2
Between	Groups	0.010	2	0.005	1.25	0.295	0.052	0.000
Subjects	Ψ_3	0.001	1	0.001	0.20	0.660	0.002	
	Ψ_4	0.010	1	0.010	2.31	0.135	0.050	
	Error$_{B\text{-}S}$	0.188	46	0.004				
Within	Session	0.000	1	0.000	0.20	0.661	0.004	0.000
Subjects	Session X Group	0.010	2	0.004	4.43	0.017	0.162	0.123
	Error$_{W\text{-}S}$	0.049	46	0.001				
Total		0.257	97					

Rule 3d Analysis		SS	df	MS	F	p	η^2	ω^2
Between	Groups	0.006	2	0.003	1.22	0.305	0.050	0.000
Subjects	Ψ_3	0.006	1	0.006	2.39	0.129	0.050	
	Ψ_4	0.000	1	0.000	0.05	0.824	0.000	
	Error$_{B\text{-}S}$	0.116	46	0.003				
Within	Session	0.001	1	0.001	0.58	0.449	0.013	0.000
Subjects	Session X Group	0.002	2	0.001	0.59	0.560	0.025	0.000
	Error$_{W\text{-}S}$	0.091	46	0.002				
Total		0.216	97					

Notes: The analyses were performed on log-transformed scores.
Ψ_3 compares Group 1 with Group 2 (i.e., the two triadic groups)
Ψ_4 compares Groups 1 and 2 with Group 3 (the triadic groups verse the proximity group).

Table 3.143: The correlation across subjects (within each group) between the utility of the different rules for predicting boundaries in the individual boundary profile of each subject.

		Stage 6			Stage 8		
		Rule 2b	Rule 3a	Rule 3d	Rule 2b	Rule 3a	Rule 3d
Group 1 n=23	Rule 2b	1.000	0.732**	-0.179	1.000	0.175	0.046
	Rule 3a		1.000	0.234		1.000	0.007
Group 2 n=18	Rule 2b	1.000	0.754**	-0.121	1.000	0.767**	0.388
	Rule 3a		1.000	0.111		1.000	0.495*
Group 4 n=8	Rule 2b	1.000	0.595	0.133	1.000	0.740*	-0.754*
	Rule 3a		1.000	0.801*		1.000	-0.445

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

When the analysis was extended between sessions (Table 3.144), the consistency of rule use (rule utility) was somewhat ambiguous. Rules 2b and 3a were used consistently in the two sessions within Groups 1 and 4, but not 2. In Group 2, it 2 was the use of Rules 3a and 3d that was consistent across sessions. One must be wary when drawing inferences from correlations because the lack of range in either variate can cause the correlation to be lost (i.e., if all subjects in Group 2 used Rule 2b to about the same degree, in both sessions, then there may not be a strong correlation within the group). One must also remember that "significance" is a reflection of group sizes and not strength of efficacy.

Table 3.144: The correlations between the utilities of the rules predicting boundaries in the average boundary profile of subjects, across sessions, for Groups 1, 2 and 4 in Stages 6 and 8, Part 1.

	Stage 6 to 8 Correlation		
	Rule 2b	Rule 3a	Rule 3d
Group 1	0.558**	0.499*	0.032
Group 2	0.136	0.685**	0.686**
Group 4	0.703	0.706	0.043

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

As in Stage 4, there were few differences between the groups on boundary profiles. Figure 3.28 indicates why this is so by demonstrating that subjects who were clustered together in the tonality analysis (i.e., those who became members of the same group) do not always become closely associated in the boundary profile cluster analysis. That is, subjects who were in the same tonality group were not necessarily neighbours in the cluster analysis (only Stage 8 is presented).

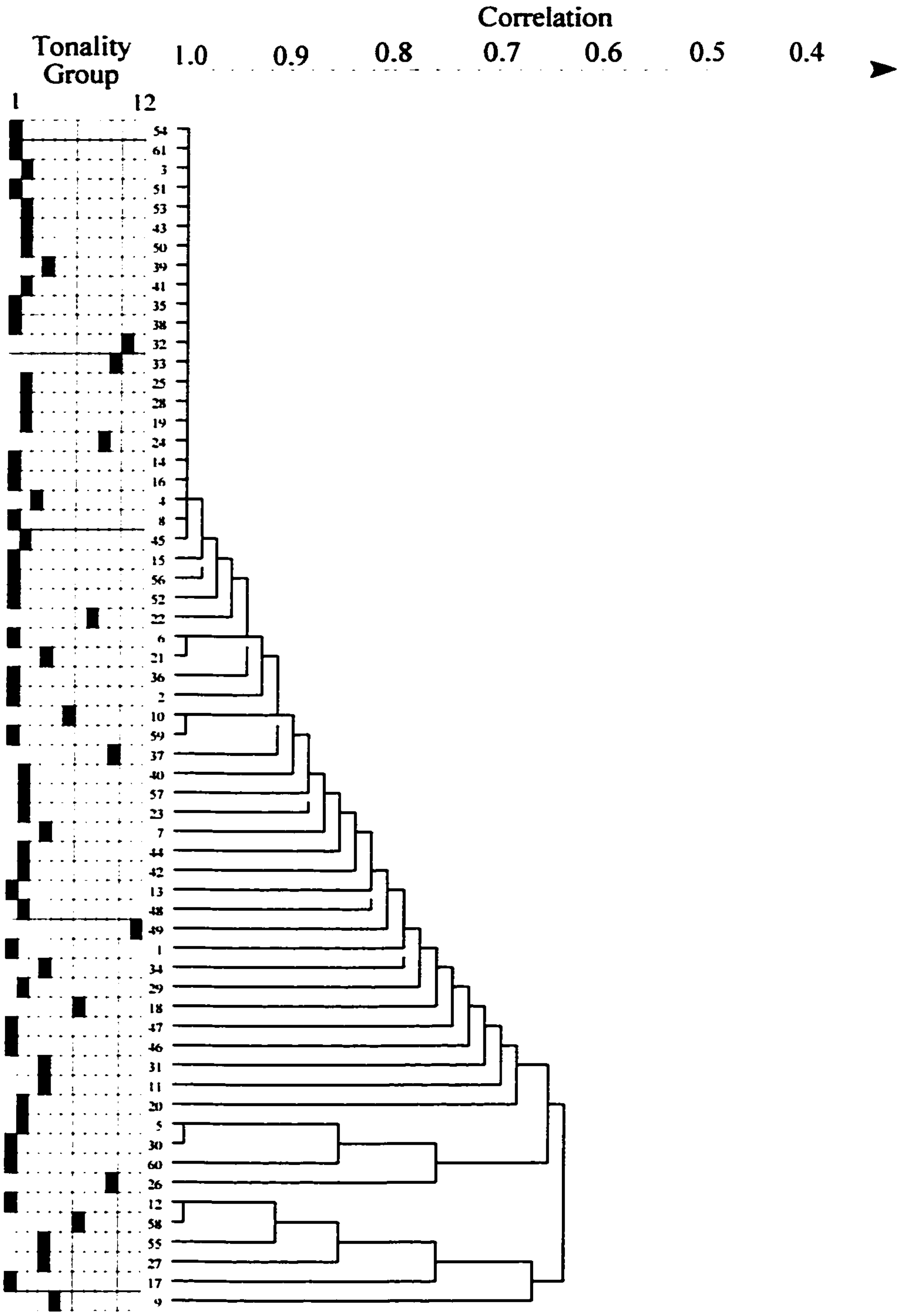


Figure 3.28: Comparison of cluster analysis for Boundary Location (phi measure, Stage 8) and group identification on the basis of Tonality Assessment (Tonality Groups 1 to 12). Note that members of the same tonality group are not necessarily adjacent in the boundary location.

Boundary Efficacy Within Groups: Stages 6 & 8, Part 2

Each subject was presented with 16 test sequences that varied on a number of dimensions: primarily seven Conditions (-1, 0, 1, 2, 3, 4 and 5) and two Types (Literal and Foil), but Type:Foil was further divided into Size, Location and Tonality. For Foil sequences, Location varied over 1, 2, 3 and 4 while Size varied over 2, 1, -1 and -2. Both Size and Location equal to 0 represented the Literal type. Tonality actually assessed the degree of change in tonality when the Foil sequence was created from its parent Literal (see Table 3.9). Assuming that the listener has a fair approximation of the true melody stored in memory, any changes that resulted in a large change in tonality should be more detectable than any changes that do not result in a large change in tonality (remember that all changes are relatively small in terms of absolute pitch). However, these predictions are only valid for those subjects demonstrating a triadic tonality profile: Tonality Groups 1 and 2. Proximity groups (Group 4) would be expected to be sensitive to changes in absolute size of change (since they, by definition, are not sensitive to tonality).

As was stated previously, the best insights are obtained by direct examination of the means, but the usual analysis was conducted. Initially, Session, Condition, Type, ConditionXType, Location, Size, Tonality, LocationXCondition, SizeXCondition and TonalityXCondition were analyzed within each group separately, using hierarchical multiple regression examining the incremental change in R^2 as the different terms were added to the equation and testing the part correlations (i.e., the semipartials) for each effect. The previous contrasts for each effect were used (see Tables 3.72 for Session, Condition and Type, 3.77 for Location, 3.80 for Size and 3.83 for Tonality). In addition to these effects, there were a changing number of subject vectors (using dummy coding) and four covariate vectors (first and second order for the serial position within the melody -- with 0 defined as the middle of the melody; first and second order for the serial position of the test sequence within the test phase -- with 0 defined as the middle of the set of test sequences). In the hierarchical analysis, subject effects were entered first, followed by the covariates, the main effects of Condition and Type, their interaction, Location, Size, Tonality and finally their interactions with Condition. Tonality was entered last because it was not controlled to the same extent as Location or Size. Size was entered just before Tonality (rather than first) because Size and Tonality represented a competition of sorts (Size and Tonality are both aspects of pitch height, while Location is separate from the two).

For the three groups, the means of Condition, Type and their interaction, collapsed over Session are presented Table 3.145 through 3.147.

Table 3.145: The means, standard deviations and cell counts for Condition, Type, ConditionXType, in Stages 6 & 8, Part 2, for Group 1.

				Group 1				
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.78±.42	.87±.34	.84±.37	.71±.46	.96±.21	.87±.34	.78±.42	.82±.38
	49	52	45	45	46	45	80	362
Foil	.67±.48	.53±.50	.60±.49	.65±.48	.62±.49	.32±.47	.95±.22	.60±.49
	24	105	43	43	42	41	42	340
Cond	.74±.44	.64±.48	.73±.45	.68±.47	.80±.41	.60±.49	.84±.37	.72±.45
	73	157	88	88	88	86	122	702

Table 3.146: The means, standard deviations and cell counts for Condition, Type, ConditionXType, in Stages 6 & 8, Part 2, for Group 2.

				Group 2				
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.68±.47	.84±.37	.92±.28	.69±.47	.89±.32	.86±.35	.83±.38	.82±.39
	38	38	36	36	36	36	64	284
Foil	.38±.49	.53±.50	.61±.50	.81±.40	.75±.44	.21±.42	.97±.18	.60±.49
	24	85	33	32	32	33	32	271
Cond	.56±.50	.63±.49	.77±.43	.75±.44	.82±.38	.55±.50	.88±.33	.71±.45
	62	123	69	68	68	69	96	555

Table 3.147: The means, standard deviations and cell counts for Condition, Type, ConditionXType, in Stages 6 & 8, Part 2, for Group 4.

				Group 4				
	Far	Before	Note 1	Note 2	Note 3	Note 4	After	Type
Liter	.94±.24	.94±.24	.88±.34	.88±.34	.94±.25	1.0±.00	.72±.45	.88±.32
	18	18	16	16	16	15	29	128
Foil	.33±.50	.28±.46	.50±.52	.43±.51	.15±.38	.43±.51	.47±.52	.36±.48
	9	32	14	14	13	14	15	111
Cond	.74±.45	.52±.50	.70±.47	.68±.48	.59±.50	.72±.45	.64±.49	.64±.48
	27	50	30	30	29	29	44	239

Session, Type and their interaction are presented for all groups in Table 3.148. Data for Session and SessionXType are presented merely so that one may see how the overall effect of Session and its interaction with Type evolved within groups (the overall effect was due to the fact that all groups moved in the same direction: Performance on Foils improved over sessions). Generally, all groups improved on Foils from Stage 6 to Stage 8, but it was Group 4 that changed the most. In fact, relative to the magnitude of changes on Literals, one could argue that Groups 1 and 2 showed no change on Foils.

Table 3.148: The means, standard deviations and cell counts for Session, Type, SessionXType, in Stages 6 & 8, Part 2, for Groups 1, 2 and 4.

Group 1

	Ses 1	Ses 2	Type
Liter	.81±.39	.84±.37	.82±.38
	186	176	362
Foil	.56±.50	.64±.48	.60±.49
	167	173	340
Sess	.69±.46	.74±.44	.72±.45
	353	349	702

Group 2

	Ses 1	Ses 2	Type
Liter	.85±.36	.79±.41	.82±.39
	142	142	284
Foil	.56±.50	.64±.48	.60±.49
	138	133	271
Sess	.70±.46	.72±.45	.71±.45
	280	275	555

Group 4

	Ses 1	Ses 2	Type
Liter	.86±.35	.91±.29	.88±.32
	63	65	128
Foil	.26±.44	.46±.50	.36±.48
	57	54	111
Sess	.58±.50	.71±.46	.64±.48
	120	119	139

The similarity of Groups 1 and 2 to the overall performance is shown in Table 3.149, using a simple correlation across the seven conditions, separated into Literals and Foils. Note that Groups 1, 2 and Overall had high correlations (Literals with Literals; Foils with Foils;), while Group 4 was distinctly different (particularly, Group 4, Foils is negatively correlated with Overall) . It is not surprising that Groups 1 and 2 were most related to the Overall ratings since they comprise the bulk of the subjects. However, as in Stage 4, the fact that Group 4 was not the same as Groups 1 and 2 (r's=0.46, 0.15, 0.02 & -0.05) is informative.

In general, Groups 1 and 2 were the same (r's=0.75 & 0.82). In contrast to Stage 4 (see Table 3.131), both Groups 1 and 4 have substantial negative correlations between responses on Literals and Foils; the simplest interpretation is that there is something of a response bias in these two groups (however, this does not explain differential responding within conditions; it is as if there is a response bias within some conditions).

Table 3.149: Correlation over the pattern of responding on Condition.

		Group 1		Group 2		Group 4		Overall	
		Literal	Foil	Literal	Foil	Literal	Foil	Literal	Foil
1	Literal	1.000	-.420	.753	-.248	.459	-.641	.967**	-.318
	Foil		1.000	-.221	.820*	-.907*	.146	-.499	.938**
2	Literal			1.000	.0133	.016	-.043	.709	.022
	Foil				1.000	-.826	.050	-.320	.936**
4	Literal					1.000	-.448	.563	.208
	Foil						1.000	-.637	-.400
O	Literal							1.000	.109
	Foil								1.000

Note: * significant at $p<.05$
** significant at $p<.01$

The presentations of Location, Size and Tonality are combined for all three groups in Table 3.150.

Table 3.150: The means, standard deviations and cell counts for Location, Size and Tonality for Groups 1, 2 and 4 in Stage 6 and 8, Part 2.

Location of Change

	Literal	Note 1	Note 2	Note 3	Note 4
Group 1	.82±.38	.40±.49	.88±.32	.65±.48	.65±.48
	362	111	52	94	83
Group 2	.82±.39	.33±.47	.82±.39	.69±.47	.74±.44
	284	92	39	74	66
Group 4	.88±.32	.23±.43	.37±.50	.50±.51	.34±.48
	128	30	19	30	32

Size of Change

	Literal	Size -2	Size -1	Size 1	Size 2
Group 1	.82±.38	.69±.47	.71±.46	.58±.49	.39±.49
	362	58	75	161	46
Group 2	.82±.39	.58±.50	.71±.41	.66±.48	.20±.41
	284	40	68	128	35
Group 4	.88±.32	.52±.51	.38±.50	.27±.45	.40±.51
	128	21	26	49	15

Tonality of Change

	Literal	Ton 1	Ton 2	Ton 3	Ton 5	Ton 7
Group 1	.82±.38	.49±.50	.66±.47	.62±.49	.52±.51	.53±.51
	362	81	192	29	29	17
Group 2	.82±.39	.44±.50	.71±.46	.52±.51	.52±.51	.47±.52
	284	63	143	29	21	15
Group 4	.88±.32	.32±.48	.30±.46	.33±.49	.83±.41	.63±.52
	128	22	50	15	6	8

The analysis within groups are presented in Table 3.151, 3.152 and 3.153. In these analyses, no group showed a significant effect of Session though it was marginal for Group 4. All groups showed a dramatic difference between Foils and Literals and the effect was significant for all groups. From the means and the associated ΔR^2 and R^2_{part}, it is clear that the effect was strongest in Group 4 and weakest in Group 2. The effect of condition is the opposite of that of Type. Group 2 showed the strongest effect and Group

4 showed the weakest effect: Condition was unambiguously significant in Group 2, ambiguously significant in Group 1 (the hierarchical analysis and the part analysis differ) and unambiguously non-significant in Group 4. No Group showed an interaction between Session and Condition or Session and Type. Only Group 1 demonstrated an interaction between Condition and Type although the same was ambiguous in Group 2 since the part correlation was not significant (though the part correlation was of the same magnitude as in Group 1). In Group 4, the Condition by Type effect was not significant, though this is likely due to the number of subjects since the part correlation was actually larger than in the other groups. Session by Condition by Type is not significant in any group. For Location was significant in all groups, while Size in Tonality were not (however, in the hierarchical analysis, Size and Tonality were entered last; this is not a consideration for the part analyses).

Table 3.151: The complete regression analysis of Session, Condition, Type, Location, Size, Tonality and their interactions, in Stages 6 & 8, Part 2 for Group 1, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(^{\Delta R2})$	R^2_{part}	$F(R^2_{\text{part}})$	$p(R^2_{\text{part}})$
Subjects	0.081	22					
Melody Pos.	0.013	2	4.930	0.008			
Sequence Pos.	0.011	2	4.028	0.018			
Session	0.003	1	1.945	0.164	0.000	0.018	0.895
Condition	0.031	6	3.947	0.001	0.009	1.270	0.269
Type	0.059	1	48.879	0.000	0.008	7.269	0.007
Sess X Cond	0.002	6	0.335	0.919	0.004	0.548	0.771
Sess X Type	0.001	1	1.113	0.292	0.000	0.003	0.960
Cond X Type	0.045	6	6.433	0.000	0.018	2.592	0.017
SessXCondXType	0.001	6	0.099	0.997	0.001	0.213	0.973
Location	0.016	3	4.604	0.003	0.012	3.546	0.014
Size	0.005	3	1.539	0.202	0.009	2.626	0.050
Tonality	0.008	4	1.758	0.136	0.008	1.758	0.136
Pooled Residual	0.725	638					
Total	1.000	701					

Notes: The effects are listed in the order in which they entered the equation.

Table 3.152: The complete regression analysis of Session, Condition, Type, Location, Size, Tonality and their interactions, in Stages 6 & 8, Part 2 for Group 2, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model							
Effect	ΔR^2	df	$F(\Delta R^2)$	$p(^{\Delta R2})$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.043	17					
Melody Pos.	0.027	2	7.848	0.000			
Sequence Pos.	0.017	2	4.971	0.007			
Session	0.000	1	0.885	0.665	0.000	0.343	0.559
Condition	0.046	6	4.609	0.000	0.022	2.548	0.019
Type	0.061	1	39.955	0.000	0.006	4.417	0.036
Sess X Cond	0.004	6	0.482	0.822	0.004	0.411	0.872
Sess X Type	0.004	1	2.419	0.121	0.000	0.014	0.907
Cond X Type	0.057	6	6.519	0.000	0.017	1.988	0.066
SessXCondXType	0.003	6	0.338	0.917	0.004	0.425	0.862
Location	0.013	3	2.973	0.031	0.020	4.588	0.004
Size	0.010	3	2.229	0.084	0.005	1.058	0.367
Tonality	0.008	4	1.391	0.236	0.008	1.391	0.236
Pooled Residual	0.707	496					
Total	1.000	554					

Notes: The effects are listed in the order in which they entered the equation.

Table 3.153 The complete regression analysis of Session, Condition, Type, Location, Size, Tonality and their interactions, in Stages 6 & 8, Part 2 for Group 4, providing the change in R^2 (ΔR^2) and the analysis of the part correlations, R^2_{part}.

Hierarchical Regression Analysis: Regression Model

Effect	ΔR^2	df	$F(\Delta R^2)$	$p(\Delta R^2)$	R^2_{part}	$F(R^2_{part})$	$p(R^2_{part})$
Subjects	0.038	7					
Melody Pos.	0.012	2	1.494	0.227			
Sequence Pos.	0.004	2	0.461	0.631			
Session	0.016	1	3.811	0.052	0.001	0.235	0.629
Condition	0.024	6	0.968	0.448	0.014	0.908	0.490
Type	0.297	1	106.682	0.000	0.031	12.185	0.001
Sess X Cond	0.011	6	0.643	0.696	0.017	1.113	0.356
Sess X Type	0.006	1	2.220	0.138	0.001	0.531	0.467
Cond X Type	0.025	6	1.529	0.177	0.021	1.385	0.223
SessXCondXType	0.017	6	1.020	0.414	0.023	1.466	0.191
Location	0.030	3	3.727	0.012	0.027	3.455	0.018
Size	0.012	3	1.583	0.195	0.013	1.718	0.165
Tonality	0.017	4	1.685	0.155	0.017	1.685	0.155
Pooled Residual	0.491	190					
Total	1.000	238					

Notes: The effects are listed in the order in which they entered the equation.

Lastly, it should be mentioned that an overall analysis the examined treated group as a variable was also done. However, as in Stage 4, Group was not significant (ΔR^2=0.00) although there were interactions of Group with other terms (Condition, Type and Location) . The previous analysis within each group shows the differences, such as they are, and that analysis is equivalent to a simple effects analysis within levels of Group (which is necessary in the event that the interactions are significant).

For Location, all groups showed a severe detriment for changes on the first note. However, Groups 1 and 2 showed best performance on the second note, while Group 4 showed best performance for changes on the third note. In Groups 1 and 2, the differences between the second, third and fourth notes were not dramatic, whereas in Group 4 the pattern is different.

For Size, Groups 1 and 2 demonstrated higher performance for changes of ±1 semitones than for changes of ±2 semitones. The result is only marginally significant (see Tables 3.150, 3.151, 3.152 and 3.153), though it was consistent enough to warrant consideration. This pattern over Size was also seen in Group 4 for the familiar melody "Three Blind Mice".

As has been stated before, Size and Tonality are confounded because both are associated with frequency or pitch. In a tonal piece (such as this melody) size changes of ±1 semitones are more associated with changes of tonality than size changes of ±2 semitones. This concept is presented in Table 3.154 for the key of C-Major. Note that changing a diatonic note by one semitone will result in a non-diatonic note on five occasions but changing a diatonic note by two semitones will result in a non-diatonic note on only two occasions. When these changes are weighted by the probabilistic occurrence of notes within a tonal piece (these probabilites are taken from Krumhansl [1990] and as such apply to the corpus of western tonal music, and not identically to the melodies used herein), then one can get a rough picture of the confounding of changes in size and tonality.

When one weighs the diatonic notes by the Krumhansl hierarchy (which is correlated with both number and duration of notes in a melody; cf., Chapter 2 this work for the application of the algorithm to this melody), changes that decrease the size are not identical with changes that increase the size in terms of change in relative tonality.

Table 3.154: The association between changes in size and changes from in-key to out-of-key. The weighted value is the relative difficulty of each size change when weighted to consider tonality. That is, a size change of +1 would produce a non-diatonic note more often than a size change of +2, and hence would result in higher performance (more detectability).

Change		Note	Change	
-2	-1		+1	+2
non-diatonic	—	C 6.35	non-diatonic	—
		$C^{\#}$ 2.23		
—	non-diatonic	D 3.48	non-diatonic	—
		$D^{\#}$ 2.33		
—	non-diatonic	E 4.38	—	non-diatonic
non-diatonic	—	F 4.09	non-diatonic	—
		$F^{\#}$ 2.52		
—	non-diatonic	G 5.19	non-diatonic	—
—	non-diatonic	A 3.66	non-diatonic	—
		$A^{\#}$ 2.29		
—	non-diatonic	B 2.88	—	non-diatonic
2 10.44	5 19.59	Number weighted value	5 22.77	2 6.36

Notes: Only change from the diatonic (key) notes were considered since the melodies used herein were limited in this manner.

Changes of +1 semitone will more often result in a non-diatonic tone (hence a change in tonality of the test sequence) than changes of -1 semitone. That is, if diatonicism matters, then changes of +2 should be the least obvious (which they are in Groups 1 and 2), changes of -2 should be the next least obvious (which they are) and changes of ±1 should be the most obvious, which they are). Think of weight as obviousness (to one who uses tonality).

Note that, as shown in Table 3.155, there was a positive correlation between Size (magnitude only) and performance in Group 4 in Stages 6 and 8. However, Size (magnitude only) was negatively correlated with performance in Group 1 in Stage 4 and in Stages 6 and 8. Size (magnitude only) was also negatively correlated with performance in Groups 2 and 4, in Stages 6 and 8. These "strange" observations (larger sizes leads to lower performance) can be reconciled by examination of the Weight correlations (Weight can be taken as an indication of obviousness). Weight was positively correlated with performance in Groups 1 and 2 in Stages 6 and 8, and Weight was positively correlated with performance in Group 4 in Stage 4. It is as if, the subjects in Group 4 could use information about tonality as a aid to memory only for a familiar melody. Hence, Group 4 with a familiar melody acted like Groups 1 and 2 with an unfamiliar melody. The correlations in Groups 1 and 2 in Stage 4 should be interpreted cautiously because these groups did not show much range in performance in that Stage (ranges = 0.20 and 0.18 versus 0.81 for Group 4). Groups 1 and 2 did show considerable range in Stages 6 and 8 (ranges = 0.32 and 0.51 versus 0.25 for Group 4). The patterns for the groups on Size seem to be explained by the confounding influence of tonality.

Hence, the difference between Stage 4 and Stages 6 and 8, as it relates to groups, could be due to the difference in familiarity. In Stage 4, the melody was familiar to everyone. For those with musical training, tonality did not matter since sequences could be compared directly with a memory representation. For those with more limited training, tonality could be used as an aid. Conversely, for the unfamiliar melody, those with musical training were forced to (could choose to?) relying on tonality as a cue, while those with less training were forced to fall back onto more basic structures.

Table 3.155: Correlations between the Size parameter (-2, -1, 1, 2), the Size weighted by tonality parameter (10.44, 19.59, 22.77, 6.59) and performance as a function of Size in Stage 4 and in Stages 6 and 8.

				Stage 4 Group			Stages 6 and 8 Group		
		Size	Weight	1	2	3	1	2	3
Size		1.000	-0.118	0.819	0.159	-0.832	-0.415	-0.736	0.761
Weight			1.000	-0.690	0.033	0.712	0.518	0.839	-0.702
Stage 4 Group	1			1.000	0.694	-0.993	0.176	-0.218	0.950
	2				1.000	-0.675	-0.218	0.551	0.654
	3					1.000	-0.157	0.243	-0.958
Stages 6 & 8 Group	1						1.000	0.897	0.244
	2							1.000	-0.201
	3								1.000

Notes: Weight is the size weighted by tonality (a measure of obviousness for those who use tonality).

For Tonality, Groups 1 and 2 show the predicted pattern (more strongly in Group 2), with the additional consideration discussed previously (Tonality 1 will not produce the highest performance due to the number and role of triad tones within a four-note sequence: See Boundary Efficacy Within Groups: Stage 4, Part 2). However, Tonality was not significant within either group (see Tables 3.151 through 3.153), though one contrast was significant and two contrasts were marginal. For Group 4 (Tables 3.151 and 3.154), the pattern of responding is very different again (not significant), showing little differentiation of levels of tonality.

Discussion

There were four main purposes contained within the design of this experiment, and in general, the experiment was successful. The first goal was the determination of empirical boundary profiles. With respect to this goal, the experiment was an unqualified success. Individuals having a wide variety of musical backgrounds were able to parse both a familiar nursery rhyme and an unfamiliar, but tonal, melody using a direct on-line procedure. The on-line procedure was developed for this assessment. Subjects' responses were highly consistent across repetitions and showed the expected dependency on familiarity. That is, subjects' responses were less consistent with an unfamiliar tune and subjects's responses became more consistent over repetitions. Furthermore, different subjects demonstrated a great deal of similarity in their responding. That is, different subjects parsed the melody in similar, though not identical, manners. In this regard, the work echoes that of Newtson and coworkers in the area of social psychology (Newtson, 1973, 1976; Newtson & Engquist, 1976; Newtson, Engquist & Bois, 1977; Newtson, Hairfield, Bloomingdale & Cutino, 1987).

Despite the high consistency and similarity, it might be fruitful for future studies to examine the boundary profiles in greater detail. Such detailed analysis might reveal subtle systematic differences (a basis for grouping on boundary profiles) and such analysis might be able to determine whether or not the minor departures from perfect reliability represent developmental change or noise. Changes in the perception of the melodies might evolve with exposure to the melodies. Following Newtson and coworkers, these evolutions should manifest as systematic changes in the number of boundaries: boundaries at each repetition should build on the boundaries of previous repetitions (i.e., some locations should be stable, while others are added). Unfortunately, based on the work of Newtson and coworkers, it is not possible to say whether this trend should be an increase in the number of boundaries (successively smaller units leading to more extracted information) or a decrease in the number of boundaries (successively larger units leading to a more global appreciation of the stimulus). The implication herein was that the tonality groups (Groups 1 and 2) seemed to move from larger to smaller units while the proximity group seemed to move from smaller to larger units, both converging on the same unit size. This was particularly true for the second unfamiliar melody.

With respect to the details of parsing, subjects tended to use groupings of six to nine notes, which is within the range of the Miller's (1956, cited in Anderson, 1990) short-term memory limit (7±2), but there was a wide range. Units consisting of a single note event and units consisting of 10 or more note events were not common, but they were not entirely absent either.

By inspection of Figures 3.16 and 3.23, it appears that subjects preferred to place boundaries at relatively longer notes that ended a smooth contour. This can be seen at Note Events 3, 5, 10 14 and 23 of the first melody ("Three Blind Mice": Figure 3.16) and at Note Events 9, 16 and 25 of the second melody (the extract from "Softly Now the Light of Day"; Figure 3.23). In the first melody, the longer notes that were associated with boundaries were more important within the tonal hierarchy (mediants, or dominants). In the second melody, the lengthened notes that were associated with empirical boundaries were all either the mediant or the tonic. There were no other instances of lengthened notes. It is not likely a coincidence that these longer notes correspond to tonal centres within the tonality of the melody. Krumhansl (1990) has provided a substantial amount of evidence indicating that tonally important notes tend be those that are sounded more often and for longer durations. It can be seen, again by inspection, that not any long, tonally-important note will do. Other equally long notes that also corresponded to tonally important notes were not associated with a boundary (e.g., Notes 16, 21, 31 and 36 of "Three Blind Mice" were the tonic, but they did not have empirical boundaries). It is difficult to determine what is the determining factor. It could be the parsing of lyrics, but then, one could parse the lyrics in a different manner. Casual inspection seems to indicate that only those lengthened notes that end of a smooth contour may lead to a boundary. Unfortunately, there are exceptions to this as well. Finally, it must be mentioned that there is a weak boundary in the first melody that is associated with no change at all (Note 34). Lacking any explanation to the contrary, I personally, would consider this to be an effect of parsing of the lyrics.

The second main purpose of the experiment concerned the empirical verification of Lerdahl and Jackendoff's (1983) Group Preference Rules (GPR). In this, the experiment can be described as successful, with some qualifications. Of the rules tested, only one was found to have strong empirical verification (Rule 2b: boundaries due to changes in the lengths of the notes) which coincides with the previous observations. This rule was associated with boundaries with a correlation of about $r \approx 0.70$ ($r^2 \approx 0.50$). The remaining rules (Rules 3a and 3d) were correlated at a much lower level: $r \approx 0.30$ ($r^2 \approx 0.09$). Rule 1 was generally supported (avoid units with only a few events). Lerdahl and Jackendoff do not provide an "ideal" size, but units of size 4 is implied as the norm; the average unit sizes were six note events or more which is not inconsistent. This size is also not inconsistent with the notion of a motive. However, it must be reiterated that there were some units having only one note event. These results support those of Deliege (1987). (Peretz [1989] did not examine Rule 2b, so comparisons cannot be made directly.)

In general, it was found that the rules generated many false positives (predicted boundaries where none were placed) and few misses (failure to indicate boundaries where one had been placed). Although the fit is not perfect, it is promising for future work. The first implication is that a new method of matching the rules to the empirical boundaries will have to be found. That is, the assessment of empirical boundaries seems to be reliable both within and between subjects. However, the match between the rules (the theoretical predictions) and the empirical boundaries is not as high as one would like. It is possible that this match can be improved with a new analysis without changing the theoretical foundation of the rules. That is, before inventing new rules to explain discrepancies (or discarding the old ones), the match between the old rules and the empirical data should be optimized. The problem is that empirical boundary placement (i.e., by the subjects) is done on a note-by-note basis. If a subject uses some internal representation of rules to parse the melody, then those rules are likely applied on an "either-or-and" basis at each note of the melody (see Figure 3.29).

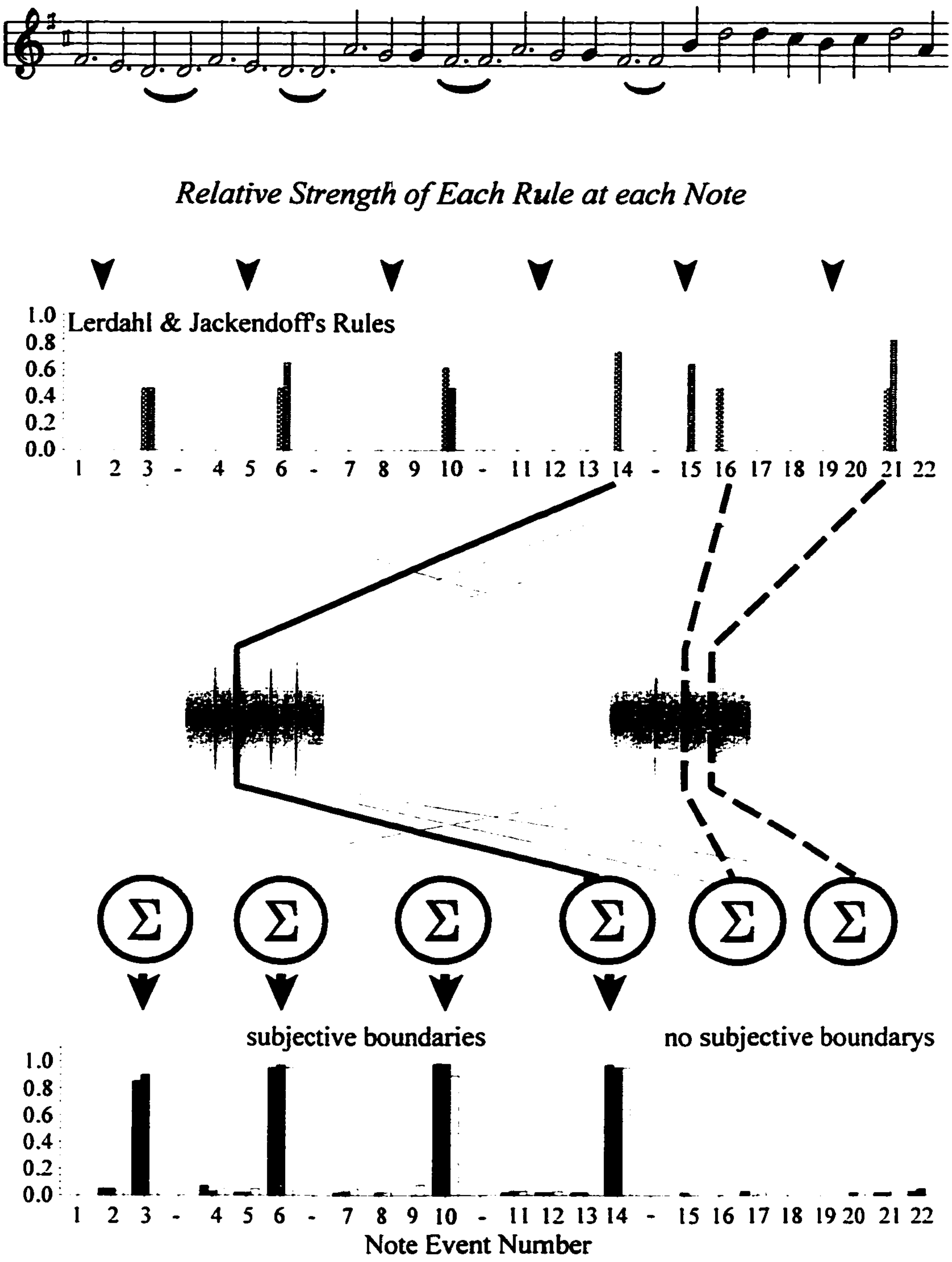


Figure 3.29: The formation of boundaries likely represents the joint action of many rules at each point in the melody. It is not difficult to explain a particular boundary as the action of a single strong rule (e.g., Note Event 14) or as the sum of several weaker rules (e.g., Note Event 3). It is more difficult to explain the lack of a boundary at very similar points (e.g., Note Events 15, 16, and 21).

That is, an empirical boundary may be "caused" by either Rule X or by Rule Y or by the combination of Rules X and Y. As shown in Figure 3.29 (data copied from previous analyses), boundaries at Note Events 3, 6 and 10 seem to reflect the combination of Rules 2b and 3a, while the boundary at Note Event 14 seems to reflect Rule 2b in isolation. Note that in the second melody ("Softly Now the Light of Day"), the boundary at Note Event 16, which is associated with two rules seems to be a little stronger than those at Note Events 9 and 25 (particularly on the first presentation). Lerdahl and Jackendoff (1983) have called this confluence.

More importantly, the analysis of the rules needs to be refined to account for the lack of an empirical boundary when similar situation have demonstrated a boundary. For example, why is there no boundary at Note Events 15, 16 and 21 (or Note Events 31 and 44 of the same melody) even though previous similar situations (equivalent rule strengths) had produced boundaries?

It could be that combinations of rules act to preclude a boundary. Lerdahl and Jackendoff (1983) seem to evoke Rule 1 (avoid small units) to resolve conflicts between rules. That is, when two different rules indicate boundaries on adjacent notes, only one boundary should be placed: Rule 1 states that a units should not contain a single event. To avoid a unit with a single event, other (higher level) rules would arbitrate the choice between locations. In contrast to this, its possible that when there are two possibilities in conflict, neither is used. This might explain the lack of boundaries at near Note Events 23-24 and 35-36 in the first melody.

Some of this refinement will necessarily evoke the higher order rules of Lerdahl and Jackendoff (1983). Some of this refinement will require better analyses to match rules in isolation and in combination to the empirically observed data. The refinement of the analysis of these rules and the extension of the analysis to higher level rules of Lerdahl and Jackendoff is the topic of the next chapter.

The second implication is that changes to the rules will have to be made. because misses represent a fundamentally different problem. A false positive can be viewed as a problem of quantity or scaling, but a miss (particularly, a complete miss) implies a qualitatively different process is involved. The miss at Note 34 in the first melody ("Three Blind Mice", Figure 3.16) is not easily explained from the rules or from an examination of the music of the melody. One could invoke Rule 1 with a minor modification to the number of notes based on the previous observation of units tending to be around eight notes long: It is as if subjects (only 30% of subjects) simply needed to place a boundary somewhere in this general area and this is where it fell. The boundary is about halfway between the end of the melody and the previous boundary: it is 11 notes after the previous boundary (subjects may have needed to close a unit or lose that part of the melody). This boundary can be explained by reference to the parsing of the lyrics of the melody. It is interesting that the proportion of subjects placing a boundary at this point decreases with repetitions. The implication is that the placement of theoretical boundaries needs to be adjusted somewhat. The probable cause is that within the scheme of Lerdahl and Jackendoff (1983), theoretical boundaries are placed in the centre of the most recent four or five note span: This requires a lot of processing: hold four notes in memory (while gathering more notes as they sound), decide where to place the boundary, assign the first two notes of the span to the first unit and store that unit, and retain the second two notes of the span as a part of the current (still unsaved) unit. Subjects, however, likely place boundaries by comparing the current note to the previous notes. That is if the current note is different from the previous (on some dimension), place a boundary. Rule 2b (Attack-point), which is the most successful, and Rule 3a (Register Change) could work by comparing the current note to previous notes. Rule 3d (Length Changes) which is less successful than Rule 2b could not work by comparing the current note to previous (i.e., the boundary is placed, retrospectively, two notes before the current note).The differences between these two approaches, while theoretically large, will not produce different parsing except in a few circumstances.

It is critical that misses be examined in more detail because misses cannot be dealt with within the framework of Lerdahl and Jackendoff's (1983) theory. That is, if an empirical boundary is not implied by one of the low level rules cited (used) herein, then there cannot be a boundary at that point (all low level rules were either controlled in the stimulus or explicitly assessed). Given the strict definition of a hierarchy that Lerdahl and Jackendoff have advocated, all high level boundaries must be a boundary at every lower level. Before it can be said that the theory fails to explain a particular empirical boundary, the analysis must be adjusted so that misses (a failure of the theory to predict a boundary) are minimized. That is, the analysis must be adjusted so that the fit between the theory and the empirical data generates no misses (implicitly, this will increase the false alarm rate). This, too, is a part of the analysis of the next chapter.

The general conclusions with respect to the theoretical parsing can be subdivided into two main concerns. The first is that a new routine for matching theory with empirical data needs to be developed so that the fit between the rules and the data can be optimized (minimizing the number of misses): This will not be an easy task. The second is that if the fit cannot be improved with optimization, then the rules may have to be revised or new rules may have to be added (minimizing the number of misses).

A third main objective for this experiment was the experimental verification that the empirically determined boundaries were more than just an artifact of the experiment. That is, if boundaries did represent a parsing of the melody into units, then retrieval from memory should demonstrate predictable effects. In particular, the retrieval of two units should demand more resources (attention, energy, time) than the retrieval of a single unit. Hence, if a subject is asked to compare a test sequence to the melody, test sequences that correspond to more than one unit should require more resources. This, in turn, should manifest as more errors in such comparisons.

Generally, such results were observed in both melodies although the results from Stage 4, Part 2 are generally easier to interpret. Results with test sequences were consistent with predictions concerning the location of change in the test sequence (all groups are sensitive to changes on the first note), the size of the change in the test sequence (the lack of an effect) and the tonality of the change in the test sequence (change from in key to out of key are more obvious). The results are generally consistent with those of Dowling, (1973), Peretz (1989) and Tan et al., (1981), but direct comparisons should be made judiciously (particularly with Dowling and Tan et al.,) because previous studies used fewer test sequences, did not examine Foils (at least, not in detail) and generally shorter melodies (or stimuli).

The observation that the results did not meet predictions identically can be explained by the fact that test sequence performance must reflect many factors such as. the complexity of the test sequence, the complexity of the melody (and, therefore, its accuracy of retention), the abilities of the subject to process melodies, the time between test sequence presentation within the melody and the test sequence presentation as a test (assessed in part, by the covariates in the analysis) and the ability of the subject to use various features of the test sequence and/or melody for comparison. Note that the differences in performance between Stage 4 and Stages 6 and 8 must be due in part to some of these factors. The melody of Stage 4 was highly familiar to all (hence one might be willing to assume that the memory representation of the melody was fairly accurate) , while the melody of Stages 6 and 8 was highly unfamiliar to all. In addition, the tonal strength of the melody of Stage 4 was higher than that of the melody of Stages 6 and 8 (see Chapter 2, this work). Another way to view these results is to consider how many different factors one might think would affect performance in this task (e.g., relative boundary location, sequence complexity, melody complexity, accuracy of memory representation of melody [hence, musical ability and melody familiarity], type of change [relative importance of change within melody and test sequence], relative tempo). If there are 10 factors, then an initial estimate would assign 10% of the variance to each factor. Hence, each factor would be expected to have a partial correlation less than about 0.3. Although this would be an interesting avenue to pursue, because this task was secondary to the overall goals, detailed analysis were not pursued[37].

A fourth main purpose of this experiment concerned the relationship between the boundary profiles and tonality (more accurately, tonality groups, based on probe-tone

---

[37] One might think that the analysis were detailed, but there are many more explorations that could be made with this data alone. For example, signal detection analysis could be applied to the analysis of condition, but such an analysis would simply replicate the analysis used.

profiles). Basically, it is known that tonality is fundamental to music and several studies have shown that an internalized representation of tonality is associated with training (Frankland & Cohen, 1990; Krumhansl, 1990). Hence, it was considered possible that tonality and boundary formation might be related. Three large groups (the focus of analyses) and several smaller groups of subjects were created and boundary profiles within each of these groups were examined. Of the three large groups, two were associated (though not significantly) with higher levels of training (Groups 1 and 2, labelled as Triadic) and one was associated with lower levels of training (Group 4, labelled as Proximity). In the empirical boundary profiles for each group, only minor differences could be seen. All groups tended to use Rule 2b (length) most consistently. There are three possible explanations for this. Firstly, if changes in duration (Rule 2b) is the dominant parsing mechanism, then the role of tonality is diminished accordingly. Secondly, it is likely that there are a number of mechanisms for the placement of boundaries and as such tonality is only one and therefore, one can only expect subtle effects of tonality in a task that does not (cannot) control other aspects of boundary formation. Thirdly, and most likely, the major mechanism for boundary formation is changes in duration (i.e., longer notes are points of rest) and tonally important notes are generally associated with notes of longer duration (cf., Lerdahl & Jackendoff, 1983; Krumhansl, 1990). That is, relatively longer notes tend to be tonally important notes (i.e., the notion of a cadence involves both; see Chapter 1 this work). Hence the two mechanisms are confounded and it might be difficult to separate the two since melodies that did not associate longer notes with tonal centres might not sound like tonal music.

With respect to test sequences, Group 4 (Proximity) tended to be different from Groups 1 and 2 (Triadic). This was most obvious in the analysis of Condition, Size and Tonality. In the analysis of Location, the groups did not differ. With respect to Type, Group 4 exhibited the largest disparity between Literals and Foils. The interesting observations within this section were that with respect to Size and Tonality (these two measures are confounded), Group 4, with the more familiar more tonal melody, seemed to be similar to Groups 1 and 2 with the unfamiliar (less tonal) melody. This has some interesting implication that could not be explored at this time (e.g., detailed individual analysis of the evolution of the probe-tone profiles and test performance that might indicate that some subjects in Group 4 were moving towards Group 1 or 2).

Generally, the experiment was successful. Subjects produced boundary profiles that were consistent and similar. These boundary profiles could be related to the rules of Lerdahl and Jackendoff in meaningful way. The efficacy tasks indicated the boundary formation did have some impact on subsequent recall. Finally, there were subtle effects of tonality group on the previous boundary location and boundary efficacy tasks.

For the future, there are main areas of refinement. Firstly, a better algorithm for fitting the rules to the data is needed. Before it can be said that Rule 2b is the most important, it is important that all the rules be placed on an equal footing. For all rules, false alarms should be common because subjects will be parsing at some level of the hierarchy that is above the level of Rules 2 and 3 (cf. Chapter 2, this work). Conversely, for all rules working in concert, misses should not happen. That is, some rule should explain every empirical boundary, even if no one rule explains all empirical boundaries. The next chapter attempts to develop a new method of analysis, based on the data of this experiment.

The second major point concerns the demand characteristics of the experiment as a whole. Essentially, after the practice trials, subject would have known that there would be a memory component to the task. Hence, boundary formation could have been a consequence of the desire to remember the melody. Boundary formation might be an artifact of the task. Although the high consistencies and similarities argue against this notion, it must be tested. Hence, Chapter 5 of this work is a description of Experiment 2 which was, basically, a repeat of Experiment 1, removing the requirement for memorization.

CHAPTER 4

Reanalysis of the Link Between the Empirical Data and the Rules of Lerdahl and Jackendoff

In this work, the predictions of Lerdahl and Jackendoff's (1983) low level Group Preference 2b (Attack Point), 3a (Register) and 3d (Length) were compared with the placement of group boundaries by subjects[38]. In the previous test (Chapter 3 of this work) Rule 2b was generally the most important while Rules 3a and 3d were relatively unimportant. Although the results were promising, no one rule, nor the simple linear combination of all rules, yielded accurate prediction. Broadly, the goal of this chapter is the optimization of the fit between the theory and the data.

It was not expected that any one rule could accurately predict the empirical data. Empirical boundaries reflect the combined action of all the low level rules and the influence of other higher level rules defined by Lerdahl and Jackendoff (1983). As well, there may be rules that Lerdahl and Jackendoff (1983) missed. For these reasons, each low level rule was expected to generate false alarms (boundaries predicted by the rules, but not indicated by subject: the definition will be refined) and misses (boundaries indicated by subjects but not predicted by the rules: the definition will be refined). However, the combination of all rules should not generate any misses. That is, theoretically, overall there should be an imbalance between false alarms and misses. In this chapter, one major goal is the development of a new analysis technique to address that imbalance between false alarms and misses. A second goal is the extension of the analysis to the next level of Lerdahl and Jackendoff. The inclusion of higher order rules will affect the overall error rate. Both these goals are directed at optimizing the fit between the theory and the data. It is important to realize that these two goals are not completely separable. That is, it is difficult to extend the analysis to new rules without some method of analysis that addresses the imbalance between false alarms and misses. Hence, the initial emphasis of this section is on the issue of the asymmetry of false alarms and misses. The reason for this should become clear in the subsequent discussions.

---

[38] In principle, the techniques discussed herein can be applied to any rules.

By design (Lerdahl & Jackendoff, 1983; cf., Chapter 2 of this work), Rules 2 and 3 represent the lowest level of a strict hierarchy. As such, any boundary within a melody must correspond to the application of some low level rule (or rules), but not all applications of these low level rules would necessarily result in boundaries at higher levels of the hierarchy. Hence, each low level rule could generate both false alarms and misses. This is not a problem for the theory. Critically however, for the theory to be correct, all the rules in conjunction should not produce *any* misses. All the rules in conjunction could generate many false alarms. The asymmetry between false alarms and misses arises from the fact that subjects likely parse the melody at some intermediate level and from the fact that there is more than one low-level rule.

Firstly, the parsings indicated by subjects must correspond to some level that is equal to, or higher than, the level indicated by Rules 2 and 3. The "higher than" is more probable given the results of the previous experiment (see Chapter 3 of this work). That is, all that we know at this point is that the level of the subjects is higher than the level of rules -- we do not know how much higher and we do not know which level of the hierarchy corresponds to the parsings of subjects. Lerdahl and Jackendoff (1983) do not explicit state which level is the most natural for a listener, but the structure of their text and the weight given to various levels implies that listeners do not "hear" (attend, structure, remember) music at the lowest levels of their hierarchy. It is likely that subjects did not naturally indicate their parsing at the lowest level of the hierarchy during the experiment (see Chapter 3 of this work). Newtson observed that subjects asked to parse "naturally" used an intermediate unit size. Subjects could be induced to parse in larger or smaller units (cf., Newtson, 1973; 1976; Newtson & Engquist, 1976; Newtson, Engquist & Bois, 1977; Newtson, Hairfield, et al., 1987). Hence, it is likely that the melody parsings by the subjects in the previous experiment corresponded to some intermediate level[39]. It is best to view the application of the low level rules as the production of a set of

---

[39] Note that subjects were asked to parse at the smallest size that made sense. The point is that we cannot not know which level of the hierarchy this actually applies to the empirical data.

possible interpretations for the melody (i.e., the low level rules define potential locations for a boundary; cf., Chapter 2 of this work, or Lerdahl & Jackendoff, 1983, p. 42). Of the set of all possible boundaries, not all will be heard as a boundary on any single occasion. The mismatch between the level of the rules and the level of the empirical data is the first complication for the analysis: it contributes to the asymmetry of the importance of false alarms and misses.

The second complication is created by the fact that at each note in the melody, any one of several rules may or may not apply. Essentially, the indication of a boundary by at least one of the low level rules is necessary, but not sufficient, for the creation of a boundary at the higher levels. Consider the simplest case: If Lerdahl and Jackendoff (1983) had proposed only one low-level rule, then every empirical boundary would have to correspond to a prediction by that rule. False alarms would be expected but misses would be forbidden, regardless of the level of the parsings of the subjects. It would be easy to determine whether or not the rule "worked": If the rule ever missed, the theory would be invalid. Figure 4.1 presents some examples: Rule "A" is a better fit to the theory than Rule "B", because Rule "A" generates no misses, even though it generates more false alarms. Now consider the slightly more complex situation. If Lerdahl and Jackendoff (1983) had predicted only two low-level rules, then each rule could generate false alarms and each rule could generate misses. However, both rules together could never generate a miss although both together could generate false alarms. In Figure 4.1, the combination of Rules "B" and "C" might be better than Rule "A" because the combination generates no misses and fewer false alarms. Rule "B" is better than Rule "C". It is more difficult to test this situation because there is a third complication.

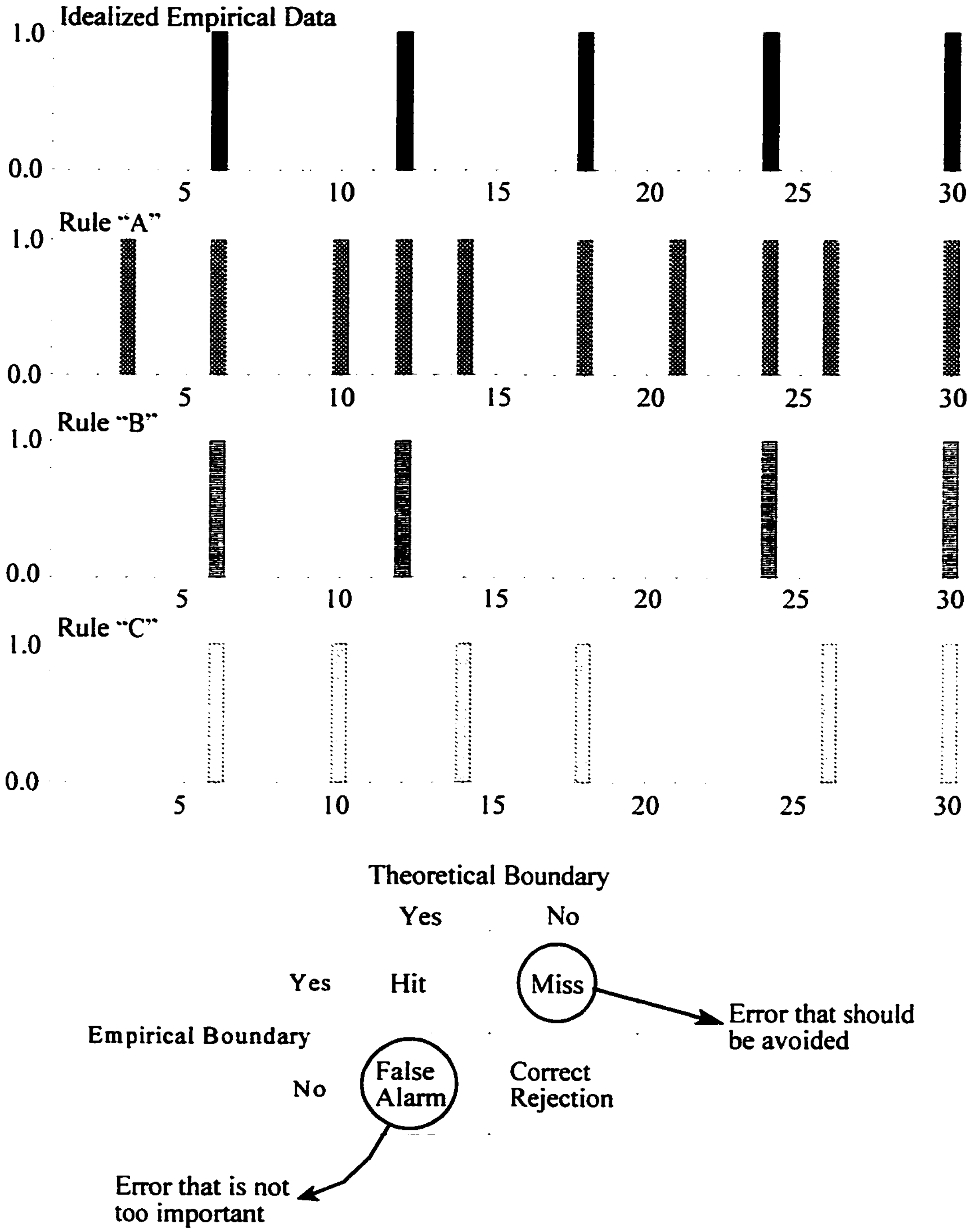


Figure 4.1: Idealized empirically determined boundaries within a "melody" of 25 notes events, a rule producing many false alarms but no misses (Rule A), a rule producing misses but no false alarms (Rule B) and a more "typical" rule producing some of both (Rule C). For this theoretical development, Rule A is less deleterious than Rule B, but the combination of Rules B and C might be better than Rule A.

The third complication is that the strength of the rule at each note of the melody is a continuous variable. As such the rule strength may or may not be sufficient for the creation of a boundary. Consider, once again, the simplest case of only one predicted rule. Every empirical boundary must correspond to a theoretical rule. Hence, fitting the rule to the data is a issue of scaling the rule (linear, or non-linear scaling) so that the false alarms are dropped while no misses are generated. If this cannot be done, then the theory is suspect. This is complicated by the fact that the empirical data is also continuous (i.e, the average boundary profile of many subjects is continuous).

Note that issue of scaling can be considered as a finer analysis of the miss / false alarm dimension (i.e., misses and false alarms can be considered a binary coding of scaling). Ideally one would like to match the strength of the theoretical rule to the strength of the empirical boundaries (i.e., the proportion of subjects indicating a boundary at that point). When the strength of the theoretically rule exceeds the strength of the empirical boundary, one would have a false alarm. When the strength of the empirical boundary exceeds the strength of the rule, one would have a miss[40]. The important point is, both the theoretical rule and the empirical data are continuous variables. This makes it more difficult to analyze misses and false alarms. The first two panels of Figure 4.2 ("Empirical Data" and "No False Alarms or Misses") depict the situation of continuous empirical data and a continuous rule. Three levels of match are provided (high, medium and low). Note that even if there are no misses or false alarms, there is still the problem of the scaling of the rule to match the empirical boundaries. To consider more realistic examples, Figure 4.2 present the same three situations as are in Figure 4.1: Rule "A" is a better match to the data than Rules "B" or "C", because only Rule "A" has no misses. The continuity of the data makes this more difficult to discern. Rule "B" is better than Rule "C": fewer misses.

---

[40] One can set, and argue about, various criteria for the disparity between the rule and the empirical boundary before one actually attaches the label "miss" or "false alarm".

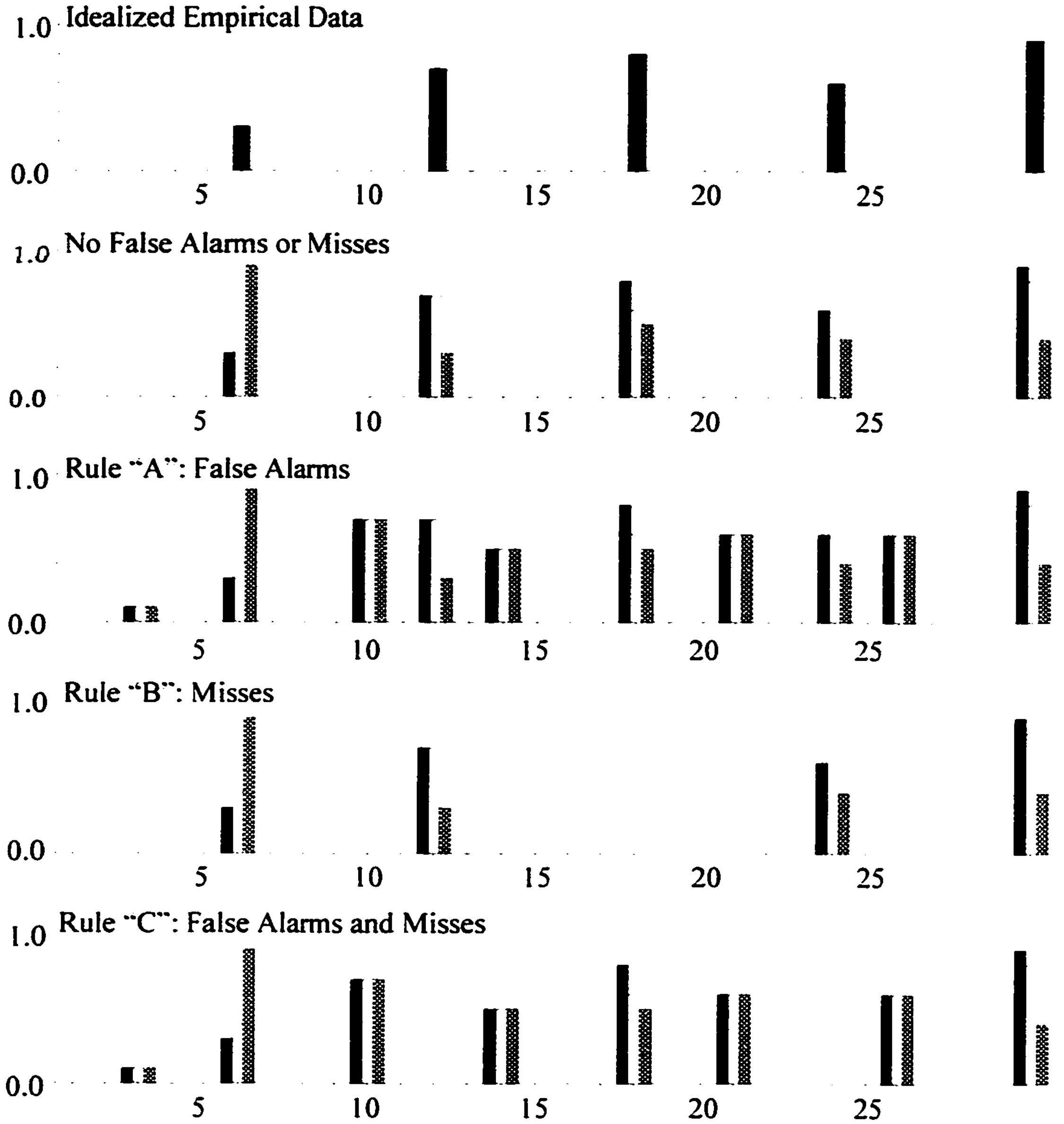


Figure 4.2: Idealized empirically determined boundaries for continuous data, and potential matches when there are, no false alarms or misses, some false alarms (but no misses), some misses (but no false alarms) and both false alarms and misses. Each rule is shown with 3 levels of correspondence. Note that, in isolation, Rule A is better then Rules B or C, but the combination of Rules B and C is better than Rule A.

■ High degree of match Medium degree of match ▨ Low degree of match

Consider, once again, the slightly more complex situation in which there are two theoretical rules. Every empirical boundary must correspond to either of the rules or perhaps to both of the rules. The problem of scaling the rules to fit the data is much more complex, even though the basic problem has remained the same: Both rules must be scaled (simultaneously) so that false alarms are dropped while no misses are generated. How can this be done when each rule is continuous, empirical boundaries are continuous, and false alarms are not equivalent to misses (for the test and verification of the theory)? This situation is depicted also in Figure 4.2. Figure 4.2, the combination of Rules "B" and "C" might be better than Rule "A" because the combination generates no misses and fewer false alarms (as was true in Figure 4.1). However, continuous data, continuous rules and the asymmetric importance of false alarms and misses makes the analysis more ambiguous. Note that the analytic solution must be applicable in the case of n rules, where n can be any number (not just 2). Lerdahl and Jackendoff (1983, pp. 45-46) predict six low level rules, while acknowledging the possibility for more.

In summary, the rules indicate possible boundaries at the lowest level of the hierarchy. The empirical parsings of subjects likely represent boundaries at some intermediate level of the hierarchy. As such, it was expected that the rules, in isolation or in combination, would generate a large number of false alarms (predicting a boundary where none had been placed empirically; see Figures 4.1 and 4.2). These false alarms would not invalidate the theory. Misses (failure to predict a boundary where one had been placed empirically), on the other hand, would be a serious concern for the theory. In fact, one could argue that only misses matter while false alarms are largely irrelevant to the theory. However, in the analysis, simply ignoring false alarms would be too extreme: As a trivial example, if false alarms were to be ignored, then any theory that predicts a boundary at every note of a melody would never generate an error. With due consideration for that caveat, it is obvious that false alarms are much less important than misses. That is, as long as the level corresponding to the parsing of subjects is greater than, or equal to, the level of the theoretical rules (which is must be, in this analysis, because the rules represent the lowest possible level of the hierarchy), there will be some imbalance between the importance of, and the expected rates of, false alarms and misses. Properly, it is only by acknowledging the imbalance between false alarms and misses can the analysis proceed. It is only by treating false alarms differently from misses that one can analyze the data if one does not have knowledge of the actual level of the hierarchy that corresponds to the parsings by subjects. When the theory has progressed to the point where it makes predictions at the same level as the empirical data, false alarms and misses should be weighed equally. Finally, the analysis is further complicated by the fact that both the rules and the empirical data are continuous and not binary.

Before continuing, it must be acknowledged that it would be possible to convert both the empirical data and the rules into dichotomies (in fact, individual subject data for a single presentation of the melody are dichotomous, although the individual data for multiple presentations of the melody are not). However, this approach was not chosen because it requires the setting of a criterion or cutoff, to define when a boundary exists both empirically and theoretically. There is simply no information to guide such criteria. Arbitrarily setting criteria could falsely support one rule over another.

Refining the Analysis

The asymmetry between false alarms and misses and the continuous nature of the data and rules complicate the analysis. These complications that were only partially addressed by the previous analyses[41]. Although useful as an initial exploratory foray, the previous analyses comparing the rules to the empirical data were not optimal because techniques were based in correlation and regression. These techniques base their error analysis on the summed squared deviations: $SS_{error} = \Sigma(X_i - X)^2$. Such an error analysis does not distinguish between false alarms and misses. To properly assess error, it is necessary to find an analysis that does distinguish between the two. Several methods were considered and in the discussion that follows one will see elements borrowed from many

---

[41] Although the previous analysis were not ideal, they did serve as a useful preliminary analysis and branch point for more complex analyses to be developed herein. One reason for the choice of those analyses was their general familiarity in the literature.

different techniques (particularly signal detection theory, Green & Swets, 1966)[42].

In this section, a new analysis of the previous data was developed with two major goals. The first goal was the development of a new computational method that could provide differential weights (importance) to false alarms and misses. This was intended to allow for the optimization of the match between the individual rules and the data. This also provided a fairer basis for comparisons between the rules.

In summary, the goals of this analysis was the development of a new method of analysis was developed that weighted misses and false alarms differently, so that the fit between each rule and the empirical data could be optimized and so that comparisons between rules could be more accurate.

As implied by the previous discussion, there are several difficulties that this new analysis will have to address. Specifically, these difficulties include:

1) The scaling of each rule is somewhat arbitrary. That is, the use of linear scaling may be inadequate or the conversion from stimulus to rule may be wrong (i.e., the slope that relates the note events in the stimulus to the measure of rule strength; cf., Chapter 2, this work).

2) False alarms are not as important as misses. That is, the rules may represent a *lower* level of the hierarchy than the empirical data. The rules *do not* represent a higher level of the hierarchy than the empirical data.

3) Rules may operate alone or in combination, but when multiple rules apply, it cannot be assumed that the application is that of simple addition (i.e., only one rule may be doing all the work even though many rules could be applied).

There are many ways to attack such problems. For example, non-linear conversions from stimulus to rule or from rule to boundary (e.g., exponential or logarithmic functions) can accomplish many of the objectives of unequally weighted false alarms and misses. Be that as it may, for this analysis, the assumptions (implicitly defining/limiting some

---

[42] Explaining why each technique was rejected would consume an unreasonable amount of space. Hopefully, those who are interested will be able see the reasons in the subsequent development.

methodological choices) are:

1) The scaling from stimulus to rule is linear.
2) The conversion from melody structure to rule strength is constant throughout the melody.
3) The conversion from rule strength to actual boundary formation is a constant throughout the melody.
4) All subjects can be represented by the average profile, or equivalently, all rules are instantiated in all subjects to approximately the same degree.
5) The strength of a boundary is proportional to the number of subjects indicating a boundary.
6) Parsimony is important for all analyses.

Obviously, Assumption 1 is necessary to make the analysis tractable. However, it is important the realize that since both the rules and the empirical data are bounded continuous values between 0.0 and 1.0, the use of non-linear relationships is problematic. In the range from 0 to 1, the linear term is highly correlated with both the odd and even powers (hence, exponential and logarithmic functions). Assumptions 2 and 3 are also necessary to make the analysis tractable and they are consistent with the theory of Lerdahl and Jackendoff (1983). Assumptions 4 and 5 seem manifestly reasonable given the empirical results presented previously (cf., Chapter 3, this work). That is, given that all subjects produced similar profiles, it is unlikely that the similarity of different subjects truly reflects the use of different rules coincidentally at the same point. For example, it is unlikely that Subject A who used only Rule 3d and not Rule 2b would produce a boundary profile similar to Subject B who used only Rule 2b and not Rule 3d. However, in future work, analyses should be conducted at the individual level and then averaged. As such, one can view these assumptions as another manifestation of analytic tractability. Assumption 6 serves to remind that the point is not just an exercise in statistics. The new method must be an improvement on the old.

As stated previously, the basic problem is the unequal importance of false alarms and misses. Ideally, in fitting the empirical data to each rule, misses should be "punished" more heavily than false alarms. As shown in Figure 4.3A, the typical regression approach does not do this. That is, the error term in regression models is:

$$SS_{error} = \Sigma(X_i - X'_i)^2$$

where: X_i is the empirical data
X'_i is the predicted value of the data based on the rule (the mean, or the prediction based on the best fit line)

As can be seen in Figure 4.3A, the function is symmetric about 0.0. False alarms (the magnitude of the rule is greater than the empirical boundary) are weighted equally with misses (the magnitude of the rule is less than the empirical boundary). What is desired is an error term that resembles that which is presented in Figure 4.3B. Designing such an asymmetric error term is straightforward. For example,

$$\begin{aligned} SS_{error} &= \Sigma(X_i - X'_i)^2 && \text{if } X'_i \leq X_i \\ &= 0 && \text{if } X'_i \geq X_i \end{aligned}$$

where: X_i is the data to be fitted (i.e., the empirical boundary)
X'_i is the predicted value of the data (i.e., the rule under consideration)

or

$$\begin{aligned} SS_{error} &= \Sigma(\text{Boundary}_i - \text{Rule}_i)^2 && \text{if Rule}_i \leq \text{Boundary}_i \\ &= 0 && \text{if Rule}_i \geq \text{Boundary}_i \end{aligned}$$

will work (false alarms are not punished at all, which may not be the most desirable; see Figure 4.3B). However, before choosing an error function it is important to realize that the analysis will have to use non-linear regression to determine the solution (an iterative approach). Such techniques generally require (or prefer) error functions that are continuous and have continuous derivatives (cf., Bevington & Robinson, 1992; SPSS, 1988). However, in fact any error function of the form:

$SS_{error} = \Sigma(X_i - X'_i)^n \pm (X_i - X'_i)^m$

where: X_i is the data to be fitted (i.e., the empirical boundary)
X'_i is the predicted value of the data (i.e., the rule under consideration)
n is even, m is odd, n>2, n>m

or

$SS_{error} = \Sigma(\text{Boundary}_i - \text{Rule}_i)^n \pm (\text{Boundary}_i - \text{Rule}_i)^m$

will work. Some examples are presented in Figure 4.3C. All these functions have an asymmetrical shape in which the rate of increase on the positive side of the minimum value is greater than the rate of increase on the negative side of the minimum. This information is contained within the first and second derivatives. The functions differ in the steepness of their slopes, in the relative amount of flatness in the area of the minimum, and the amount of asymmetry. Note that as n increases the functions get steeper. Also note that when n and m are close in value, the function exhibits more than one inflection point in the range of interest (-1.0 to 1.0). All this information is available in the first and second derivatives, but the graphs in Figure 4.3C make the point much more obviously. For the purposes of this analysis, n was set to 4.0 and m was set to 1.0[43]. This function is displayed in Figure 4.3D.

---

[43] Note that the choice of other values of n and m will not really alter the results. This choice represents the minimal deviation from the usual error function, a moderate slope for large deviations and a lack of multiple inflection points within the important range.

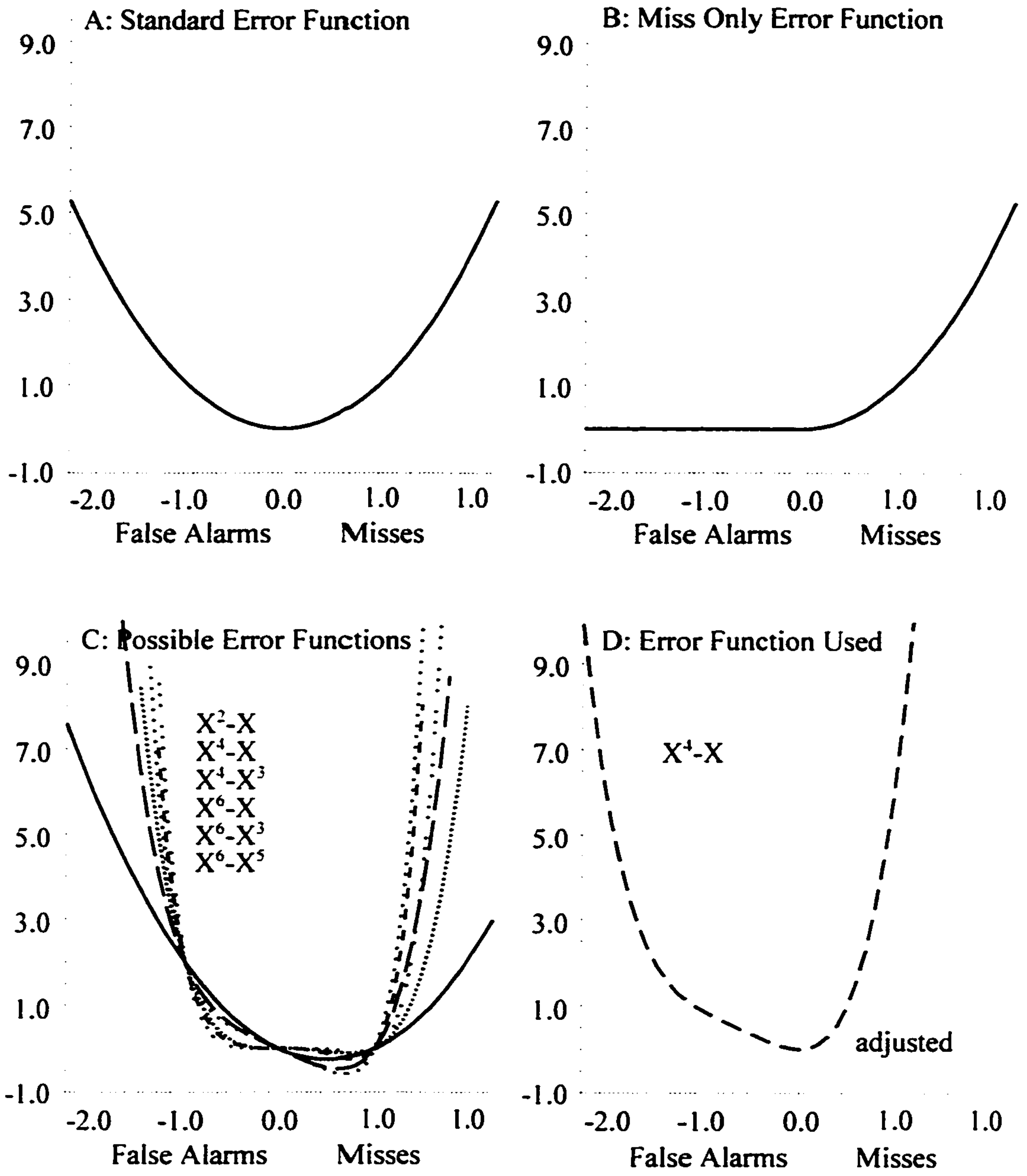


Figure 4.3: The standard error function (A, effectively X^2), the asymmetric errror function to punish misses but not false alarms (B), possible error functions (C) to model the asymmetric error function (note that all are similar, though some are flatter in the trough and some have stepper slopes), and the final choice of error function (D: effectively X^4 -X, adjusted to a zero minimum at zero error).

The function was adjusted so that the minimum error occurred when $X_i = X'_i$ (i.e., the error is minimized when prediction matches empirical data) and the function as adjusted to produce a minimum error value with a magnitude of 0.0 at the point $X_i = X'_i$. Hence the error function is:

$$SS_{error} = \Sigma[(\Delta X_i + \sqrt[3]{0.25})^4 - (\Delta X_i + \sqrt[3]{0.25})^1 - (\sqrt[3]{0.25^4} - \sqrt[3]{0.25})]$$

$$= \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$$

$$\text{where: } \Delta X_i = X_i - X'_i = \text{Boundary}_i - \text{Rule}_i$$

Since the goal of this analysis was descriptive (not inferential), the only quantity of interest is the relative amount of error produced by each rule. Rules that produce lower errors fit the empirical data better (i.e., "significance" is not an issue here; However, those who are interested could use Monte-Carlo techniques to refine the analysis). Another way to view this error term is to realize that those rules producing lower error values predict more of the actual boundaries, with due regard for the continuous nature of the empirical data and rules. Of course, rules generating lower errors may also be generating more false alarms. Using this error function, the optimal fit between each rule and the empirical data can be determined for the equation:

$$X_i = b*X'_i + error_i$$

where X_i is the empirical data (i.e., the boundary)
X'_i is the rule value (i.e., the rule)
$error_i$ is the error in the fit

or

$$Boundary_i = b*Rule_i + error_i$$

The value of b is the slope, which can be viewed in the same manner as any slope (subject to the constraint that inferences to populations should not be made). Note that no intercept is included. The intercept would amount to a constant reflecting, for the most part, the average value of empirical responses at the points where the rule did not apply. That is, the intercept would reflect the value of X_i when $X'_i = 0$ (when the rule did not apply). From the perspective of an individual rule, empirical boundaries should be ignored (i.e., considered equal to 0) when the rule does not apply: The fit should reflect the ability of the rule to predict boundaries when the rule is designed or expected to predict boundaries. Mathematically, the operation is equivalent to regression in which the solution is constrained to pass through the origin. This analysis was applied to the concocted data presented in Figure 4.2. The results are intended to demonstrate how the error function separates false alarms and misses.

Table 4.1: Fit Between Theoretical and Empirical Data of Figure 4.2 Using Asymmetrical Error Terms.

Theory		*r*	Equation	SS_{error}
No False Alarms or Misses	High	1.000**	b = 1.000*rule	0.000
	Medium	0.984**	b = 1.072*rule	0.159
	Low	0.754**	b = 1.557*rule	1.402
Rule "A": False Alarms	High	0.744**	b = 0.828*rule	1.568
	Medium	0.708**	b = 0.875*rule	1.869
	Low	0.450**	b = 1.168*rule	4.083
Rule "B": Misses	High	0.834**	b = 1.000*rule	3.224
	Medium	0.825**	b = 1.000*rule	3.293
	Low	0.580**	b = 1.548*rule	4.264
Rule "C": False Alarms and Misses	High	0.545**	b = 0.778*rule	4.723
	Medium	0.541**	b = 0.781*rule	4.798
	Low	0.290	b = 1.019*rule	7.115

Notes: *r* = simple correlation

$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$

High: Predictions by rules, when they apply, exactly equal the empirical data.

Medium: Predictions by rules, when they apply, approximate the empirical data.

Low: Predictions by rules, when they apply, do not approximate the empirical data.

Consider the situation of No False Alarms or Misses (compare Table 4.1 and Figure 4.2). The analysis identified the high, medium and low correspondence conditions appropriately. Note that, in general, Rule "A" is better than Rule "B" which in tern in better than Rule "C" by the new method, but Rule "A" is not the best by the old (correlational) method. Generally, the profiles containing false alarms do not produce the same magnitude of errors as the profiles containing misses. Also note that the high match "No False Alarms or Misses" profile produces a slope of 1.0 and an error of 0.0 (and a correlation of r=1.0; it was designed this way). The important point is that the corresponding high match Rule "B" (Misses) profile also has a slope of 1.0. This slope maximizes the potential of each theoretical boundary to predict the empirical boundary. The fact that there is a miss does not affect the slope (it does affect the error). Usually regression/correlation cannot do this. Finally, note that the Rule "A" (False Alarms) does not produce a slope of 1.0, but it does produce a lower error value (if false alarms were to be ignored in the error term, then the slope would be 1.0).

Applying the New Analysis

This new method was applied to the data from the previous experiment (see Chapter 3, this work) to insure that the new method was at least as good as the old method (correlation and regression), but one major simplification was made. In the analysis of actual data, the last two presentations within each stage were averaged. That is, Repetitions 2 and 3 of the melody in Stage 4 were averaged, Repetitions 2 and 3 of the melody in Stage 6 were averaged and Repetitions 5 and 6 of the melody in Stage 8 were averaged (Stages 6 and 8 used the same melody. The rules were compared with those averages using the asymmetric error terms (correlations are provided to simplify comparisons). This tack, taken to eliminate redundant analyses, is valid because of the high correlations between the respective repetitions for all subjects (see Chapter 3, this work). In addition, an initial attempt was made to extend the analysis to the higher level rules (Rules 4 [Intensification], 5 [Symmetry] & 6 [Parallelism]).

Low-Level Rules (Rules 2 and 3)

Given the asymmetric error function, the results for "Three Blind Mice" (Stage 4, Part 1), for each rule are presented in Table 4.2 and Figure 4.4. Only Rules 2b (Attack-point change) and 3a (Register change) are actually shown because only those rules applied to this melody. In Figure 4.4, the empirically determined boundaries are shown in the top panel, to be compared with the best fit values for each of the rules in the middle panel (Rule 4 is include in the lower panel of this figure to optimize the utilization of space; Rule 4 will be discussed momentarily).

In this analysis, for each rule, the overall error for the entire melody was computed. This error value represents the ability of the rule to account for empirical boundaries in the general sense. Rule 2b (Attack-Point Change) was much more effective than Rule 3a (Register Change): The overall error for Rule 2b was much less than the overall error for Rule 3a. The results are similar to the simpler correlational analyses (see Chapter 3 of this work for details). Note that the difference between the rules exists even though both rules were applicable at the same total number of total positions (11).

To refine the analysis, the error rates for misses and false alarms are presented separately. These values were only computed at the points where the rules applied so as to assess the fit in those cases when the rule actually applied. In this analysis, one can assess the relative contributions of false alarms and misses to the total error. It must be remembered that the point was to minimize misses and let false alarms rise to whatever level was necessary. This allows for a direct comparison of the different rules at those points where they were expected to work. In this analysis, Rule 2b was still much superior than Rule 3a. The error for misses of Rule 2b was half the value of the error for misses of Rule 3a, despite the fact that both rules had the same number of misses (places where the best-fit theoretical rule was greater than the empirical data). The same was true of errors for false alarms. Rule 2b was much better than Rule 3a, producing about one-half the error value on the same number of occurrences. In conclusion, it can be said that the better fit of Rule 2b was not due to misses or false alarms alone. Rather, Rule 2b was simply better overall. The relatively higher error value for misses in Rule 3a is a statement that it is not possible to align the predictions of Rule 3a with the data, even if one effectively ignores false alarms.

Table 4.2: Rules 2 and 3 for "Three Blind Mice".

Overall Results				
Rule	n	r	Equation	SS_{error}
2b	11	0.770**	b = 1.187*2b	3.592
3a	11	0.340**	b = 0.926*3a	12.777

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b	4	0.907	0.852	7	1.933	2.935
3a	4	1.987	1.316	7	4.166	5.258

Notes: r = simple correlation (see Chapter 3, this work)
n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

It is important to realize that in the overall error term, every position where there was an empirical boundary, but the rule did not apply, would be included as a miss. Hence, the sum of the errors for misses and false alarms would not necessarily produce the total error. Given the equation to be fitted, all points at which a rule did not apply would not be a factor in the fit. That is, all points at which the rule did not fit would simply contribute an invariant constant (all as misses) to the error term, regardless of the final parameters chosen. The prediction b*2b would produce 0 at all points where Rule 2b did not apply (because the strength of Rule 2b at those points would be 0.0). Hence, the miss at each of those locations would simply be the value of the empirical boundary. This error would be a constant regardless of the value of b. This analysis is different from the previous correlational/regression analysis (see Chapter 3 of this work).

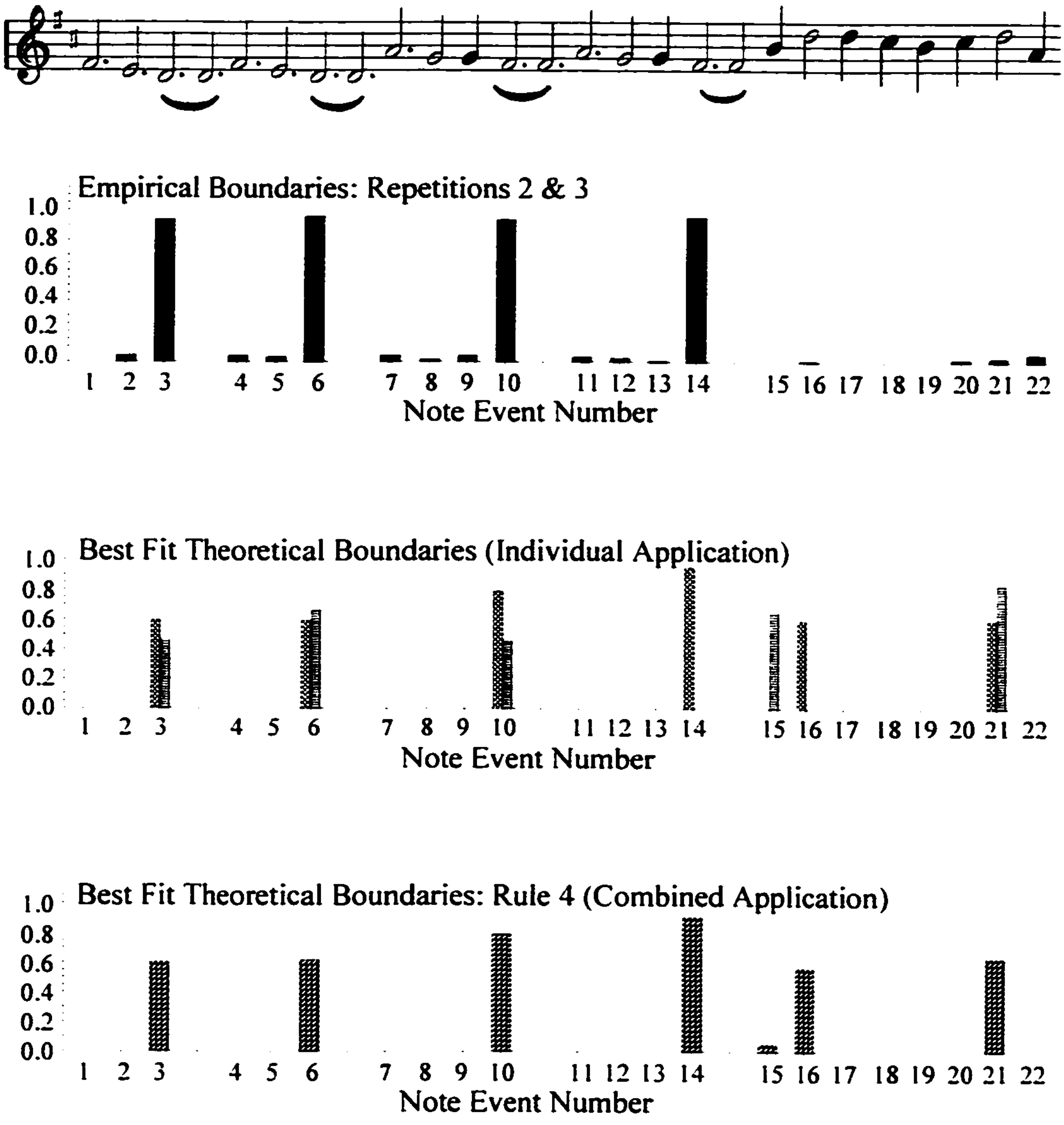


Figure 4.4a: The first part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and 3a. Rule 4 represents the combination of Rules 2 and 3 (see later discussion in text).The rules can be compared to the location of empirical boundaries based on the average of all subjects over the last two repetitions of the melody.

▩ Rule 2b ▩ Rule 4

▩ Rule 3a

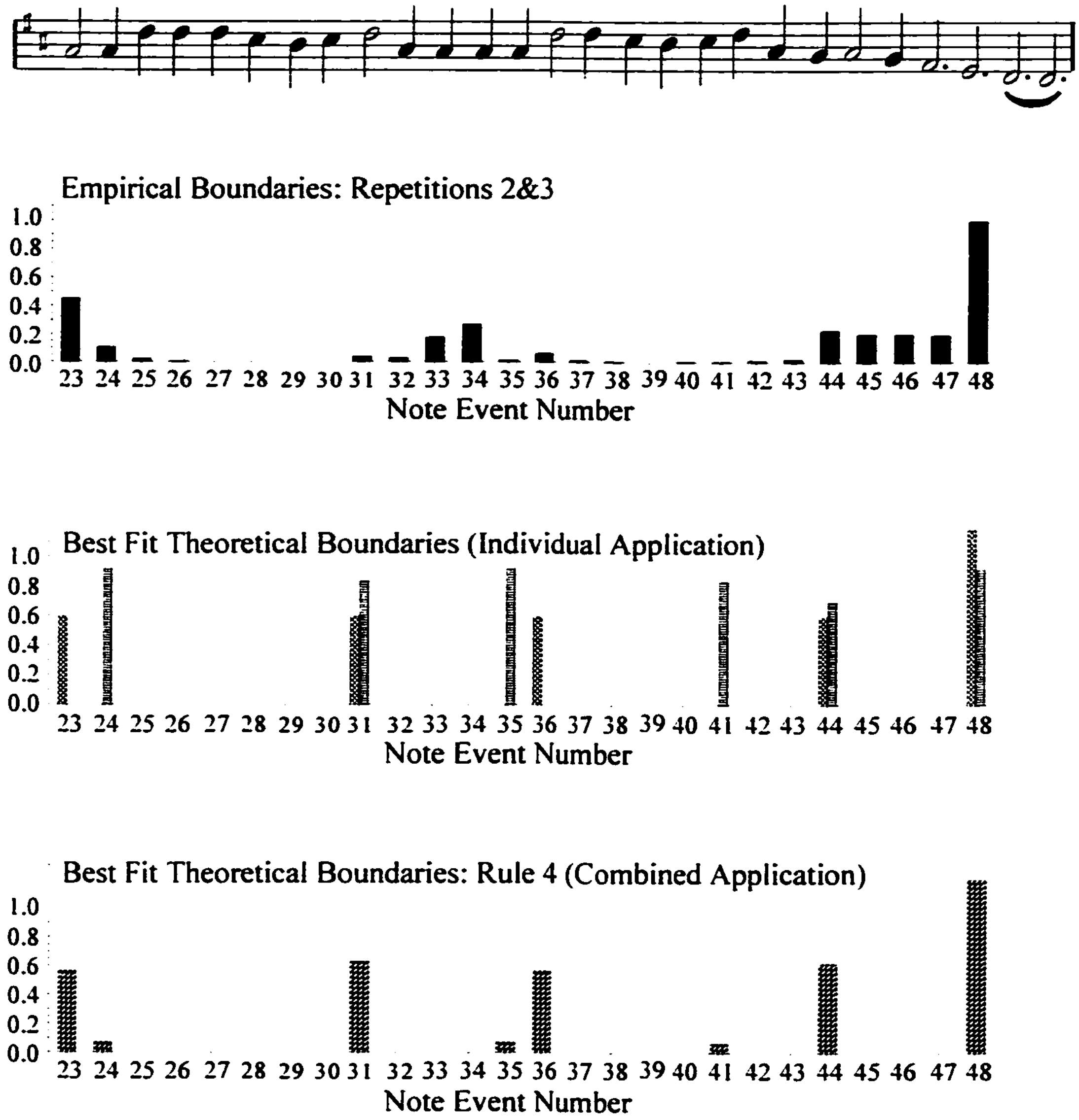


Figure 4.4b: The second part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and 3a. Rule 4 represents the combination of Rules 2 and 3 (see later discussion in text).The rules can be compared to the location of empirical boundaries based on the average of all subjects over the last two repetitions of the melody.

▨ Rule 2b ▨ Rule 4

▨ Rule 3a

For the extract from the melody "Softly Now the Light of Day", the situation was very similar (Tables 4.3 & 4.4 and Figure 4.5), although there were three rules to consider and two different presentations (Stage 6 and Stage 8). In Figure 4.5, the empirically determined boundaries are shown in the top panel, to be compared with the best fit values for each of the rules in the middle panel. Again, the lower panel refers to Rule 4, which is to be discussed later.

As in Stage 4 and the previous analyses (see Chapter 3 of this work), the results indicated that Rule 2b was the most important rule even though it only applied at five locations within the melody. Rule 3a could actually be applied at more locations (9), but still produced a much higher overall error than Rule 2b. Rule 3d produced the higher level of overall error on five potential locations. By inspection of Figure 4.5, it is obvious that Rule 2b applied at all the major empirical boundaries, while the other rules did not. In comparison to Stage 4 ("Three Blind Mice") the total number of times that Rule 2b could be applied dropped to 5 from 11 but the error rate showed a similar decline (i.e., In Stage 6, both number and total error declined to 45% of their value in Stage 4; the average error rates are the same). The implication is that Rule 2b was equally effective in both melodies. However, it must be remembered that the melody of Stages 6 and 8 contained 34 notes in comparison to the melody of Stage 4 which contained 48 notes: Direct comparisons between the two melodies should be made judiciously.

Results for misses and false alarms separately provided much the same story (see Table 4.3). However, from this breakdown, it would seem that Rule 3d was preferable to Rule 3a: It produced a smaller error for misses and false alarms (bearing in mind that it occurs at fewer positions). It appears that the reason Rule 3d did not perform well is that it simply did not apply at the more informative parts of the melody (i.e., points where empirical boundaries were placed): For Rule 3d, the high value for the overall error, when compared to the miss and false alarm errors, implies that Rule 3d did not apply when the empirically determined boundaries were high (i.e., it did not apply in a lot of real situations). However, when Rule 3d did apply, it can be aligned quite well with the empirical data. Rule 3a, on the other hand, produced a lot of false alarms, but even given the bias to ignore false alarms, it still produced a large error for misses. That is, on the occasions when Rule 3a did apply, it was not possible to match the value of the rule to the empirically determined value with a precision that is even remotely close to that of Rule 2b or 3d.

Table 4.3: Rules 2 and 3 for the extract from "Softly Now the Light of Day" (Stage 6).

<table>
<tr><th colspan="5">Overall Results: Stage 6</th></tr>
<tr><th>Rule</th><th>n</th><th><i>r</i></th><th>Equation</th><th>SS_{error}</th></tr>
<tr><td>2b</td><td>5</td><td>0.877**</td><td>b = 1.186*2b</td><td>1.650</td></tr>
<tr><td>3a</td><td>9</td><td>0.313</td><td>b = 0.695*3a</td><td>9.104</td></tr>
<tr><td>3d</td><td>5</td><td>0.367*</td><td>b = 0.778*3d</td><td>11.110</td></tr>
</table>

<table>
<tr><th colspan="7">Breakdown for Misses and False Alarms: Stage 6</th></tr>
<tr><th>Rule</th><th colspan="3">Misses</th><th></th><th colspan="2">False Alarms</th></tr>
<tr><th></th><th>n</th><th>SS_{error}</th><th>|Deviation|</th><th>n</th><th>SS_{error}</th><th>|Deviation|</th></tr>
<tr><td>2b</td><td>2</td><td>0.237</td><td>0.338</td><td>3</td><td>0.454</td><td>0.741</td></tr>
<tr><td>3a</td><td>2</td><td>0.689</td><td>0.645</td><td>7</td><td>1.708</td><td>2.756</td></tr>
<tr><td>3d</td><td>1</td><td>0.147</td><td>0.222</td><td>4</td><td>0.610</td><td>1.105</td></tr>
</table>

Notes: *r* = simple correlation (see Chapter 3, this work)
n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|\text{Deviation}| = \Sigma(X_i - X'_i)$

For Stage 8 (Table 4.4, see also Figure 4.5), the results were essentially unchanged, producing the same slopes and errors as in Stage 6, as would be expected given the high similarities of the averaged boundaries in these two stages. Note that Figure 4.5 presents the fit of the rules to the empirical data for Stages 6 and 8 in separate panels. Inspection demonstrates very little difference.

Table 4.4: Rules 2 and 3 for the extract from "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8				
Rule	n	*r*	Equation	SS_{error}
2b	5	0.883**	b = 1.265*2b	1.520
3a	9	0.310	b = 0.722*3a	10.286
3d	5	0.334	b = 0.768*3d	13.023

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b	3	0.271	0.361	2	0.567	0.864
3a	2	0.736	0.654	7	1.844	2.899
3d	1	0.162	0.232	4	0.633	1.160

Notes: *r* = simple correlation (see Chapter 3, this work)
n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|\text{Deviation}| = \Sigma(X_i - X'_i)$

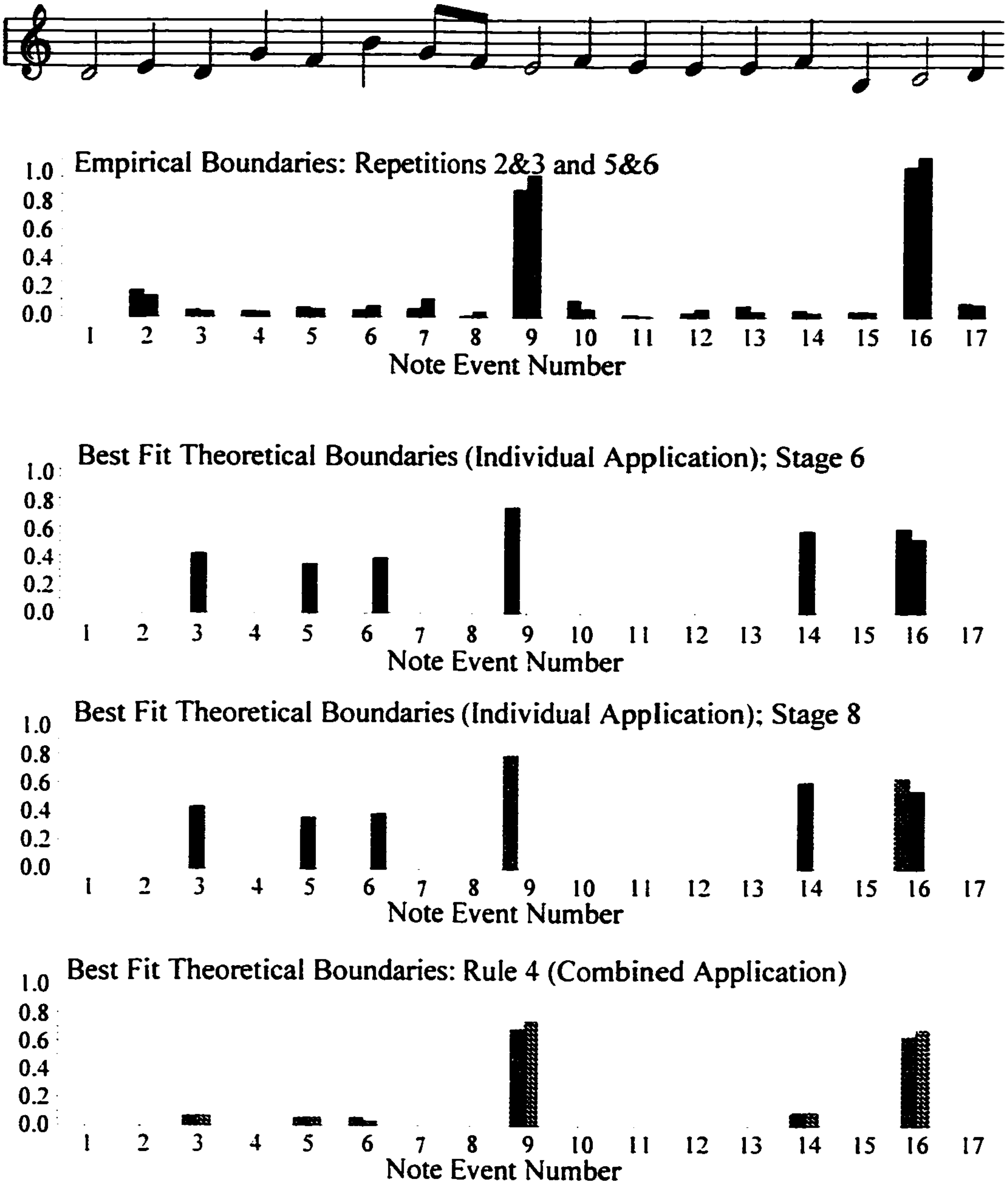


Figure 4.5a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given the best individual fits of Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d to the empirical data (based on the average of Repetitions 2 and 3 and average of Repetitions 5 and 6). Rule 4 represents the best combination of Rules 2 and 3 (see later discussion in text).

▨ **Rule 2b**

▨ **Rule 3a**

▨ **Rule 3d**

▨ **Rule 4**

■ **Stage 6 (Presentations 2 & 3)**

■ **Stage 8 (Presentations 5 & 6)**

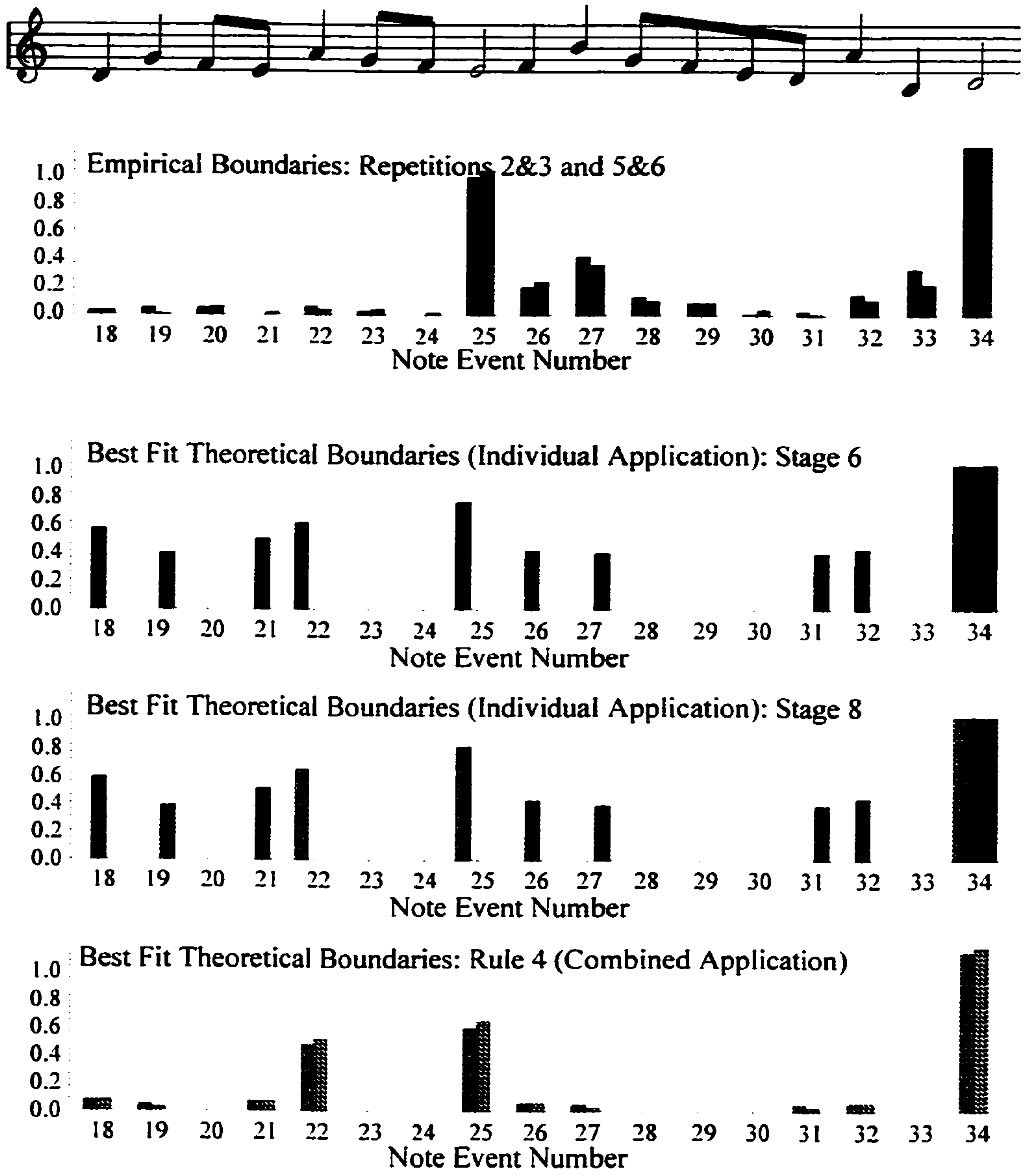


Figure 4.5b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given the best individual fits of Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d to the empirical data (based on the average of Repetitions 2 and 3 and average of Repetitions 5 and 6). Rule 4 represents the best combination of Rules 2 and 3 (see later discussion in text).

▩ Rule 2b
▩ Rule 4
■ Stage 6 (Presentations 2 & 3)
■ Rule 3a
■ Stage 8 (Presentations 5 & 6)
▨ Rule 3d

Extension of the Analysis to the Higher Level Rules

(Rules 4, 5, and 6)

Although interesting on their own merits, the main point of these re-analyses was not the confirmation of previous results. Rather, the intent was to provide a firmer (fairer) basis for comparisons between rules, and for extensions to more complex analyses. The analyses provide a scaling for each rule that maximizes the potential for each rule to predict boundaries (effectively ignoring false alarms). This goal was actually a subgoal in the more general desire to extend the analysis to the higher level rules: Rule 4 (Intensification), 5 (Symmetry) and 6 (Parallelism).

These higher level rules build on the potential boundaries identified by the lower level rules. In principle, the use of the higher level rules in conjunction with the lower level rules, should allow for a very close match between the theory and empirical data. This match may not be perfect because subjects may be parsing at a level beyond that which is captured by all of these rules (Rule 2, 3, 4, 5 and 6), but it should be better than any single rule in isolation.

On the basis of the previous analysis, the individual rules were adjusted so that there would be a minimal number of misses. That is, the strength of each rule was adjusted so that each rule would predict as many boundaries as possible. Hence, the combination of Rules 2a, 2b, 3a and 3d should not miss any boundaries evident in the data. Rules 2 and 3 represent the lowest level of the hierarchy. No empirical boundary should exist at any level of the hierarchy if there was not a corresponding boundary at every lower level. Hence, Rules 2 and 3 should define every possible boundary. That is, the combination (i.e., either rule A or rule B) of these low level rules should result in a prediction having no misses. Actually, the prediction would be no misses if the data were binary. Since the data are continuous, the prediction for the combination should be a small value for the error of misses. In this analysis, false alarms were essentially ignored (there was a small penalty for false alarms). Hence, the combination of Rules 2a, 2b, 3a and 3d would generate a lot of false alarms.

As one moves up the hierarchy from the low level rules to the actual parsing by subjects, potential boundaries are eliminated. That is, false alarms are eliminated. Hence, at this point in the analysis, the goal is to winnow out the false alarms by using Rules 4, 5 and 6 without generating any new misses. The difficulty that must be acknowledged is that boundaries may be based on any one of a number of possibilities. That is, a boundary may be caused by:

1) a strong single rule (this might be Intensification)
2) the joint action of several weak rules (this might be Intensification)
3) a weak rule boosted by Symmetry or Parallelism
4) the joint action of several weak rules boosted by Symmetry or Parallelism

Each of these is dealt with in turn. In this discussion, it must be recognized that strong and weak are relative terms defined on the basis of a continuum, and that, in fact, strong and weak refer to the degree to which the music instantiates a rule (not necessarily the degree to which a rule causes a boundary).

Rule 4: Intensification

According to Lerdahl and Jackendoff, "Where the effects picked out by GPRs 2 and 3 are relatively more pronounced, a larger-level group boundary may be placed" (1983, p. 49). The ambiguity leaves wide room for interpretation. On first inspection, the rule seems to imply that stronger instances of Rules 2a, 2b, 3a, 3b, 3c or 3d lead to a higher probability of a boundary. This is, in fact, how each of the low level rules has been implemented in the algorithm of this thesis, with a resulting numerical assessment of rule strength (which, should lead to boundary strength). As has been noted, this rule strength may not be exactly as envisioned by Lerdahl and Jackendoff (interestingly, they refused to provide a quantitative assessment of their own [1983, p. 55], but then alluded to stronger instances having a role in boundary formation [1983, p. 49]). As such, conceiving Rule 4 (Intensification) as the strength of a single rule does not seem adequate as a basis for a new rule. It would serve no function.

At another point, Lerdahl and Jackendoff comment "...the various cases of GPRs 2 and 3 may reinforce each other, producing a stronger sense of boundary, as in...Alternatively, different cases of the rules may come into conflict, as in..." (1983, p. 46). Lerdahl and Jackendoff also discuss this idea of the joint effects somewhat obliquely when discussing confluence (1983, p. 67). Unfortunately for the quantification of Rule 4, Lerdahl and Jackendoff never explicitly state whether or not the joint action of rules follows an additive factors logic (or an interaction).

From exemplars provided by Lerdahl and Jackendoff, a more reasonable conception for Rule 4 would be a statement that when two or more rules (from the set of Rules 2 and/or 3) indicate boundaries at the same location, the net effect is to increase the probability for a boundary in proportion to the sum of the rules (worrying, later, about how to do that sum). That is, rules act in an additive manner. This is an additive factors model. It must be admitted that it is possible for two or more rules to interact. That is, when two rules indicate boundaries at the same location, the emergent effect is greater than (or different from) the simple sum of the two rules.

On this basis, then, in this analysis, Rule 4 (Intensification) was considered first as the simple sum of the rules. Empirical data were fitted to an equation of the form:

$$\text{Boundary}_i = a*2a_i + b*2b_i + c*3a_i + d*3b_i + e*3c_i + f*3d_i + \text{error}_i$$

where: Boundary_i is the empirical boundary
a, b, c, d, e, f are the slopes
2a, 2b, 3a, 3b, 3c, 3d are the theoretical rules
error_i is the error of the fit
i is the note event number of the melody

Note that this analysis is simply that which has been discussed from the beginning of this chapter. That is, previously with reference to Figure 4.1 and 4.2, it was noted that Rules "B" and "C" in conjunction might represent a better fit to the data than Rule "A" alone, event though Rule "A" was the best single rule. The analysis was designed from the start to ask the question, "What combination of rules is best?", given that false alarms and misses are not to be weighted equally and given that both the empirical data and the rules are continuous.

In actuality, only the rules that applied to each melody were entered into the equation (i.e., Rules 2b and 3a for "Three Blind Mice" and Rules 2b, 3a and 3d for the extract from the melody "Softly Now the Light of Day"). This assessment produced the best fit between all the rules and the empirical data, still granting more weight to misses than false alarms.

To assess interactions, one typically creates an interaction vector as the product of two main vectors. In this case, that is not the optimal solution. The problem is that the number of interaction vectors increases rapidly with the number of main effects vectors. In the case of the melody "Three Blind Mice", there are two main effects vectors (Rules 2b and 3a) leading to one interaction vector (2b*3a). Since the melody contains 48 notes, the analysis would consist of three vectors predicting 48 data points, which is a reasonable 16 data points per vector. However, for the melody "Softly Now the Light of Day", there are three main effects vectors (Rules 2b, 3a and 3d) producing four interaction vectors (2b*3a, 2b*3d, 3a*3d, and 2b*3a*3d). Since the melody is only 34 notes in length, there would be only five data points per vector. The latter inclusion of additional rules would further reduce this number. To minimize these problems a priori, only the second order interactions were entered. In all analyses, the interaction vectors were created from the best fit versions of the basic rules (i.e., the rules after multiplying by their respective slopes).

In the melody "Three Blind Mice" (Table 4.5), the basic summation of the two rules increase the total number of points predicted to 15 (from 11 for Rule 2b alone, or 11 from Rule 3a alone). This was because Rules 2b and 3a overlap at seven points. The use of both rules simultaneously improved the overall fit slightly (3.592 for Rule 2b alone down to 3.540) as well as the misses (0.907 for Rule 2b alone down to 0.828) while false alarms rose slightly (1.931 for Rule 2b up to 1.991) from the previous best case. The decrease in error for misses is good, especially when one considers that the total number of positions accounted for has risen from four to five. The increase in errors for false alarm is not a worry since the total number of false alarms has risen from seven to eleven. Note that the error for false alarms is still less than the previous worst case. From the equation, one can see that the best fit of the combination emphasized Rule 2b and restricted Rule 3a to a very subsidiary rule. The slope for Rule 2b was almost unchanged from its value when used as the only predictor (now: 1.113; previously 1.187), while the slope for Rule 3a was greatly reduced (now: 0.072; previously: 0.926).

It important to realize that the improvement in fit is the result of two separate factors. Firstly, by use of the combination, more empirical boundaries are accounted for than by any one rule alone. Secondly, the use of the addition of two rules allows the slope of each rule to be reduced. This has a major effect on false alarms (it reduces the overprediction when there is no empirical boundary). However, the analysis clearly indicates that most of the effects that were previously granted to Rule 3a were better explained by Rule 2b. That is, Rule 3a and 2b overlapped on seven points, and most of that overlap was best explained by Rule 2b alone. In the area where Rule 3a acted alone, the algorithm did not find any evidence for predictive efficacy. The evidence for this is that the slope for Rule 3a declined dramatically when the rule was used in conjunction. Figure 4.4 allows one to compare this combined effect (bottom panel) with the individual effects (middle panel) and the empirical data. Note visually, that the prediction based on the combined rules is a better match to the data, particularly in regard to false alarms, but the combined data is very similar to Rule 2b alone.

Table 4.5: The Combination of Rules 2 and 3 (Rule 4) for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.131*2b + 0.072*3a	3.540
2b+3a+2b*3a	15	b = 1.131*2b + 0.072*3a + 0.000*2b*3a	3.540

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	5	0.828	0.900	10	1.991	3.123
2b+3a+2b*3a	5	0.828	0.900	10	1.991	3.123

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

The interaction term was entered in a separate analysis, in the manner of hierarchical (theory-driven) regression. The introduction of the interaction term did not improve the fit at all (the error did not change at all), and the slope for the interaction term was 0[44]. The implication is that there is nothing special about the simultaneous occurrence of two rules that cannot be explained by the simple sum of their individual contributions.

For the second melody, the extract from "Softly Now the Light of Day" (Table 4.6 and 4.7), the results again mimiced the earlier correlational analysis in both Stage 6

---

[44] Note that all slopes were constrained to be greater than 0, since negative slopes had questionable interpretation. In the discussion that follows, 0 can be taken to mean anything less than .0005, and some were not identically 0. However, since all values are bounded by 0 and 1 (rules and empirical data), all equally meaningful slopes, even interactions, should be within the same magnitude or so.

(Repetitions 2 and 3 averaged) and Stage 8 (Repetitions 5 and 6 averaged).

In Stage 6 (Table 4.5), the overall error value, the misses value and the false alarm value are quite low. In addition, the inclusion of interactions did not aid in prediction. For the analyses of interaction, first each interaction was included in isolation, and then all three interactions, were included in the model. In no case, did the interactions produce a significant contribution. One can also see the effect of the combined action of the rules for Stage 6 in Figure 4.5; note that the combined action is a very good predictor of boundaries. In Stage 8, the general result is very similar (Table 4.7).

Generally, as in the previous melody, Rule 2b is clearly the most important. For Stage 6, Rules 3a and 3d are approximately equivalent. For Stage 8, the emphasis given to Rule 3d is somewhat decreased. If anything, subjects seem to rely more on Rule 2b and slightly less on Rule 3d in the later repetitions of the melody. The implication is that Rule 3d tended to indicate boundaries at the same location as Rule 2b and as such , it added nothing to the prediction. Rule 3a, on the other hand, helped by creating predictions where Rule 2b did not apply. Nonetheless, in neither case was the rule important.

It is also interesting that the fit is slightly better overall in Stage 8 than in Stage 6, implying that Lerdahl and Jackendoff's notion of the ideal listener may be more closely approximated after one has some familiarity with the tune. Figure 4.5 also displays the results for Stage 8 graphically.

It is important to note that the fit in Stages 6 and 8 is better than the fit in Stage 4. By inspection of the figures, this makes sense: All empirical boundaries are identified in Stage 6 and 8, but not all empirical boundaries are identified in Stage 4. The previous correlational analysis did not capture this (both melodies returned multiple squared-correlations in the range of R^2=0.50 to 0.60; in fact, Stages 6 and 8 were slightly lower than Stage 4).

Table 4.6: The Combination of Rules 2 and 3 (Rule 4) for the extract from "Softly Now the Light of Day", in Stage 6.

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b = 1.093*2b+0.115*3a+0.116*3d	1.467
2b+3a+3d+2b*3a	16	b = 1.131*2b+0.072*3a+0.116*3d+0.000*2b*3a	1.467
2b+3a+3d+2b*3d	16	b = 1.131*2b+0.072*3a+0.116*3d+0.000*2b*	1.467
2b+3a+3d+3a*3d	16	b = 1.131*2b+0.072*3a+0.116*3d+0.000*3a*3d	1.467
all	16	b = 1.131*2b+0.072*3a+0.116*3d +0.000*2b*3a+0.000*2b*3d+0.000*3a*3d	1.467

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.500	0.841	10	0.551	1.118
2b+3a+3d+2b*3a	6	0.500	0.841	10	0.551	1.118
2a+3a+3d+2b*3d	6	0.500	0.841	10	0.551	1.118
2a+3a+3d+3a*3d	6	0.500	0.841	10	0.551	1.118
all	6	0.500	0.841	10	0.551	1.118

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Generally, the results support a simple additive factors model for the effect of the different low level rules on the parsing of a melody. The general lack of interactions between the low level rules was consistent for both melodies. As such, it implies that these interactions can safely be ignored in the subsequent analysis of parallelism.

Table 4.7: The Combination of Rules 2 and 3 (Rule 4) for the extract from "Softly Now the Light of Day", in Stage 8

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b = 1.188*2b + 0.115*3a + 0.060*3d	1.380
2b+3a+3d+2b*3a	16	b = 1.131*2b+0.072*3a+0.000*3d+0.000*2b*3a	1.380
2b+3a+3d+2b*3d	16	b = 1.131*2b+0.072*3a+0.000*3d+0.000*2b*3a	1.380
2b+3a+3d+3a*3d	16	b = 1.131*2b+0.072*3a+0.000*3d+0.000*2b*3a	1.380
all	16	b = 1.131*2b+0.072*3a+0.000*3d+0.000*2b*3a	1.380

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.486	0.868	10	0.650	1.214
2b+3a+3d+2b*3a	6	0.486	0.868	10	0.650	1.214
2a+3a+3d+2b*3d	6	0.486	0.868	10	0.650	1.214
2a+3a+3d+3a*3d	6	0.486	0.868	10	0.650	1.214
all	6	0.486	0.868	10	0.650	1.214

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Rule 5: Symmetry

Rule 5 states "Prefer grouping analyses that most closely approach the ideal subdivision of groups of two parts of equal length" (Lerdahl & Jackendoff, 1983, p. 49).

After some consideration, it was decided that this rule was not amenable to analysis at this point. The problem is, as stated, that the rule only comes into play once one has firm boundaries at two locations and a question about the location between those two points. That is, Rule 5 acts in a manner like Rule 1 (avoid groups consisting of a single note). Rule 5 is best used to resolve disputes between two equally viable candidates. Alternatively, if one takes the listener's perspective, to use Rule 5 online, it would have to be cast as "given that the last boundary was X musical events from its previous boundary, prefer to place a new boundary at that same number of musical events from the last boundary". That is, imagine the situation in which we know that there is a boundary at point A and we know that there is a boundary at point B and we have a choice of a boundary at points C or D (suppose that placing a boundary at both C and D would violate Rule 1). Measure the distance between boundaries A and B. Measure the distance from boundary B to C. Measure the distance from boundary B to D. If the distance from B to C matches the distance from A to B, place the boundary at C. If the distance from B to D matches the distance from A to B, then place the boundary at D. Of course, all this presupposes a notion of distance (beats?, notes?, milliseconds?) and all this should properly be phrased in terms of relative degrees of match. Also note that since the possible locations, C and D, are in close proximity to each other (else there would not be any conflict between the two), proper definitions of distance matter. Most importantly, both Rule 1 and Rule 5 require that we *know* where other boundaries actually are (as opposed to where other boundaries *might* be given low level rules). Therefore, the use of Rule 5, like Rule 1, will have to wait until boundary locations can be stated with more certainty. In this regard, Rule 5 is distinct from Rules 2, 3, 4 (discussed previously) and 6 (discussed momentarily). When listening to music, the subject possibly uses something like Rule 5 to aid in the parsing of the music. However, this is because the subject has parsed the music just heard previously, and as such can use those previous boundaries as a constraint on the determination of current boundaries. When modeling the actions of the subject, we do not have the luxury of knowing those previous boundaries.

Rule 6: Parallelism

Lerdahl and Jackendoff's statement of Rule 6, "Where two or more segments of the music can be construed as parallel, they preferably form parallel parts of groups." (1983, p. 51) leaves open the definition of parallel.

This is a common problem in both the theoretical and empirical study of music. It is possibly an overstatement to claim that everyone acknowledges the importance of parallelism in music, but the strange corollary is that there are few operational (useful) definitions (see Chapter 1, this work). For example, Lerdahl and Jackendoff (1983, p. 52) claim:

> The importance of parallelism in musical structure cannot be overestimated. The more parallelism one can detect, the more internally coherent an analysis becomes, and the less independent information must be processed and retained in hearing or remembering a piece.
>
> 1983, p. 52

However, they go on to add:

> However, our formulation of GPR 6 still leaves a great deal to intuition in its use of the locution "parallel."
>
> When two passages are identical they certainly count as parallel, but how different can they be before they are judged as no longer parallel? Among the factors involved in parallelism are similarity of rhythm, similarity of internal grouping, and similarity of pitch contour. ...But we are not prepared to go beyond this, and we feel that our failure to flesh out the notion of parallelism is a serious gap in our attempt to formulate a fully explicit theory of musical understanding.
>
> 1983, pp. 52-53

Although Lerdahl and Jackendoff (1983) do not define parallelism they, and many others (see Chapter 1, this work) have provided several insights that are the basis for this approach. Herein, based on a review of some of the literature (see Chapter 1 of this work), parallelism was considered as two separable constructs: pitch pattern and time pattern. The discussion presented previously will not be repeated here (see Chapter 1 of this work), but basically, these two concepts were intended to capture the general notion of pitch contour with due regard to interval size and the notion of rhythm within a short span of notes (often called meter).

Pitch pattern is the operationalized notion of changes in pitch (frequency) over time. Basically, within a fixed time period, the overall pitch contour was assessed using two different methods that reflect the dominant trends in the literature. Both methods explicitly involve pitch height over time, but they differ in the treatment or degree of involvement of the temporal dimension. Note that both methods explicitly include interval sizes. The simple notion of contour as the pattern of up and down without regard to interval sizes was not considered because the literature seems to have evolved to the recognition that this is not too informative (see Chapter 1 of this work).

Time pattern was operationalized as the pattern of note durations within a fixed time period. The overall time pattern was also analyzed using two slightly different methods that were distilled from the literature (see Chapter 1 of this work).

In total, there were eight types of parallelism analysis: two pitch pattern types, and two time pattern types, but each of these had two slightly different flavours (each of these will be discussed in turn). Furthermore, each analysis was done with the data of three different melodies (Stages 4, 6 and 8; Stages 6 and 8 used the same melody, but they were analyzed separately) producing 24 analyses. Finally, for every analysis, maximum beat spans from 2 to 10 were checked (this will be explained later) totaling 216 analyses. Finally, in addition, nine additional analyses used a different measure of similarity (i.e., Euclidian distance) to assess pattern parallelism. Only the most pertinent results are reported[45].

Parallelism (Rule 6) in "Three Blind Mice"

In the subsequent discussion of the four main versions of parallelism , it seemed most convenient to discuss each type of parallelism and its application to the melody "Three Blind Mice" as consecutive units. That is, the melody served as the main example for the analysis. Such a presentation makes it easier to explain the different types of parallelism, although such a presentation does make it more difficult to directly compare the different types of parallelism, particularly at the abstract level.

---

[45] Those who are interested in other analyses may contact the author to obtain more details of the other analyses.

Pitch Pattern Parallelism A (Rule 6a)

In Pitch Pattern Parallelism A, pitch contour was assessed as pitch as a function of time with the duration of individual notes included as a factor in the temporal dimension. The pitch contour (see Figure 4.6) was represented as a string of pitch values spaced at equitemporal intervals (i.e., a one-dimensional array). The equitemporal intervals used were 1/96 the value of the whole note. This notion of pitch contour is analogous to the MIDI presentation of the music and, in particular, to the representation of the MIDI idea of tics per whole note.[46] The actual pitch or frequency of the notes was represented by its MIDI note values (i.e., a semitone scale represented by whole numbers -- no considerations of tonality), rather than actual frequency for computational simplicity. Given this structure, a simple correlation coefficient[47] could then be use to compare two such pitch contours from different parts of the melody. For this assessment of parallelism, only adjacent sequences of notes were compared (i.e., the AA of musical patterns of the AABC form). Figure 4.6 provides some examples. The main point is that in this contour, a whole note would contribute four times as much information to the contour as a quarter note. Note that this type of parallelism will capture transposition, but not necessarily as a key transpositions (e.g., c-e-g in the key of C major would be MIDI notes 60-64-67 while c-e♭-g in the key of C minor would be MIDI notes 60-63-67, which are not perfectly parallel structures even though both represent the tonic triad of their respective keys). This type of parallelism would not necessarily capture within-key transpositions (e.g., c-d-e [MIDI notes 60-62-64] is not perfectly parallel with f-g-a [MIDI notes 65-67-69] even though both represent three adjacent notes within the key of C major). It captures transpositions on a chromatic scale.

---

[46] Admittedly, in MIDI tracks, the number of tics per whole note is usually much higher, but in fact any equitemporal interval that is less than the smallest note duration (or the "smallest difference that matters") will suffice for this type of analysis.

[47] The correlation is an accepted measure of pattern similarity, but within this work, it is far from ideal. However, there are not many options. Discussion of measures of similarity is delayed until latter in order to maintain focus and continuity.

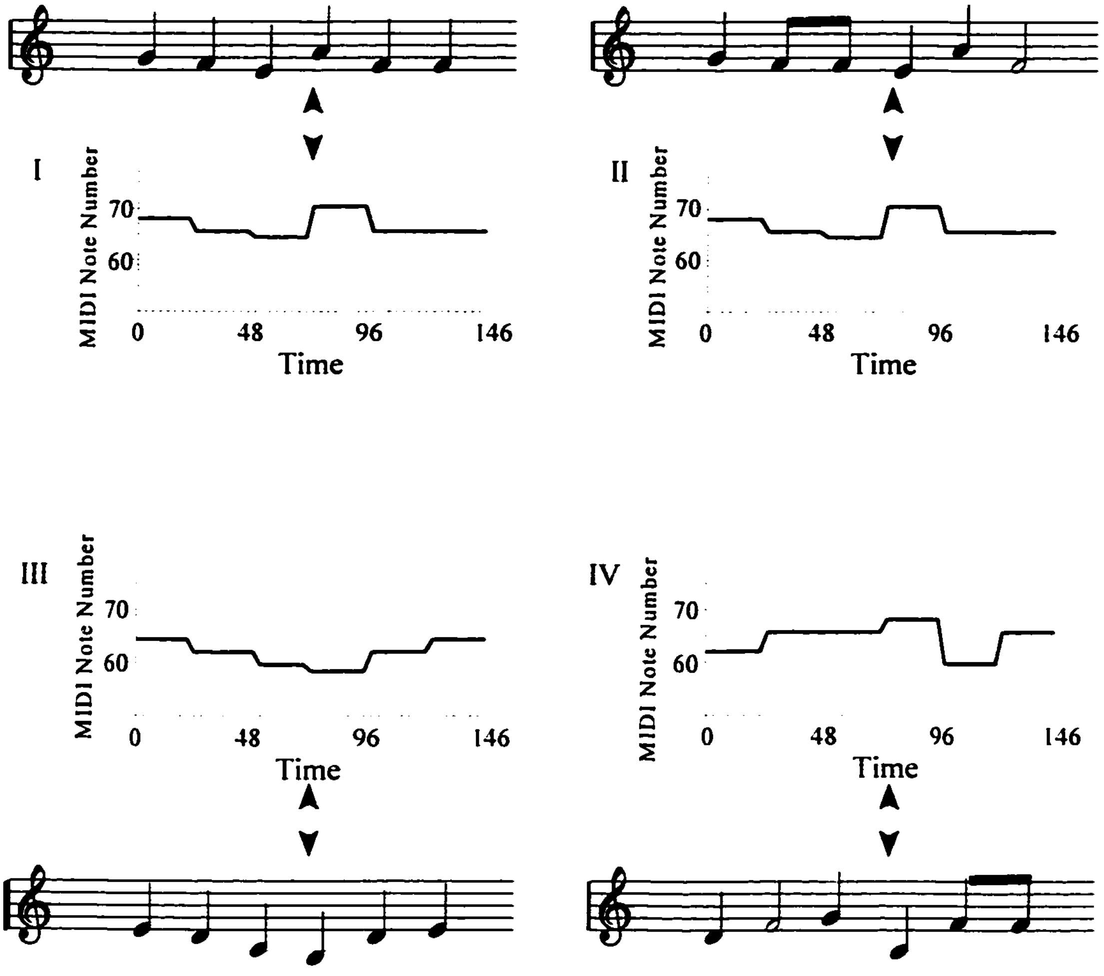


Figure 4.6: Examples of Pitch Pattern Parallelism, Type A. In this type, the duration of individual note events is a factor, but the internote intervals are not (see Time Pattern Parallelisms A & B, and to a lesser extent, Pitch Pattern Parallelism B for consideration of note onsets). I and II have the same pitch contours despite different temporal patterns, I and III have different pitch pattern despite the same temporal patterns, and IV is different from the rest. Also note that the initial part of I is not the same as the initial part of III due to the notes involved (the semitone steps are different). Note that pitch height is contained as an approximation to a linear scale within the score.

This analysis raised a second issue: How many notes should be used in the comparison (i.e., in the construction of the array of pitch values)?. This creates yet another question: Can the count of notes (regardless of duration) be considered a fair basis for comparison? When considering parallelism based, in part, on note duration, it seems that simply comparing 5 notes (of undetermined duration) to another set of 5 notes (of undetermined duration) would not capture parallelism. A sequence that consists of four whole notes has four times the duration as a sequence of four quarter notes. Conversely, a sequence of two whole notes has the same duration as a sequence of eight quarter notes. The first example has two sequences with the same number of notes. The second example has two sequences of equal duration. To circumvent the number of notes issue, sequences having the same number of beats (equal durations) were compared in this version of Pitch Pattern Parallelism. The equal number of notes situation is considered in Pitch Pattern Parallelism B

The third question was how many beats should be compared (the beat-span of the comparison). From the perspective of the subject, the comparison of two sequences for parallelism requires some note-by-note comparison that is performed in real time. Hence, it is implied that literal versions of both sequences have to be stored in short term memory (though, it is arguable that only one sequence -- the current sequence -- need be in short term memory, while the other is in long term memory). This in turn, implies that these sequences cannot be too long. It must be remembered that this comparison occurs before unitization has proceeded (i.e., it is a part of the basis for creating the current unit) so short term storage must be allocated to the individual notes. Somewhat arbitrarily, a limit of ten beats was chosen. If one considers the time unit indicated by the time signature (e.g., 4/4 2/4, 3/8) then the beat note of a time signature can be considered the modal note duration of a melody (i.e., the most commonly occurring duration). Hence, ten beats encompasses the traditional short term memory limit of 7±2 items (i.e., 10 beats is 10 notes). Since many melodies contain notes that have durations longer than the beat note, the actual number of notes stored could be less than ten (i.e., 10 beats could be only 2 notes), but some short term capacity must be used for the details of comparison.

Conversely, sections of melodies that contain many notes having durations shorter than the beat note will have more than ten notes within a ten beat span (i.e., 10 beats could be 20 notes). However, ten beats was considered as a good compromise for this exploratory work.

There was one final consideration. Within the ten beat limit, the best example of parallelism may occur for two three-beat sequences or for some other number of beats. Hence, in the analysis, parallelism was compared for all values of beat span up to the ten beat limit, and the best parallelism within that beat span was used. That is, the similarity of two adjacent sequences of one beat was assessed. Then, the similarity of two adjacent sequences of two beats was assessed. Then, the similarity of two adjacent sequences of three beats was assessed. This process was continued until the ten beat limit had been reached. The beat span having the highest degree of similarity (assessed as a correlation) was used as the measure of parallelism.

To conduct the analysis, at each note of the melody, a subsequence spanning X beats was taken (where X = 1 to 10), projected backwards in time (i.e., the X beats preceding the current note; see Figure 4.7). If this span of beats ended in the middle of a note, the span of beats was rounded up to the closest note onset (move further back in time to the beginning of the note), so that the subsequence would contain only complete notes. Next, a second subsequence of notes was taken from those notes that immediately preceded the current span. This subsequence also spanned X beats, rounded up as necessary. The two subsequences had to be adjacent. These two sequences were converted into arrays of note values on the aforementioned time base, and the correlation between the two arrays was assessed, moving back in time (note that because of rounding, the two subsequences could have slightly different lengths). In this analysis, it important to realize that the correlation coefficient is not affected by the choice of time base (i.e., one would produce the same correlation using a basic duration of 1/200 of a whole note, or 1/1000 of a whole note), although the significance of the correlation, if computed, would be. All that matters is that the time base be less than the smallest time unit of importance.

This process was repeated for all beat spans from one to ten, producing 10 correlations between adjacent spans. The beat span producing the highest positive correlation was used as the measure of parallelism (note that a negative correlation would would imply a descending and ascending comparison).

In this analysis, there was the additional restriction that there had to be some potential for a boundary at the point between the two spans and at the current note (based on Rules 2 and/or 3; see Figure 4.7). This restriction was imposed to reflect the notion that parallelism was construed to build on the basic boundaries identified by other rules. That is, parallelism could not create a boundary (this seems consistent with Lerdahl & Jackendoff's [1983] conception of parallelism): Parallelism could only strengthen a potential boundary. There had to be a potential boundary at the end (or beginning, depending one's perspective) of each subsequence.

There was one final restriction. Beats spans had to contain at least five notes in total for a correlation to be computed (e.g., 3 versus 2, or 3 versus 3, etc): neither sequence could contain a single note (which would result in a correlation of r=0.0 anyway) and both sequences could not contain only two notes (which would always produce a correlation of r=1.0). Observe that it is, in principle, possible for a beat span of only one beat to contain sufficient notes for a correlation to be computed. As such, the analysis really used beat span (time) constrained by a minimum number of notes (but no maximum number of notes). The full algorithm is presented in Appendix: Programs.

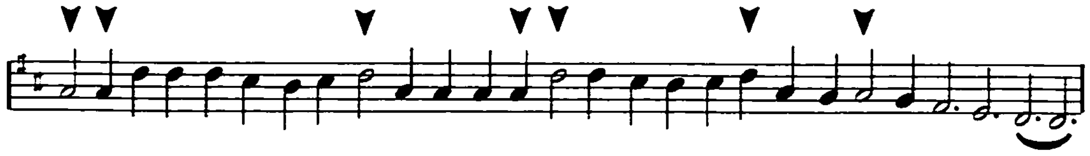
Notes having a potential for a boundary (see Figure 4.4)

B span of 1 (1 quarter note)	1 2	Spans 1 & 2 have only 1 note Span 1 does not end on a boundary
Beat spans of 2 (2 quarter notes)	1 2	Spans 1 & 2 have only 2 notes. Span 1 does not end on a boundary
Beat span of 3 (3 quarter notes)	1 2	Span 1 does not end on a boundary
Beat span of 4 (4 quarter notes)	1 2	Span 1 does not end on a boundary
Beat span of 5 (5 quarter notes)	1 2	Span 1 rounded up to include full notes Span 2 rounded up to include full notes
Beat spans of 6 (6 quarter notes)	1 2	Note that this is the same as Five beat spans (by chance)
Beat span of 7 (7 quarter notes)	1 2	

Figure 4.7: Example of the assessment of Pitch Pattern Parallelism A, at one note of the melody, for beat-spans of 1 to 7. Note that beat-spans of 1, 2, 3,& 4 would not produce values for parallelism for reasons cited, but beat-spans of 5, 6 & 7 would. Also note that because of rounding to included complete notes, beat-spans of 5 and 6 were the same. This type of analysis was conducted at each note of the melody for beat spans of 1 to 10 (8, 9 and 10 are not shown here).

To recap, at each note of the melody[48], adjacent subsequences were extracted. The subsequences were composed of notes that preceded the note being analyzed. The subsequences contained the same number of beats, but may have contained different numbers of notes. Parallelism was assessed for the pair of subsequences. The comparison was computed at each note of the melody using subsequences that contained from 1 to 10 beats. One might question the need to use a maximum span of 10 beats (i.e., would a maximum of 7 be sufficient?). For example, the melody "Three Blind Mice" only contains 32 bars or 96 beats in total. Hence, 10 beats is a fair proportion of the entire melody, and two 10-beat subsequences actually covers about 20% of the entire tune. As a check on the utility of this maximum value, the analysis of parallelism was repeated using a maximum beat span of 2 beats (subsequences compared with only 1 and 2 beats each), 3 beats (subsequences compared with 1, 2 & 3 beats each), 4 beats (subsequences compared with 1, 2, 3 & 4 beats each) and so on up to 10 beats (subsequences compared with 1 to 10 beats each). The goal was to determine which maximum beat span would produce the greatest amount of parallelism. These analyses were designed to determine whether or not 10 beats was necessary to capture all the parallelism that was evident in the melody (a lower maximum span is potentially more efficient for the analysis of a large body of work). It must be remembered that even though the maximum was set to one particular value, the actual best span (highest parallelism) at each note of the melody could be at any span from 1 to 10 beats.

To simplify comparisons, interstimulus durations were not included (i.e., there were no gaps between notes). Also, the question of how to treat rests within the construction of a pitch contour was not addressed. In this work, the issue of rests is moot since none of the melodies contained rests.

For "Three Blind Mice", the value of the best example of Pitch Pattern Parallelism A (Rule 6a) at each point in the melody is presented in Figure 4.8 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and

---

[48] This could have been done at only those points that had some potential for a boundary, but it was easier to create the algorithm in this manner.

3 (second Panel), which are necessary for Pitch Pattern Parallelism A to operate.

Table 4.8 presents a the simple correlations and the corresponding SS_{error} for the fit of Pitch Pattern Parallelism A (Rule 6a) to the empirical data for all beat spans assessed. Note that the correlation was highest and the error the lowest for beat spans of ten beats. This implied that a maximum beat span of 10 was necessary to capture the Pitch Pattern Parallelism A that is available within the melody. A maximum of 10 beats was needed because of the opening 14 notes of the melody. Only beat spans of ten captured the parallelism within that section (i.e., the first three notes consist of 12 beats, which was captured by a 10 beat span rounded up to the previous note onset: a 3/4 time signature). Also note that when the maximum beat span was only four or three, there were no examples of parallelism (hence the correlation could not be computed.

Table 4.8: Beat Span Analysis for "Three Blind Mice", Pitch Pattern Parallelism A.

Beat Span	Empirical Data	SS_{error}
10	.507**	3.376
9	-.120	3.540
8	-.120	3.540
7	-.107	3.540
6	-.102	3.540
5	-.073	3.540
4	.	3.540
3	.	3.540

Notes: * $p < .05$
** $p < .01$

Since a maximum beat span of ten proved to be the best, only those results are presented. It should be noted that even though the maximum beat span was ten beats, the best examples of parallelism were not always at the ten beat limit (i.e., a non-analytical appraisal of the data showed that the best parallelism occurred at all beat spans from 5 to 10 beats).

In the third panel of Figure 4.8 (Rule 6a), observe that there is no parallelism at Note 3 because one cannot have parallelism until one has a previous sequence to compare to. Also observe that the parallelism at Notes 6 and 10 are perfect (for spans of 10 beats), but that the parallelism at Note 14 is not perfect (for spans of 10 beats) because the last notes of the two sequences in the comparison are not the same (dotted half plus dotted half versus dotted half plus half). There is an additional parallelism at Note 15 (span of 10 beats) because Notes 7 through 10 are parallel with Notes 11 through 15. This is, in fact, the same parallelism as detected at Note 14: The addition of a single note at Note 15 does not change the structure of the second sequence (Notes 11 through 14 + Note 15) enough to destroy their parallelism (the rounding up of the number of beats to the onset of the note is also a factor here). In addition, Rule 3a proposed a boundary at Note 15 (a prerequisite for parallelism) even though few people did. By comparison, there is no similar parallelism at Note 11 because Note 11 does not have a low level boundary associated with it. The parallelism detected at Note 21 is for a span of six beats (17 through 21 verse 14 through 16). This does not look very parallel, and in fact the correlation is only r=0.56 (r^2=0.31). This implies that a cutoff based on a minimum value of r or r^2 might be fruitful (such mechanisms were not used here so that the data could fully express itself). The same comment applies to the parallelism detected at Notes 31, 36 and 41.

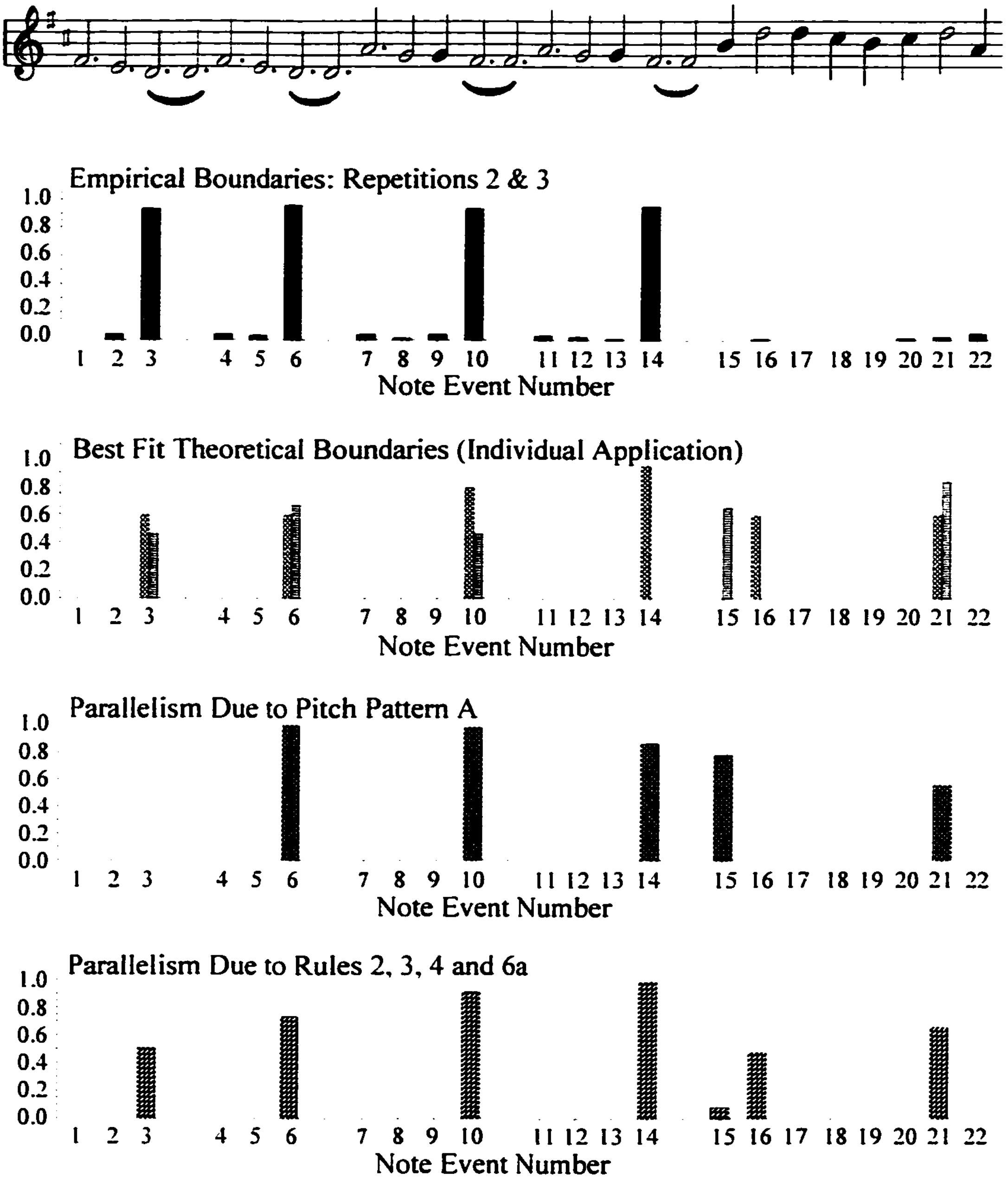


Figure 4.8a: The first part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and 3a and Rule 6a (parallelism due to Pitch Pattern A). This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**	■ **Rule 6a**
▨ **Rule 3a**	▦ **Rules 2b, 3a 4, 6 and interactions**

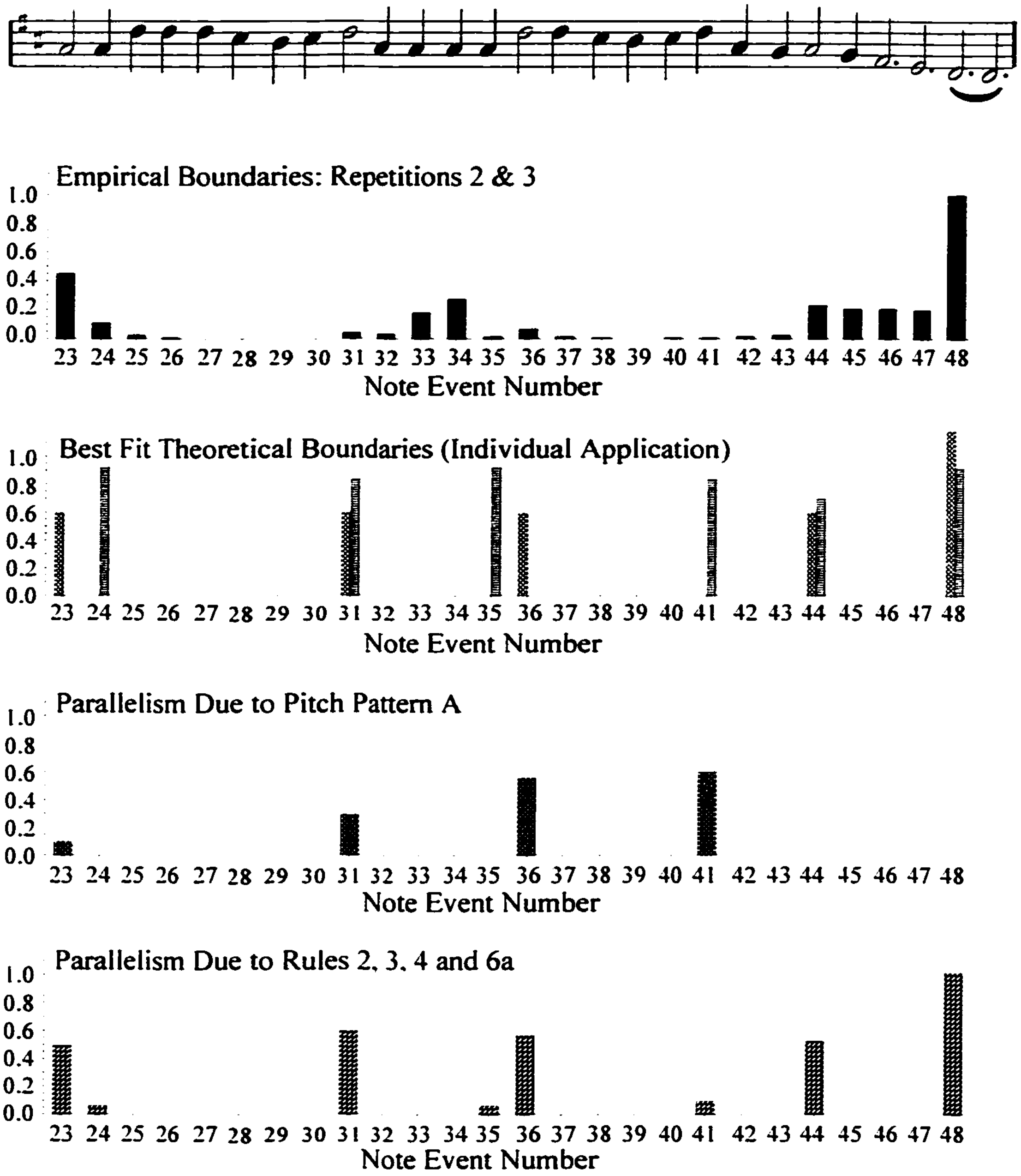


Figure 4.8b: The second part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and 3a and Rule 6a (parallelism due to Pitch Pattern A). This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▩ **Rule 2a**	■ **Rule 6a**
▤ **Rule 3a**	▦ **Rules 2b, 3a 4, 6 and interactions**

In the analysis, Rule 6a (Pitch Pattern Parallelism A) was entered in conjunction with Rules 2 and 3. The logic is that Rule 6a is built on putative boundaries identified by Rules 2 and 3, and as such can only exist where Rules 2 and 3 already exist. Rule 6a is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6a with Rules 2b and 3a were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a and 6a, and then the complete model containing all terms was examined. Table 4.9 provides the details. In Table 4.9, the previous fit for the combination of Rules 2 and 3 (i.e., Rule 4) is included for comparison.

In this analysis, one can see that parallelism did add to the predictability of the model. That is, the different terms, Rule 6a (parallelism due to Pitch Pattern Parallelism A) or the interactions of Rule 6a with Rules 2b and 3a, reduced the error below the value of 3.54 that was achieved by Rule 4 (Intensification; the optimal combination of Rules 2a and 3d) to 3.27 (a 7.6% improvement). Parallelism helped by allowing the constants preceding Rules 2a and 3a to be reduced in magnitude, which helped to decrease false alarms.

The interactions of Rule 6a with Rules 2b and 3a were also somewhat helpful. Note that the interaction of Rules 2b and 3a was not included in the model, because it had been previously found to be ineffectual. The higher order interactions were also not included by design.

As can be seen in Table 4.9, although Rule 6a or its interactions were helpful to some degree, deciding which combination is the best is difficult. The combination of Rules 2, 3, 6a and the interactions of Rules 2 and 3 with 6a resulted in a 7.63% improvement overall (from 3.54 to 3.27), but the combination of Rules 2, 3 and 6a alone (no interactions) resulted in a 4.63% improvement overall. It is difficult to decide whether the extra interaction terms warrant the small improvement. It is interesting to note that the combination of Rules 2b, 3a, 6a with the interaction of 6a and 2b actually reduced the contribution of 6a to 0.0. The implication is that the interaction of 6a and 2b explains the same variance as 6a alone (remember that Rule 6a was constrained to only occur where

Rules 2 and/or 3 had occurred): It would seem that the interaction is the better choice. given the overall error terms (the interaction is also closer to the notion of parallelism acting in a manner to boost boundaries defined by Rule 2 and/or 3).

Table 4.9: The Combination of Rules 2. 3 (Rule 4) and 6a for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.131*2b+0.072*3a	3.540
2b+3a+6a	15	b = 1.023*2b+0.034*3a+0.165*6a	3.376
2b+3a+6a+2b*6a	15	b = 0.929*2b+0.084*3a+0.000*6a+0.327*2b*6a	3.277
2b+3a+6a+3a*6a	15	b = 1.059*2b+0.010*3a+0.072*6a+1.995*3a*6a	3.362
all	15	b = 0.951*2b+0.061*3a+0.000*6a + 0.276*2b*6a+0.898*3a*6a	3.270

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	5	0.828	0.900	10	1.991	3.123
2b+3a+6a	4	0.820	0.809	11	1.835	2.945
2b+3a+6a+2b*6a	4	0.847	0.711	11	1.709	2.738
2b+3a+6a+3a*6a	6	0.776	0.850	9	1.865	2.967
all	4	0.827	0.740	11	1.723	2.750

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Given this, it was decided that the best model is that which included all the terms. This is the model that is plotted in Figure 4.7 in the bottom panel. Note that regardless of this analysis, the application of the theory still misses some relatively minor boundaries near Note 34.

In addition, the simple correlations between the rules are shown in Table 4.10. Note that there are high correlations between Rule 6a (Pitch Pattern Parallelism A) and Rules 2b (Attack-Point) and 3a (Register). This is simple consequence of the analysis since Rule 6a was constrained to generate boundaries at only those locations that had low level rule applications. Note the high correlation between Rule 6a and the interaction of Rule 6a and 2b. This is why only one of the two terms was not necessary in the final analysis.

Table 4.10: Correlation Matrix for "Three Blind Mice", Pitch Pattern Parallelism A.

	Rule						
	2b	3a	4	6a	2b*6a	3a*6a	2b*3a
data	.770**	.340*	.766**	.507**	.656**	.295*	.595**
Rule 2b		.466*	.997**	.529**	.645**	.312*	.756**
Rule 3a			.533**	.428**	.222	.600**	.662**
Rule 4				.543**	.637**	.341*	.780**
Rule 6a					.849**	.831**	.277
Rule 2b*6a						.522**	.296*
Rule 3a*6a							.361*

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6a was helpful for predicting boundaries. However, the effect was not a dramatic improvement on the predictions afforded by the best combination of Rules

2b and 3a (particularly 2b). Where it failed most strikingly was at Note 15. This failure was likely due to the difference between the human system and the algorithm. That is, the human system, processing music in real time, placed a boundary at Note 14. Having placed a boundary at Note 14, there would be no point in placing a boundary at Note 15. This can be seen as an aspect of Rule 1 -- groups must contain more than 1 note event -- but it is more likely that, having created a unit that ends on Note 14, the subject is disinclined to look for parallelism that crosses a unit boundary. Note that the potential presence of overlapping parallel structures does not contradict the theory of Lerdahl and Jackendoff (1983). They would simple argue that one or the other is the preferential structure. In this case, it seems clear that subjects prefer to have a boundary at Note 14 rather than at Note 15. Note 14 has a very strong instance of Rule 2b. Note 15 has a fairly strong instance of Rule 3a. Both Notes 14 and 15 have some parallelism, though parallelism is stronger at Note 14. At this point, one cannot use the algorithm to explain why subjects chose Note 14 over Note 15, but the algorithm does show that Note 14 has a stronger instance of a stronger rule with a stronger reenforcement by parallel structure.

Pitch Pattern Parallelism B (Rule 6b)

In Pitch Pattern Parallelism A, the pitch contour contains information of both pitch height and time. In Pitch Pattern Parallelism B, considerations of the temporal durations of pitches were removed. Hence, Pitch Pattern Parallelism B represents the note-to-note pattern as the contour without regard to pitch duration (this is equivalent to a pitch pattern comparison that depends on note onsets).

The basic method was the same as in Pitch Pattern Parallelism A. That is, adjacent subsequences of notes were represented as arrays of MIDI note values. Instead of using beat spans, however, the actual number of notes was used. By this method, all note durations are given equal weight. These two arrays were compared using a correlation coefficient. Note spans ranging from one to ten notes were used in the comparison (note spans from 1 to 10 are comparable to the previous beat spans from 1 to 10) and the highest positive correlation was used as a measure of the similarity (see Figure 4.10).

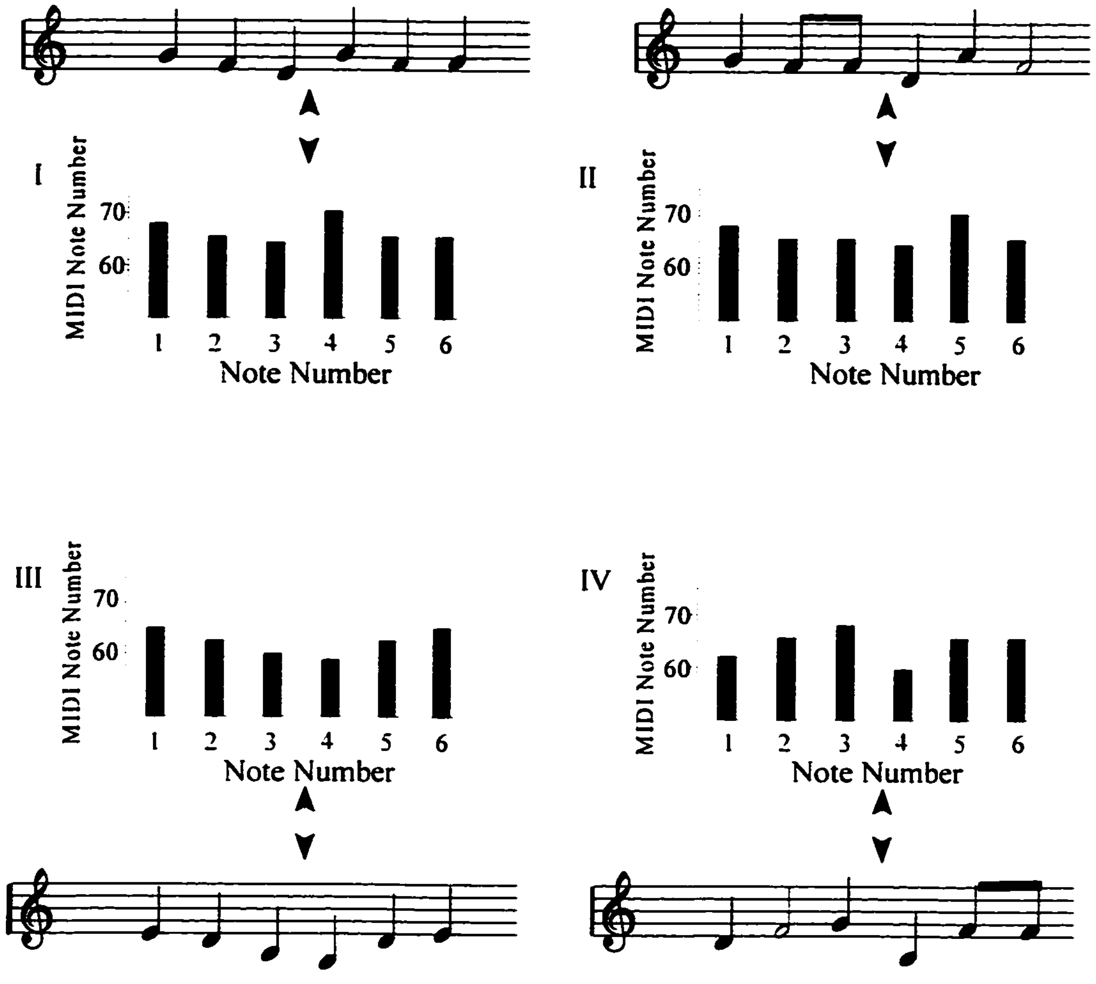


Figure 4.9: Examples of Pitch Pattern Parallelism, Type B. In this type, the pitch patterns are compared on the basis of note onsets, ignore any contribution of duration. Note that in contrast to Pitch Pattern Parallelism A, sequences I and II are not the same.

In this analysis, adjacent notes having the same frequency were retained as two separate notes. As before, sequences had to end (or begin, depending on one's perspective) on boundaries, had to contain at least two notes in each subsequence, had to contain at least five notes in total (the correlation between two pairs of notes would always be 1.0) and rests were skipped in the comparison.

For "Three Blind Mice", the value of the best example of Pitch Pattern Parallelism B (Rule 6b) at each point in the melody is presented in Figure 4.10 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel), which are necessary for Pitch Pattern Parallelism B to operate. For this melody, the best match between the assessed parallelism occurred for a note span of four notes. This value tended to capture the same information as the beat span of ten in Pitch Pattern Parallelism A (Rule 6b) because it covered the strong parallelism at the beginning of the melody. Table 4.11 presents a the simple correlations and the corresponding SS_{error} for the fit of Pitch Pattern Parallelism B (Rule 6b) to the empirical data for all note spans assessed. Note that the correlation was highest and the error the lowest for note spans of four notes. In the actual analysis, the note span that maximized the reduction in total error was used. Observe that the note span was not the maximum tested (for this reason, it was thought that the previous beat span of 10 in Pitch Pattern Parallelism A did represent the best example, and not a premature cutoff).

Table 4.11: Note Span Analysis for "Three Blind Mice", Pitch Pattern Parallelism B.

Note Span	Empirical Data	SS_{error}
10	.513**	3.540
9	.583**	3.540
8	.563**	3.500
7	.563**	3.500
6	.584**	3.457
5	.595**	3.451
4	.719**	3.204
3	.409**	3.250

Notes: * $p < .05$
** $p < .01$

It should be noted that even though the maximum note span was four notes, the best examples of parallelism were not always at the four-note limit.

In the third panel of Figure 4.10 (Rule 6b), observe that there is no parallelism at Note 3 because, as in Pitch Pattern Parallelism A (Rule 6a), one cannot have parallelism until one has a previous sequence to compare to. Also observe that the parallelism at Notes 4 and 14 are perfect (span of 3 and 4 notes respectively), but that there is no parallelism at Note 10. This is because when counting just notes, Notes 8 through 10 are not parallel with Notes 4 through 6, or conversely, Notes 7 through 10 are not parallel with Notes 3 through 6 (the 2 subsequences must have the same number of notes, must be adjacent and both must end on a potential boundary defined by the lower level rules). In fact, although it is not obvious to the eye, the correlation of note values for Notes 3 through 6 and Notes 7 through 10 is negative (r=-0.21).

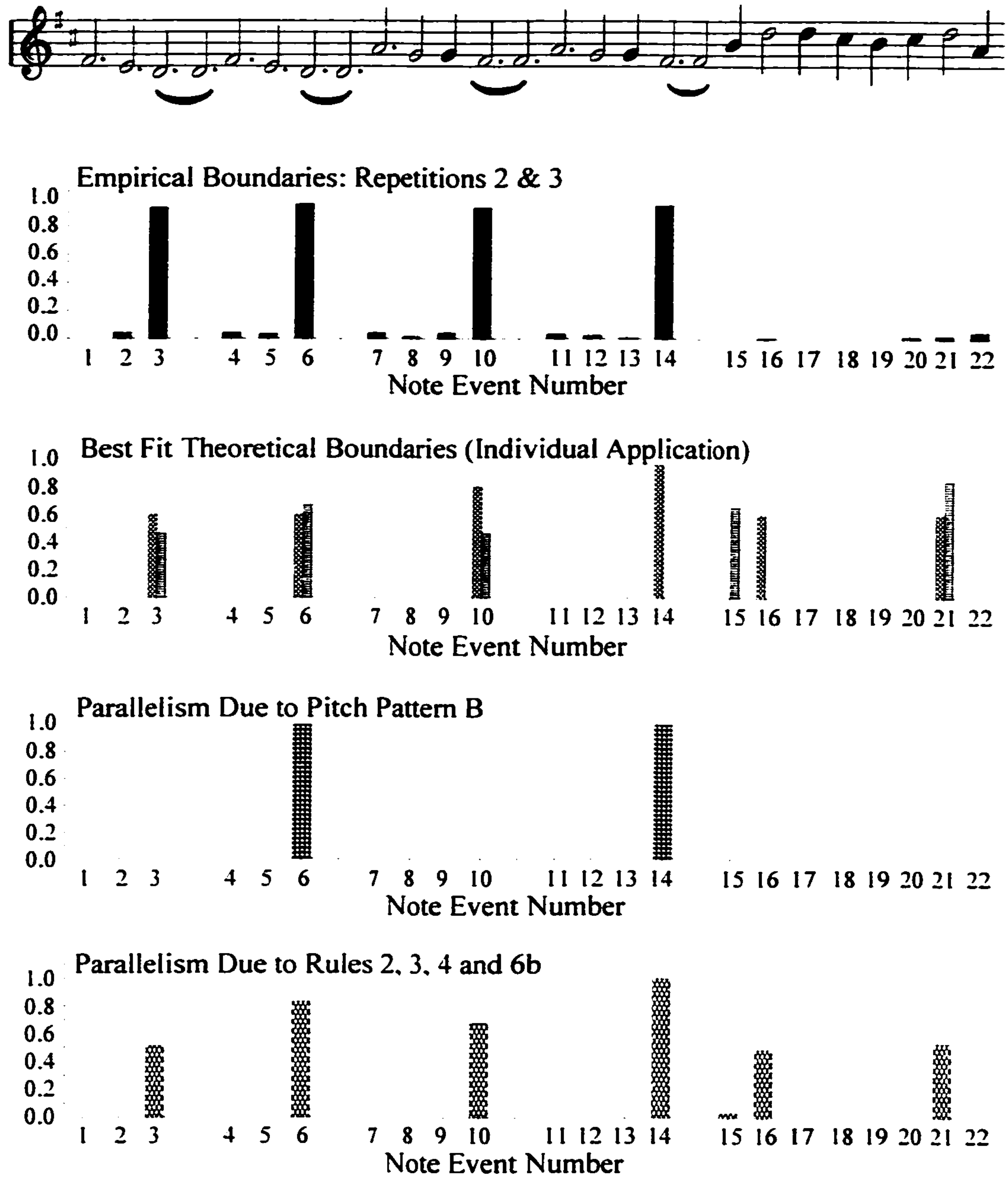


Figure 4.10a: The first part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a. and Rule 6b (parallelism due to Pitch Pattern B).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**

▨ **Rule 3a**

▨ **Rule 6b**

▨ **Rules 2b, 3a 4, 6 and interactions**

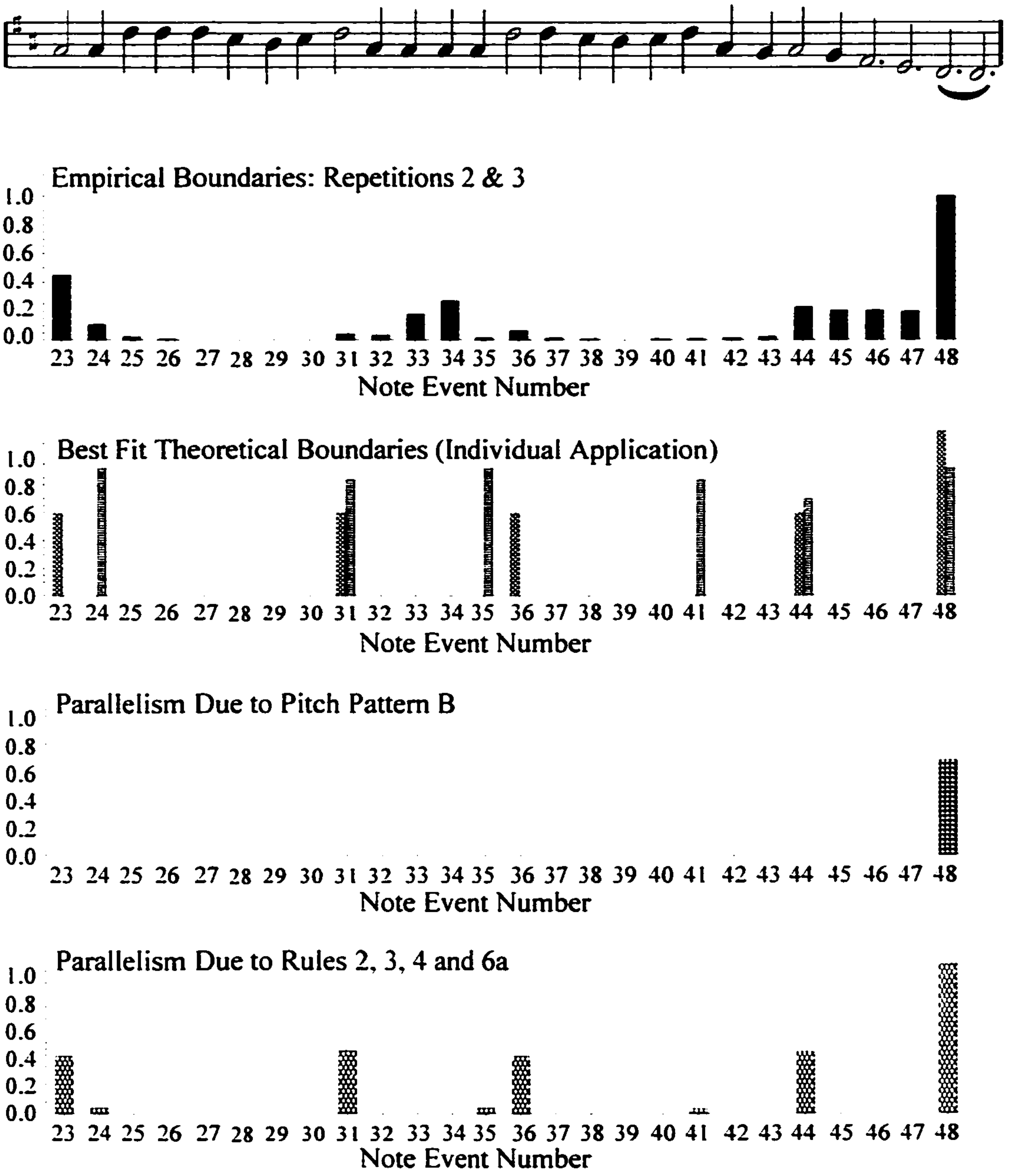


Figure 4.10b: The second part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a. and Rule 6b (parallelism due to Pitch Pattern B).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**
▨ **Rule 3a**
▨ **Rule 6b**
▨ **Rules 2b, 3a 4, 6 and interactions**

There is no simple way to capture the parallelism at Note 10 using this system of pitch pattern parallelism (i.e., parallelism for two adjacent sequences based on a fixed number of notes; this type of parallelism will be captured when the AA parallelism of AABC restriction is relaxed so that the AA parallelism of ABAC can be seen). This version of parallelism does not capture much detail throughout the remainder of the melody (cf., Pitch Pattern Parallelism A). It is true that the constraints are a part of the problem (e.g., there must be a boundary at the end of each subsequence based on Rules 2 and 3; the subsequences must be adjacent; the subsequences must be of equal length), but the constraints are not the main problem. The rule simply does not capture what we think of when we speak of parallelism in a general sense. The rule captures perfect parallelism quite well, but it performs much more poorly as the parallelism drifts away from perfect parallelism, when the note onsets do not correspond well to the temporal pattern of sound.

As in the analysis of Pitch Pattern Parallelism A (Rule 6a), the analysis of Pitch Pattern Parallelism B (Rule 6b) considered this rule in conjunction with Rules 2 and 3, because Rule 6b is built on putative boundaries identified by Rules 2 and 3. Like Rule 6a, Rule 6b is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6b with Rules 2b and 3a were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a and 6b, and then model containing all terms was examined. Table 4.12 provides the details.

In this analysis, one can see that this form of parallelism appears to add a great deal to predictability (Rule 6b or the interactions of Rule 6b with Rules 2b and 3a, reduced the error from 3.54 to 3.19: a 9.77% improvement) but this predictability was achieved by actually increasing the misses while decreasing the false alarms. This is evident in Figure 4.10: If the rule generates few boundaries, then it cannot generate many false alarms. In addition, if the rule is constrained to produce boundaries at only those values that have high potentials for boundaries (i.e., at Rules 2 and/or 3) then it cannot make many false alarms. It however, will generate misses. Hence, overall the Rule 6b helps, but on detailed analysis, this is not really the kind of help that is sought.

Table 4.12: The Combination of Rules 2, 3 (Rule 4) and 6b for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.131*2b + 0.072*3a	3.540
2b+3a+6b	15	b = 0.963*2b+0.061*3a+0.284*6b	3.204
2b+3a+6b+2b*6b	15	b = 0.963*2b+0.061*3a+0.284*6b+0.000*2b*6b	3.204
2b+3a+6b+3a*6b	15	b = 0.973*2b+0.049*3a+0.221*6b+1.538*3a*6b	3.194
all	15	b = 0.973*2b+0.049*3a+0.221*6b + 0.000*2b*6b+1.537*3a*6b	3.194

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	5	0.828	0.900	10	1.991	3.123
2b+3a+6b	4	0.929	0.881	11	1.554	2.676
2b+3a+6b+2b*6b	4	0.929	0.881	11	1.554	2.676
2b+3a+6b+3a*6b	4	0.915	0.871	11	1.558	2.631
all	4	0.915	0.871	11	1.558	2.631

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|\text{Deviation}| = \Sigma(X_i - X'_i)$

The interactions of Rule 6b with Rules 2b and 3a were also somewhat helpful. Note that the interaction of Rules 2b and 3a was not included in the model, because it had been previously found to be ineffectual. The higher order interactions were also not included by design. As can be seen in Table 4.12, the interaction of Rule 6b and 3a is more helpful than that of Rule 6b and 2b. This is interesting because it implies that the rule acts to boost the weaker of the two level rules. For comparison with the previous Rule 6a (Pitch Pattern Parallelism A), Figure 4.10 (bottom panel) provides the best fit combination of Rules 2, 3, 6b and their interactions.

Finally, the simple correlations between the rules are shown in Table 4.13. Note that there are high correlations between Rule 6b (Pitch Parallelism B) and Rules 2b (Attack-Point) and 3a (Register). This is simple consequence of the analysis since Rule 6b was constrained to generate boundaries at only those locations that had low level rule applications. The fact that both Rule 6b and the interaction of Rule 6b with 2b are not necessary is self evident within the table.

Table 4.13: Correlation Matrix for "Three Blind Mice", Pitch Pattern Parallelism B.

	Rule						
	2b	3a	4	6b	2b*6b	3a*6b	2b*3a
data	.770**	340*	.766**	.719**	.720**	.599**	.595**
Rule 2b		.466**	.997**	.593**	.645**	.490**	.756**
Rule 3a			.533**	.249	.246	.399**	.661**
Rule 4				.589**	.638**	.504**	.780**
Rule 6b					.964**	.748**	.440**
Rule 2b*6b						.699**	.503**
Rule 3a*6b							.682**

Notes: * p < .05
** p < .01

Generally, Rule 6b was helpful for predicting boundaries. As with Rule 6a, the effect was not a dramatic improvement on the predictions afforded by the best combination of Rules 2b and 3a (particularly 2b). However, in contrast to Rule 6a, this rule did not generate many potential boundaries and actually increased the miss rate.

Time Pattern Parallelism A (Rule 6c)

Time Pattern Parallelism Analysis is more difficult to address, because time pattern similarity within the notion of rhythm as distinct from meter has not been dealt with as often or as explicitly in the literature. The basic methodology was the same as that of the previous analyses of pitch pattern. The only difference was that in place of an assessment of pitch pattern similarity, there was an assessment of time pattern similarity.

To assess time pattern, one can argue that it is the breaks between notes that is important. That is, two sequences that contained breaks (i.e., the offset of the one note to the onset of the next note) between notes at the same (relative) times, would be considered to have the same time pattern. If two sequences contained all the same breaks between notes but one sequence had one extra break (i.e., Sequence 2, relative to Sequence 1, had divided one note into two smaller durations), then those two sequences would be considered very similar. It would not likely matter too much if the extra break had created two subdivisions of equal or unequal length (i.e., dividing a half note into two quarters would be approximately equal to dividing a half note into a dotted-quarter and an eighth). Two sequences that differed at two locations would likely not sound as similar as the previous (e.g., half-half-quarter compared to half-quarter-half). From these general notions, an idea of temporal similarity arose based on the timing of breaks between notes. Figure 4.11 provides examples of this notion.

In this analysis, the time contour was represented as a string of note break points spaced at equitemporal intervals (i.e., a one dimensional array). The equitemporal intervals used were 1/96 the value of the whole note. That is, in the time based array, the points where one note ended and the next note began were represented as a binary value (i.e, a 1.0 represented a break between notes, a 0.0 represented the time within a note).

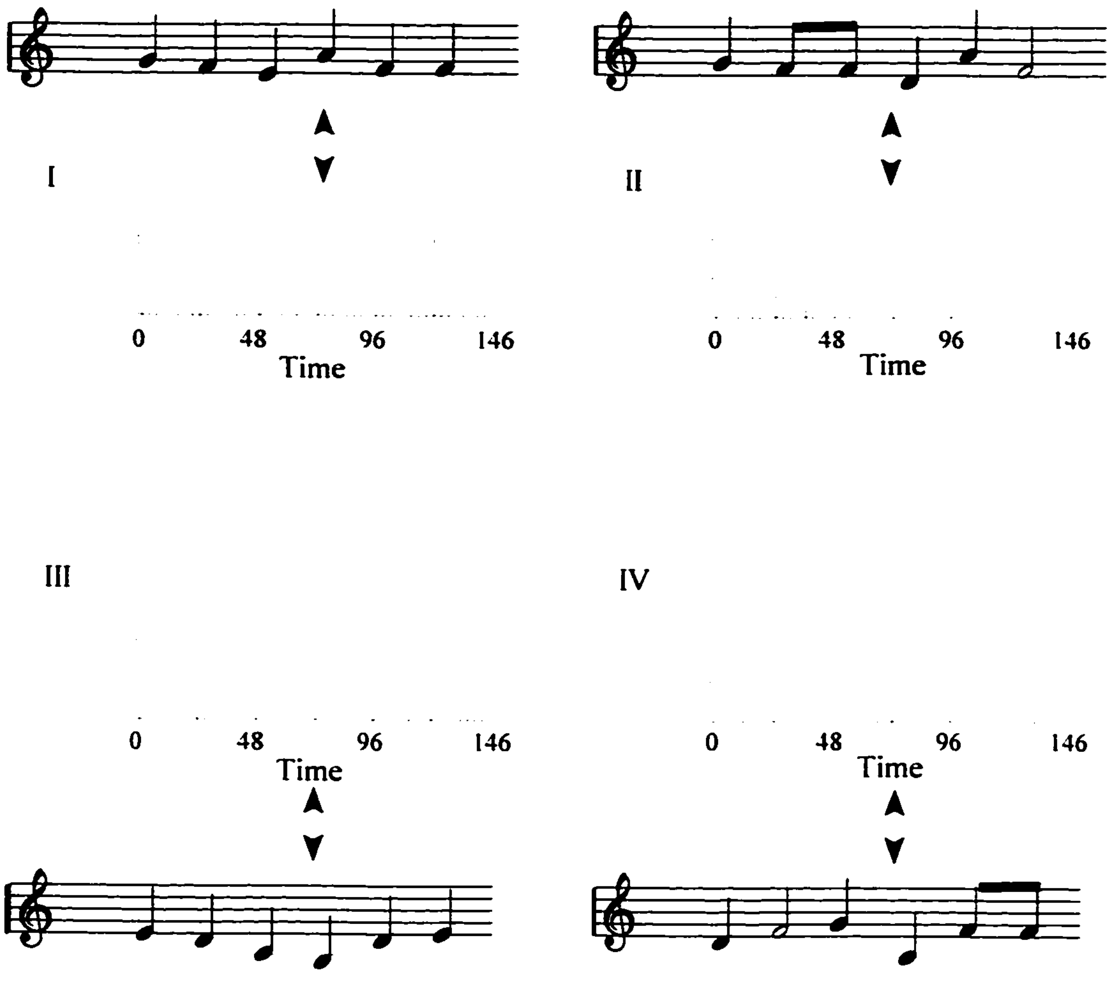


Figure 4.11: Examples of Time Pattern Parallelism, Type A. In this type, the temporal locations of the gaps between note events are tagged and compared on an absolute time scale. Note that the score does not represent time linearly (i.e., a quarter note uses as much horizontal space as a 16th note). The vertical scale is irrelevant.

Using this coding, a Jaccard binary correlation (also known as the similarity ratio: See Table 3.8) could be used to compare two sequences. This measure has the advantage that it is insensitive to the time base used for the assessment as long as the time base is fine enough to delineate onset differences that one wishes to call different (i.e., if onset differences smaller than the equivalent of a 64th note are not important, then the time base can be anything up to a 64th note). The Jaccard similarity measure could then be used in a manner like the correlation coefficient (see Figure 4.11).

Given this measure, the analysis of parallelism for Time Pattern Parallelism A (rule 6c) was conduction within "Three Blind Mice" using the same basic method as in Pitch Pattern Parallelism A. That is, two adjacent subsequences of notes were converted into onset arrays (temporal arrays marking the onsets of notes). These two arrays were then compared using a Jaccard similarity measure. Beat spans ranging from one to ten beats were used in the comparison (identical to the previous beat spans from 1 to 10 used in Pitch Pattern Parallelism A [Rule 6a], and unlike the note spans form 1 to 10 used in Pitch Pattern Parallelism B [Rule 6b]). The assessment of similarity was the Jaccard similarity coefficient returned by the beat span producing the highest value for the Jaccard similarity. As before, both subsequences had to end (or begin, depending on one's perspective) on boundaries, had to contain a total of at least five notes, and had to contain at least two notes in each subsequence (one note will only produce 1 onset time). However, rests were not skipped (since the duration of a rest is on the same quantitative scale as the duration of a note).

For "Three Blind Mice", the value of the best example of Time Pattern Parallelism A (Rule 6c) at each point in the melody is presented in Figure 4.12 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel), which are necessary for Time Pattern Parallelism A to operate. For this melody, the best match between the assessed parallelism (this form of pitch pattern parallelism) occurred for a beat span of ten notes (like that in Pitch Pattern Parallelism A [Rule 6a], and unlike that in Pitch Pattern Parallelism B [Rule 6b]). This beat span value captured the strong parallelism at the beginning of the melody. Table 4.14 presents a the simple correlations and the corresponding SS_{error} for the fit of Time Pattern Parallelism A (Rule 6c) to the empirical data for all beat spans assessed. Note that although the correlation was highest for beat spans of ten beats, there was no difference in the error values. In the actual analysis, the beat span that maximized the reduction in total error would normally be used, but in this case, it was the maximum correlation that was the determinate (this also kept this analysis comparable to Pitch Pattern Parallelism A). The beat span of ten beats provided the best correlation because of the opening 14 notes of the melody. Only a beat span of ten could capture the parallelism within that section. Also note that there were no examples of Time Pattern Parallelism A for beat spans of three.

Table 4.14: Beat Span Analysis for "Three Blind Mice", Time Pattern Parallelism A.

Beat Span	Empirical Data	SS_{error}
10	.374**	3.540
9	-.065	3.540
8	-.113	3.540
7	-.087	3.540
6	-.087	3.540
5	-.066	3.540
4	-.029	3.540
3	.	3.540

Notes: * $p < .05$
** $p < .01$

In the third panel of Figure 4.12 (Rule 6c), observe that there is no parallelism at Note 3 because one cannot have parallelism until one has a previous sequence to compare to. Also observe that the parallelism at Notes 4 and 14 are perfect (for spans of 10 beats), but that the parallelism at Note 10 is not perfect (for spans of 10 beats) because the last notes of the two sequences in the comparison are not the same (dotted half plus dotted half versus dotted half plus half). Note that this is slightly different from Pitch Pattern Parallelism A where Notes 4 and 10 had perfect parallelisms, and Note 14 did not. Time Pattern Parallelism A (Rule 6c) captured parallelism at Notes 35 and 44 that was not capture by Pitch Pattern Parallelism A (Rule 6a).

Generally, however, Time Pattern Parallelism A produced a similar pattern to Pitch Pattern Parallelism A: Both explicitly include the same aspects of time.

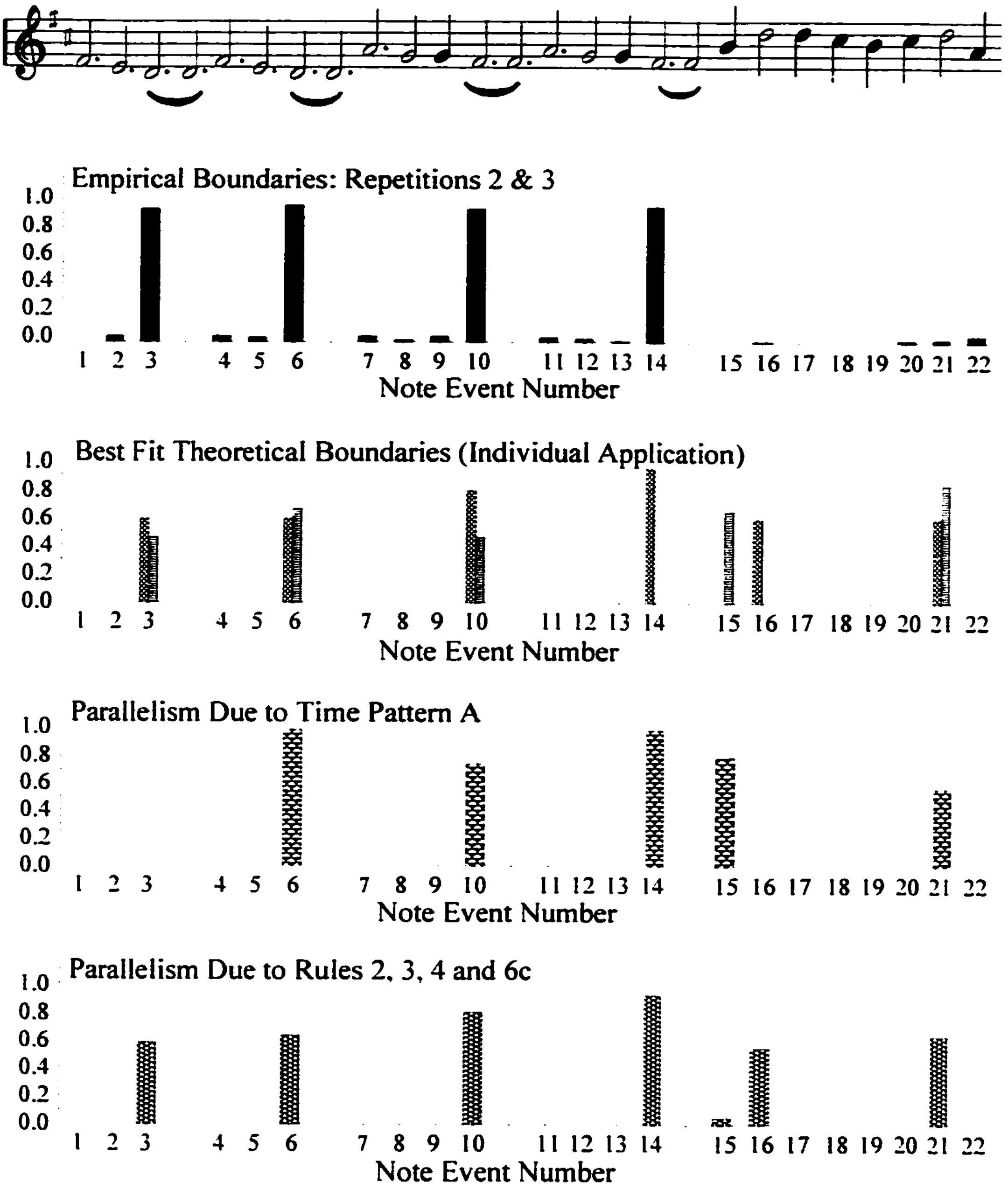


Figure 4.12a: The first part of the song used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a. and Rule 6c (parallelism due to Time Pattern A).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▩ **Rule 2a**	▦ **Rule 6c**
■ **Rule 3a**	▩ **Rules 2b, 3a 4, 6 and interactions**

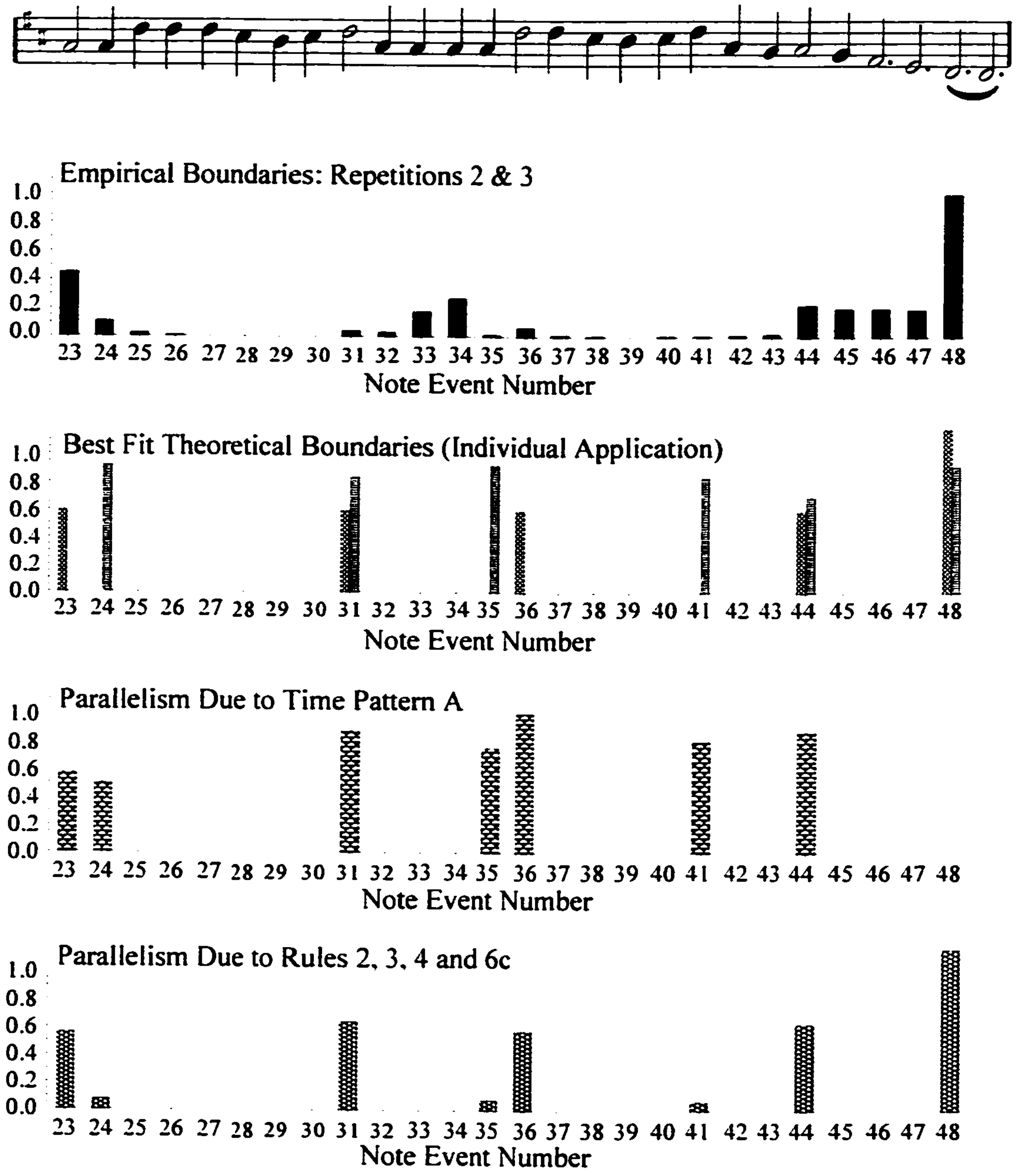


Figure 4.12b: The second part of the song used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a. and Rule 6c (parallelism due to Time Pattern A).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**	▨ **Rule 6c**
▨ **Rule 3a**	▨ **Rules 2b, 3a 4, 6 and interactions**

The analysis of Time Pattern Parallelism A (Rule 6c) considered this rule in conjunction with Rules 2 and 3, because Rule 6c is built on putative boundaries identified by Rules 2 and 3. Like Rule 6a, Rule 6c is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6c with Rules 2b and 3a were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a and 6c, and then model containing all terms was examined. Table 4.15 provides the details.

In this analysis, it is apparent that Time Pattern Parallelism A did not add predictability to the model beyond that which could be explained by the low level rules. The change was from 3.54 to 3.53 (an improvement of only 1.69%). This was surprising given the manifest similarity between Pitch Pattern Parallelism A and Time Pattern Parallelism A. The reason is that the inclusion of Rule 6c increases the false alarm rate because of the rather strong boundaries suggested at Notes 35 and 44 (notes that were not so pronounced in Pitch Pattern Parallelism A [Rule 6a]). Rule 6c did not help because the amount that it could help to reduce misses (Notes 6, 10 and 14) did not compensate for the increase in false alarms (Notes 35 and 44). This is because the low level rules were fairly accurate in terms of misses (Notes 6, 10 and 14) at the same points as Rule 6c.

Table 4.15: The Combination of Rules 2, 3 (Rule 4) and 6c for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.131*2b+0.072*3a	3.540
2b+3a+6c	15	b = 1.131*2b+0.072*3a+0.000*6c	3.540
2b+3a+6c+2b*6c	15	b = 1.089*2b+0.076*3a+0.000*6c+0.056*2b*6c	3.534
2b+3a+6c+3a*6c	15	b = 1.131*2b+0.072*3a+0.000*6c+0.000*3a*6c	3.540
all	15	b = 1.089*2b+0.076*3a+0.000*6c + 0.056*2b*6c+0.000*3a*6c	3.534

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	5	0.828	0.900	10	1.991	3.123
2b+3a+6c	5	0.828	0.900	10	1.991	3.123
2b+3a+6c+2b*6c	5	0.831	0.874	10	1.982	3.107
2b+3a+6c+3a*6c	5	0.828	0.900	10	1.991	3.123
all	5	0.831	0.874	10	1.982	3.107

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Finally, as before, the simple correlations between the rules are shown in Table 4.16. Note that there were high correlations between Rule 6c (Time Parallelism A) and Rules 2b (Attack-Point) and 3a (Register). As before, this is, in part, a consequence of the analysis since Rule 6c was constrained to generate boundaries at only those locations that had low level rule applications.

Table 4.16: Correlation Matrix for "Three Blind Mice", Time Pattern Parallelism A

	Rule						
	2b	3a	4	6c	2b*6c	3a*6c	2b*3a
data	.770**	340*	.766**	.374**	.557**	.134	.595**
Rule 2b		.466**	.997**	.564**	.736**	.271	.756**
Rule 3a			.533**	.647**	.288*	.870**	.661**
Rule 4				.596**	.792**	.335*	.780**
Rule 6c					.807**	.782**	.305*
Rule 2b*6c						.402**	.357*
Rule 3a*6c							.381**

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6c was not too helpful for predicting boundaries. However, it should not be disregarded on the basis of a single melody.

Time Pattern Parallelism B (Rule 6d)

A second form of Time Pattern Parallelism was analyzed. Basically, the relationship between Time Pattern Parallelisms A and B (Rule 6c and 6d) is the same as the relationship between Pitch Pattern Parallelisms A and B (Rule 6a and 6b).

In Time Pattern Parallelism B (Rule 6d), the note-by-note durations were recorded. The durations of two adjacent subsequences of notes were then compared using a correlation coefficient (see Figure 4.13). This is similar to the method of Pitch Pattern Parallelism B (Rule 6b). Hence, two subsequences would be considered similar if the notes in the corresponding sequential positions of the two subsequences had the same durations. As the difference between the durations of corresponding notes increases, subsequences would be considered less similar. Other aspects of this analysis were the same as previous analyses (particularly Pitch Pattern Parallelism B [Rule 6b] since both were based on the values of particular notes). That is, two adjacent subsequences of notes were represented as arrays of note durations. These two arrays were then compared using a correlation similarity measure. Note spans ranging from one to ten notes were used in the comparison (identical to the previous note spans from 1 to 10 used in Pitch Pattern Parallelism B [Rule 6b], and unlike the beat spans from 1 to 10 used in Time Pattern Parallelism A [Rule 6c] and Pitch Pattern Parallelism A [Rule 6a]). The assessment of similarity was the correlation coefficient returned by the note span producing the highest value for the correlation. As before, both subsequences had to end (or begin, depending on one's perspective) on boundaries, had to contain a total of at least five notes, and had to contain at least two notes in each subsequence (the correlation between 2 pairs of numbers is always 1.0). However, as in Time Pattern Parallelism A (Rule 6c), rests were not skipped since the duration of a rest is on the same quantitative scale as the duration of a note.

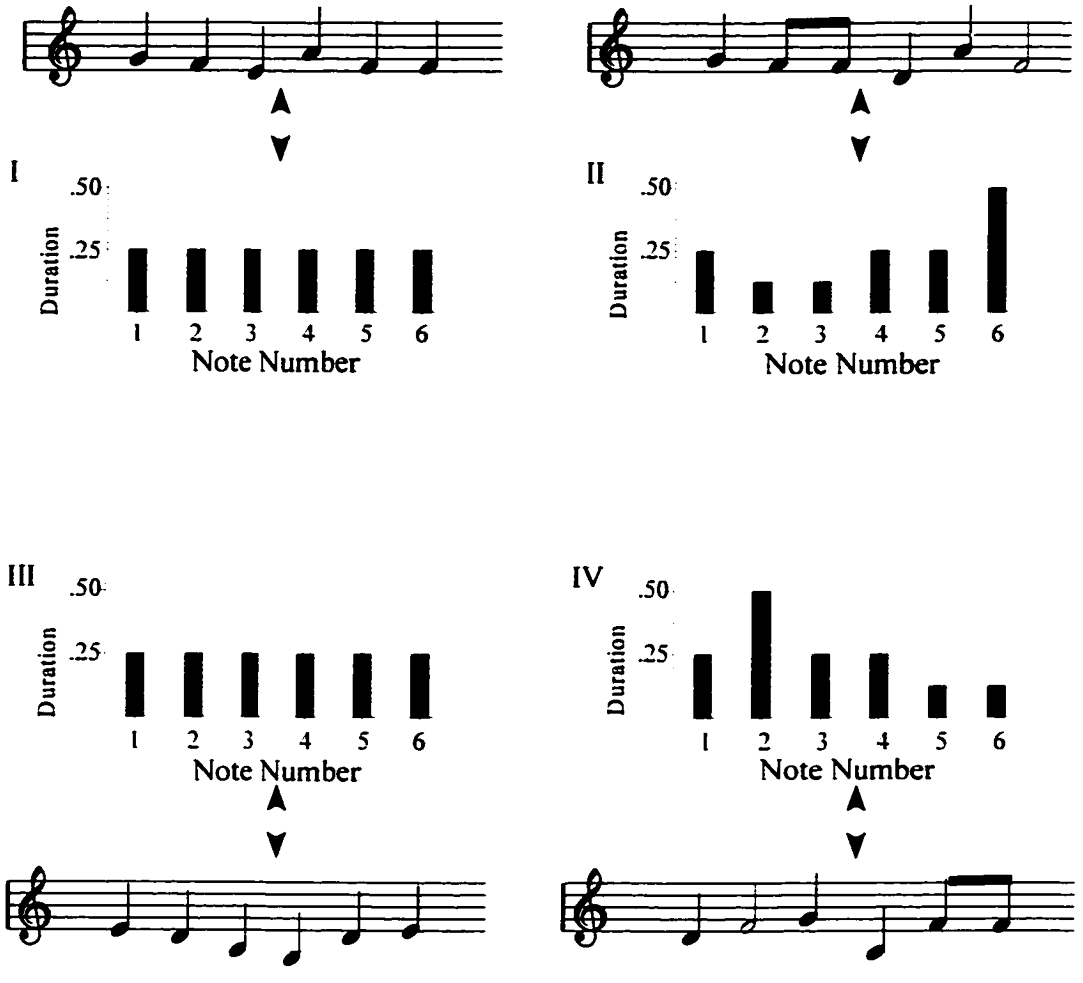



Figure 4.13: Examples of Time Pattern Parallelism, Type B. In this type, temporal patterns represent the durations of each note event. Note that temporal pattern within a score is contained within note structure.

For "Three Blind Mice", the value of the best example of Time Pattern Parallelism B (Rule 6d) at each point in the melody is presented in Figure 4.14 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel), which are necessary for Time Pattern Parallelism B to operate. For this melody, the best match between the assessed parallelism (this form of pitch pattern parallelism) occurred for a note span of four notes (like that in Pitch Pattern Parallelism B [Rule 6b], and unlike that in Time Pattern Parallelism A [Rule 6c]). This note span value captured the strong parallelism at the beginning of the melody. Table 4.17 presents the simple correlations and the corresponding SS_{error} for the fit of Time Pattern Parallelism B (Rule 6d) to the empirical data for all note spans assessed. Note that the correlation was highest and the error the lowest for note spans of four notes. In the actual analysis, the note span that maximized the reduction in total error was used. Note that the although the maximum beat span was four notes, not all examples of Time Pattern Parallelism B (Rule 6d) occurred at a span of four notes.

Table 4.17: Note Span Analysis for "Three Blind Mice", Time Pattern Parallelism B.

Note Span	Empirical Data	SS_{error}
10	.702**	3.508
9	.702**	3.508
8	.705**	3.512
7	.740**	3.319
6	.723**	3.312
5	.729**	3.264
4	.844**	2.932
3	.409**	3.050

Notes: * $p < .05$
** $p < .01$

In the third panel of Figure 4.14 observe that, as before, there is no parallelism at Note 3 because, as in the other assessments of parallelism, one cannot have parallelism until one has a previous sequence to compare to. Also observe that the parallelism at Notes 4 and 14 are perfect (spans of 3 and 4 notes respectively). Also note that there is a value for Time Pattern Parallelism B (Rule 6d) at Note 10 (unlike the case in Pitch Pattern Parallelism B [Rule 6b]) because Notes 3 through 5 are parallel with Notes 7 through 10 due to the relatively longer notes at the beginning and ends of each subsequence (span of 4 notes). In the rest of the melody, Time Pattern Parallelism B (Rule 6d) produces boundaries at the same points as Pitch Pattern Parallelism B (Rule 6b). That is, few boundaries are placed. This is likely due, in part, to the constraints on the analysis (particularly that both subsequences must end on a boundary).

The analysis of Time Pattern Parallelism B (Rule 6d) entered this rule in conjunction with Rules 2 and 3, because Rule 6d is built on putative boundaries identified by Rules 2 and 3. Like Rules 6a, 6b, and 6c, Rule 6d was conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6d with Rules 2b and 3a were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a and 6d, and then all rules and then the model containing all rules and their interactions was examined. Table 4.18 provides the details.

Surprisingly, in this analysis, it is apparent that Time Pattern Parallelism B (Rule 6d) did add predictability to the model beyond that which could be explained by the low level rules. The improvement was from 3.54 to 2.93 (a 17.18% change). This predictability was due mainly to the rule itself (i.e., not the interactions). However, note that although the overall error drops, the miss rate actually increases. The effect is essentially the same as that of Pitch Pattern Parallelism B (Rule 6b). The inclusion of Rule 6d allows some of the existing boundaries (empirical) to be explained, in part, by Rule 6d, thereby allowing the false alarm rate for Rules 2 and 3 to decline. Rule 6d did not induce many false alarms. The interactions do not aid predictability beyond that of the Rules 2, 3 and 6d.

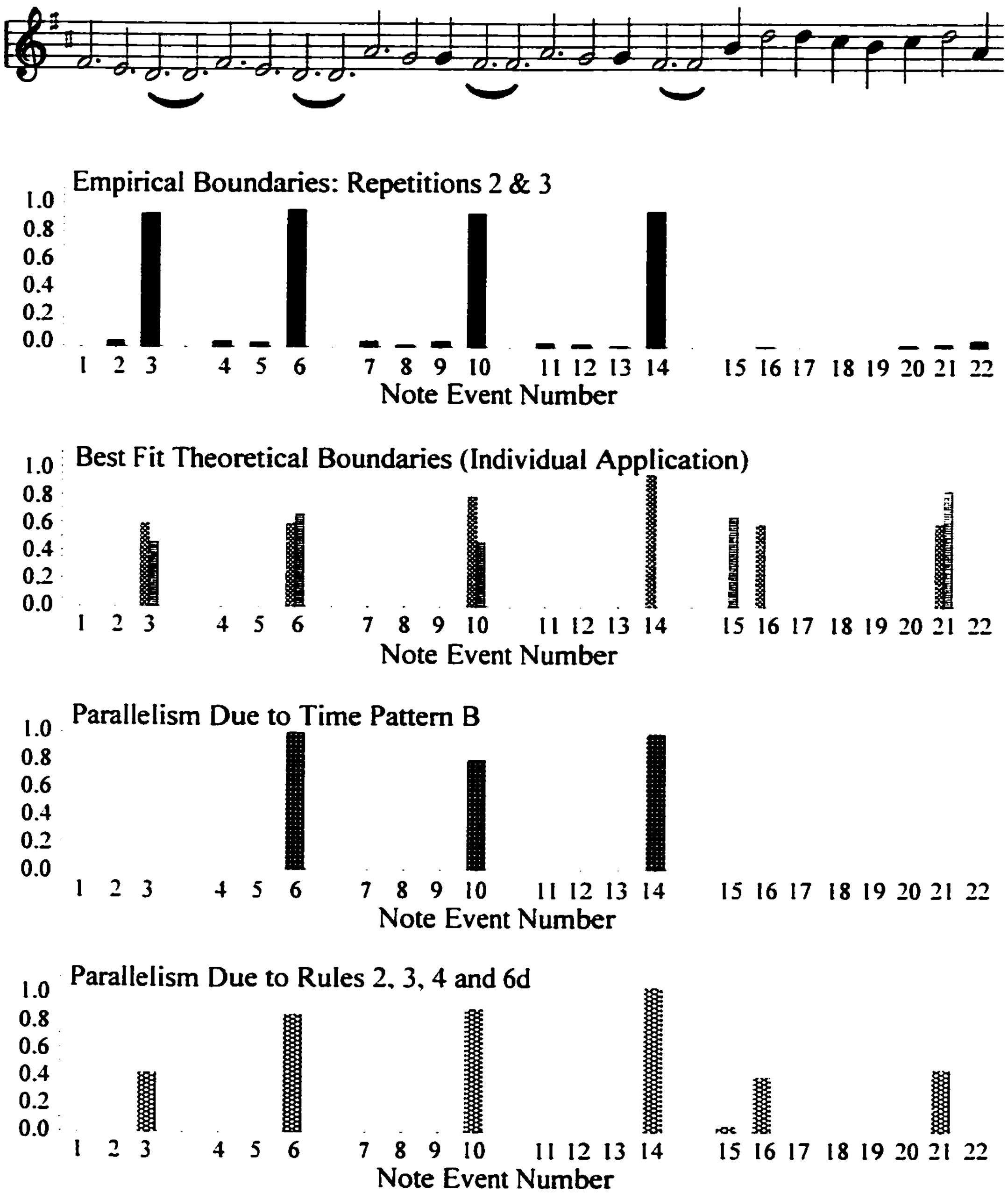


Figure 4.14a: The first part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a and Rule 6d (parallelism due to Time Pattern B).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**	■ **Rule 6d**
▤ **Rule 3a**	▦ **Rules 2b, 3a 4, 6 and interactions**

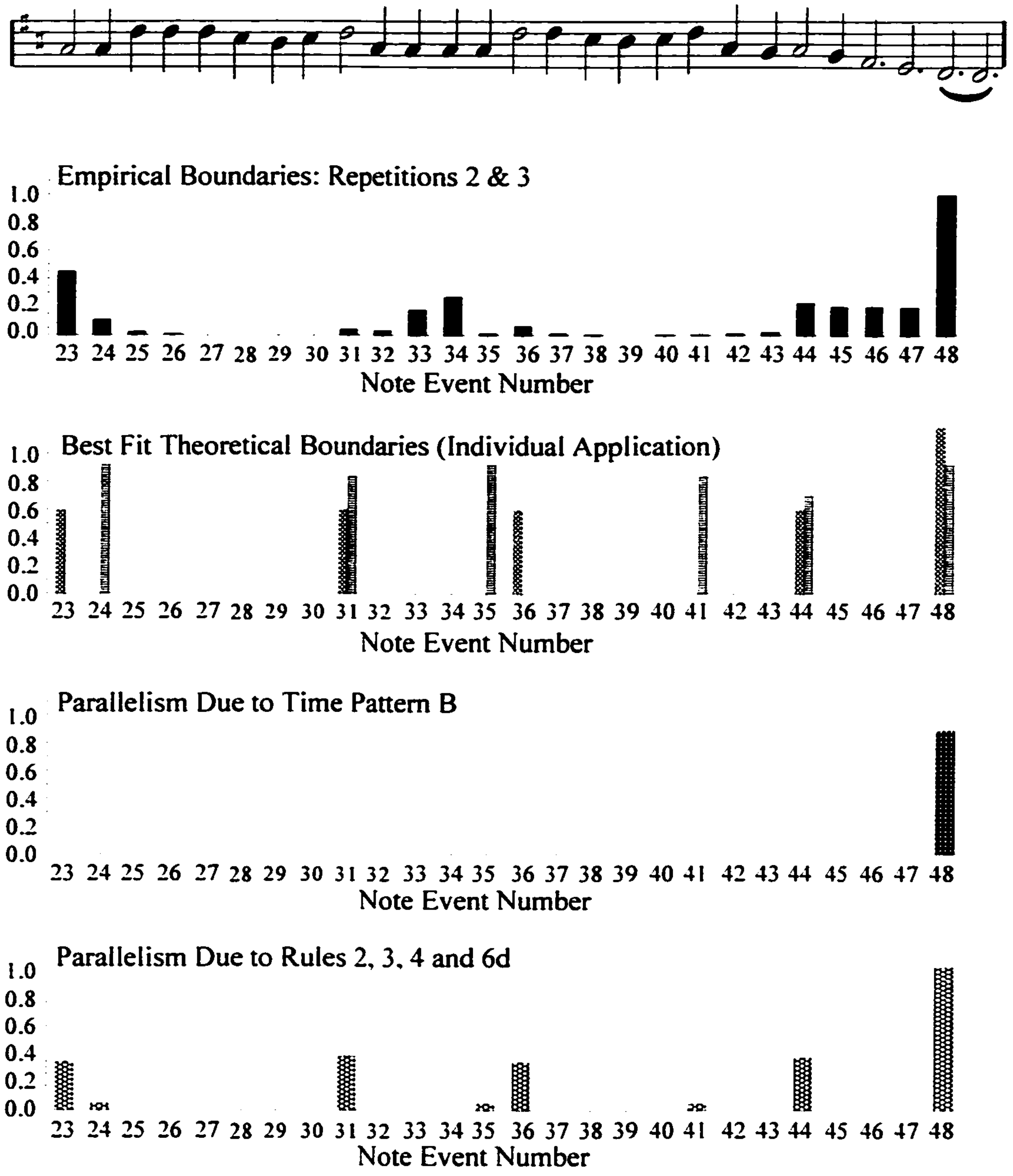


Figure 4.14b: The second part of the melody used in Stage 4, Part 1, and the best individual fits for Lerdahl and Jackendoff (1983) Group Preference Rules 2b and3a and Rule 6a (parallelism due to Time Pattern B).This can be compared with the location of subjective boundaries based on the average of all subjects over the last two repetitions of the song.

▨ **Rule 2a**
■ **Rule 6d**
▨ **Rule 3a**
▩ **Rules 2b, 3a 4, 6 and interactions**

Table 4.18: The Combination of Rules 2, 3 (Rule 4) and 6d for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.131*2b+0.072*3a	3.540
2b+3a+6d	15	b = 0.788*2b+0.057*3a+0.403*6d	2.932
2b+3a+6c+2b*6d	15	b = 0.788*2b+0.057*3a+0.403*6d+0.000*2b*6d	2.932
2b+3a+6d+3a*6d	15	b = 0.788*2b+0.056*3a+0.399*6d+0.084*3a*6d	2.932
all	15	b = 0.788*2b+0.056*3a+0.399*6d + 0.000*2b*6d+0.084*3a*6d	2.932

Breakdown for Misses and False Alarms						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	5	0.828	0.900	10	1.991	3.123
2b+3a+6d	5	1.095	0.799	10	1.116	2.150
2b+3a+6d+2b*6d	5	1.095	0.799	10	1.116	2.150
2b+3a+6d+3a*6d	5	1.095	0.798	10	1.117	2.148
all	5	1.095	0.798	10	1.117	2.148

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 4.19. Note that there were high correlations between Rule 6d (Time Parallelism A) and Rules 2b (Attack-Point) and 3a (Register). As before, this was, in part, a consequence of the analysis since Rule 6d was constrained to generate boundaries at only those locations that had low level rule applications.

Table 4.19: Correlation Matrix for "Three Blind Mice", Time Pattern Parallelism B

	Rule						
	2b	3a	4	6d	2b*6d	3a*6d	2b*3a
data	.770**	340*	.766**	.844**	.825**	.702**	.595**
Rule 2b		.466**	.997**	.700**	.733**	.591**	.756**
Rule 3a			.533**	.309**	.311*	.433**	.661**
Rule 4				.696**	.728**	.603*	.780**
Rule 6d					.969**	.804**	.560**
Rule 2b*6d						.792**	.620*
Rule 3a*6d							.763**

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6d was helpful for predicting boundaries, but this was due to, as in Rule 6b, its lack of false alarms in conjunction with its overlap with Rules 2 and 3.

Parallelism (Rule 6a) in "Softly Now the Light of Day"

All the previous analyses that were applied to the melody "Three Blind Mice" were applied to the extract from the melody "Softly Now the Light of Day". It was thought that the information provided by one melody should not be taken as grounds for rejecting a single rule. That is, while it seemed that Rule 6c (Time Pattern Parallelism A) did not add predictability (or that only Rules 6b and 6d added predictability), this might be unique to the particular melody.

On the other hand, in the interests of brevity, the presentation of the analyses for this melody are compressed wherever possible. That is, it is hoped that the previous discussions pertaining to the individual analysis were sufficiently detailed that I may be permitted to say that all aspects of the analysis were the same: The only change was the stimulus and its associated empirical boundaries.

Pitch Pattern Parallelism A (Rule 6a)

For "Softly Now the Light of Day", the value of the best example of Pitch Pattern Parallelism A (Rule 6a) at each point in the melody is presented in Figure 4.15 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel; in comparison with Figure 4.5, it was necessary to place Rule 2 and 3 for Stages 6 and 8 within the same panel), which are necessary for Pitch Pattern Parallelism A to operate. For this melody, the best match between the assessed parallelism was not obvious. The correlations between Rule 6a and the empirical boundaries were not dramatic (see Table 4.20): Reductions in the total error were similarly unimpressive. Hence, a beat span of ten was used for compatibility with the subsequent analysis of Time Pattern Parallelism A. Table 4.20 presents the simple correlations and the corresponding SS_{error} for the fit of Pitch Pattern Parallelism A (Rule 6a) to the empirical data for all beat spans assessed. Note that the correlations and the errors do not differ (errors are reported only for Stage 8 data: Empirical Data 2).

Table 4.20: Beat Span Analysis for "Softly Now the Light of Day", Pitch Pattern Parallelism A.

Beat Span	Stage 6	Stage 8	SS_{error}
10	.199	.168	1.353
9	.199	.168	1.353
8	.198	.166	1.361
7	.198	.166	1.361
6	.070	.042	1.326
5	.070	.042	1.326
4	.027	.012	1.380
3	.034	.020	1.380

Notes: * $p < .05$
** $p < .01$

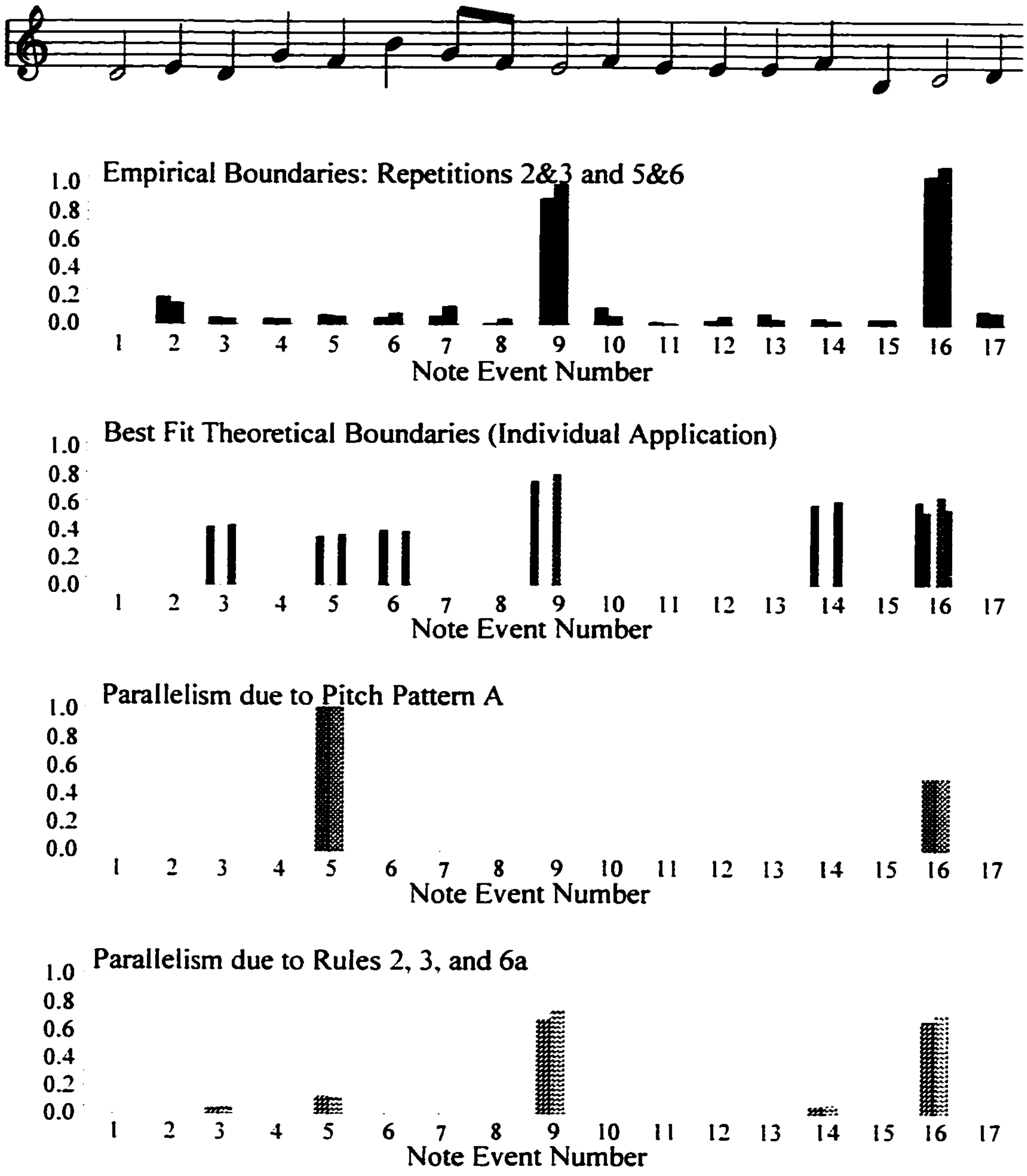


Figure 4.15a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6a. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▩ Rule 2b
▤ Rule 3a
□ Rule 3d
■ Rule 6a
▦ Rules 2b, 3a 4, 6 and interactions
■ Stage 6 (Presentations 2 & 3)
■ Stage 8 (Presentations 5 & 6)

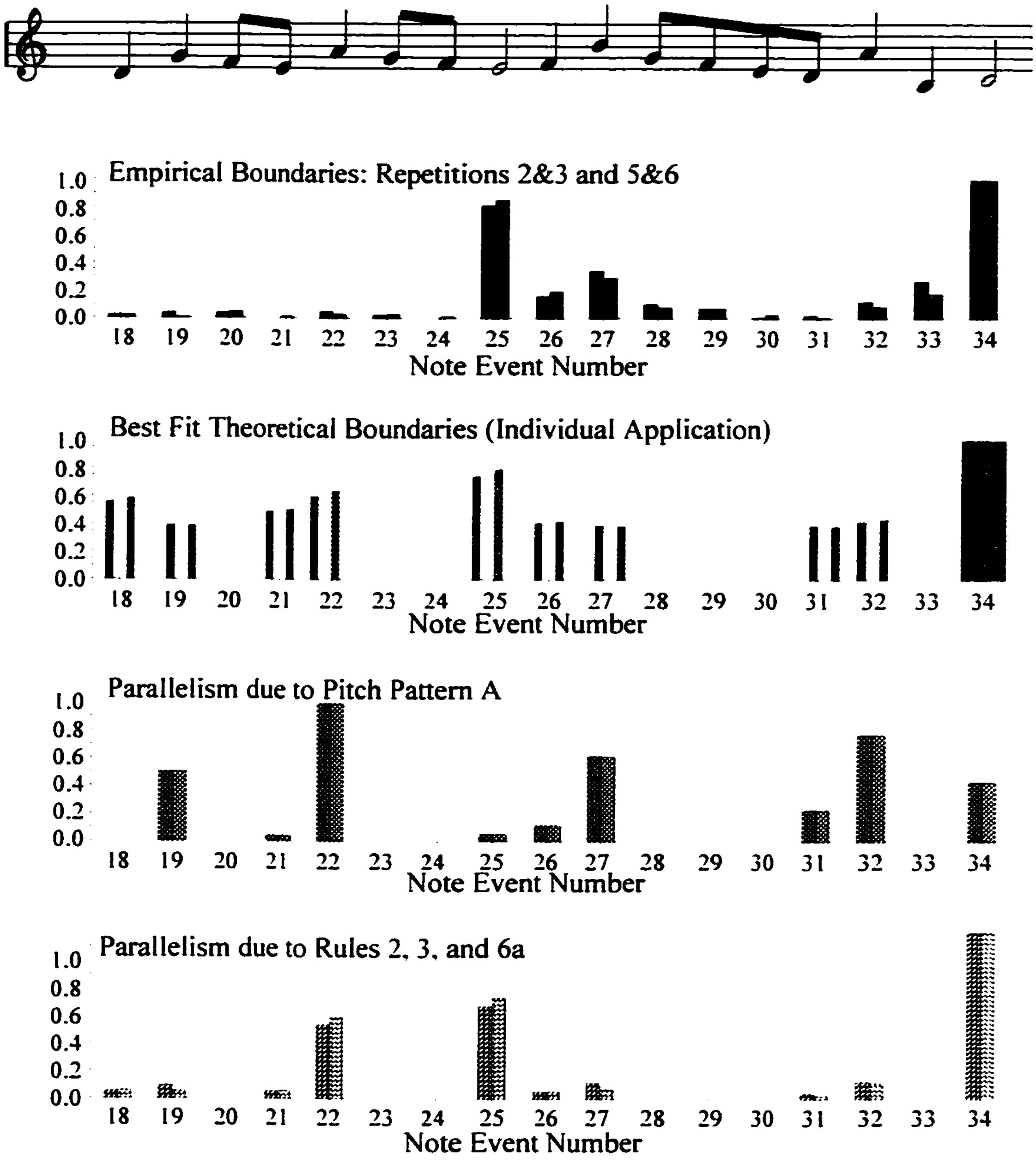


Figure 4.15b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6a. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b** ■ **Rule 6a** ■ **Stage 6 (Presentations 2 & 3)**
▤ **Rule 3a** ▦ **Rules 2b, 3a 4, 6 and interactions** ■ **Stage 8 (Presentations 5 & 6)**
□ **Rule 3d**

In the third panel of Figure 4.15, observe that parallel structures aligns with only one of two strong boundaries in the first half of the melody (and the weaker of the two). The low level rules however, aline with both of these boundaries, so parallelism has nothing to add. In the second half of the melody, the situation is worse, because the parallel structures (there are a number of them) do not aline with the one strong boundary while the low level Rule 2b does. Hence, Rule 6a generates a lot of false alarms. From the figure, it would seem that Pitch Pattern Parallelism A (Rule 6a) is redundant.

In the analysis of Pitch Pattern Parallelism A (Rule 6a), Rule 6a was entered in conjunction with Rules 2 and 3, because Rule 6a is built on putative boundaries identified by Rules 2 and 3. Rule 6a is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6a with Rules 2b, 3a and 3d were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a, 3d and 6a, and then all rules and then the model containing all rules and their interactions was examined. Table 4.21 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 4.22 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged). Note that the combination of all rules and interactions did provide some improvement overall (Stage 6: 5.86%; Stage 8: 2.97%) and in the miss rate despite more data points being classified as misses (Reminder: A miss is any point where the prediction is less than the empirical data.). The improvement was mainly attributable to the parallelism at Note 16.

Table 4.21: The Combination of Rules 2, 3 (Rule 4) and 6a for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.093*2b+0.115*3a+0.116*3d	1.467
2b+3a+3d+6a	16	b=1.087*2b+0.072*3a+0.043*3d+0.114*6a	1.416
2b+3a+3d+6a+2b*6a	16	b=1.087*2b+0.072*3a+0.043*3d+0.114*6a + 0.000*2b*6a	1.416
2b+3a+3d+6a+3a*6a	16	b=1.087*2b+0.072*3a+0.043*3d+0.114*6a + 0.000*3a*6a	1.416
2b+3a+3d+6a+3d*6a	16	b=1.083*2b+0.084*3a+0.000*3d+0.076*6a + 2.169*3d*6a	1.393
all	16	b=1.066*2b+0.066*3a+0.000*3d+0.000*6a +0.000*2b*6a+1.594*3b*6a+3.256*3d*6a	1.381

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.500	0.841	10	0.551	1.118
2b+3a+3d+6a	6	0.403	0.762	10	0.598	1.129
2a+3a+3d+6a+2b*6a	6	0.403	0.762	10	0.598	1.129
2a+3a+3d+6a+3a*6a	6	0.403	0.762	10	0.598	1.129
2a+3a+3d+6a+3d*6a	7	0.376	0.784	9	0.603	1.159
all	7	0.393	0.801	9	0.573	1.093

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 4.22: The Combination of Rules 2, 3 (Rule 4) and 6a for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.189+2b+0.115*3a+0.060*3d	1.380
2b+3a+3d+6a	16	b=1.182*2b+0.082*3a+0.006*3d+0.084*6a	1.353
2b+3a+3d+6a+2b*6a	16	b=1.182*2b+0.082*3a+0.006*3d+0.084*6a + 0.000*2b*6a	1.353
2b+3a+3d+6a+3a*6a	16	b=1.179*2b+0.079*3a+0.015*3d+0.070*6a + 0.317*3a*6a	1.353
2b+3a+3d+6a+3d*6a	16	b=1.181*2b+0.090*3a+0.000*3d+0.062*6a + 1.833*3d*6a	1.348
all	16	b=1.167*2b+0.075*3a+0.000*3d+0.000*6a +0.000*2b*6a+1.331*3b*6a+3.617*3d*6a	1.339

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	7	0.486	0.868	9	0.650	1.214
2b+3a+3d+6a	6	0.432	0.846	10	0.676	1.247
2a+3a+3d+6a+2b*6a	6	0.432	0.846	10	0.676	1.247
2a+3a+3d+6a+3a*6a	6	0.440	0.851	10	0.668	1.235
2a+3a+3d+6a+3d*6a	6	0.417	0.840	10	0.687	1.269
all	6	0.428	0.862	10	0.666	1.231

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 4.23 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown). Note that in contrast to the previous melody, "Three Blind Mice" , the correlations between Rule 6a (Pitch Pattern Parallelism A) and Rules 2b (Attack-Point), 3a (Register) and 3d (Length) were not high. This difference between the melodies arises because Rule 6a was constrained on the basis of any single application of any one of three rules (Rules 2b, 3a or 3d) and those three rules (Rules 2b. 3a and 3d) themselves are not highly correlated. In the melody "Three Blind Mice", Rule 6a was constrained by only two rules, which were moderately correlated. As a consequence Rule 6a was more likely to be highly correlated with both.

Table 4.23: Correlation Matrix for "Softly Now the Light of Day", Pitch Pattern Parallelism A.

	Rule							
	2b	3a	3d	4	6a	2b*6a	3a*6a	3d*6a
Stage 6	.867**	.313	.367*	.876**	.199	.487**	.430*	.428*
Stage 8	.883**	.310	.334*	.882**	.168	.474**	.416**	.382*
Rule 2b		.292	.386*	.988**	.296	.712**	.377*	.410*
Rule 3a			.178	.412*	.278	.312	.610**	.212
Rule 3d				.467**	.285	.367*	.243	.898**
Rule 4					.333	.762**	.439**	.451**
Rule 6a						.538**	.671**	.369**
R 2b*6d							.382*	.386*
R 3a*6d								.266

Notes: * $p < .05$
** $p < .01$

Generally, Pitch Pattern Parallelism A (Rule 6a) did add some predictability to the model beyond that which could be explained by the low level rules. The same pattern was evident in both stages. This predictability was due to both the rule and its interactions, particularly with Rules 3a and 3d (not Rule 2b), although in the final equation Rule 6a alone was not a factor (i.e., Rule 6a was dropped while the interactions were retained).

Pitch Pattern Parallelism B (Rule 6b)

For "Softly Now the Light of Day", the value of the best example of Pitch Pattern Parallelism B (Rule 6b) at each point in the melody is presented in Figure 4.16 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel; again, Rules 2 and 3 within Stages 6 and 8 are presented in a single panel), which are necessary for Pitch Pattern Parallelism B to operate. For this melody, as in Pitch Pattern Parallelism A (Rule 6a), the best match between the assessed parallelism was not obvious. The correlations between Rule 6b and the empirical boundaries were not dramatic (see Table 4.23) and reductions in the total error were similarly unimpressive. Hence, a note span of ten notes was used for compatibility with the subsequent analysis of Time Pattern Parallelism B. Table 4.23 presents the simple correlations and the corresponding SS_{error} for the fit of Pitch Pattern Parallelism B (Rule 6b) to the empirical data for all note spans assessed. Note that the correlations and the errors do not differ (errors are reported for Stage 8 data only).

Table 4.24: Note Span Analysis for "Softly Now the Light of Day", Pitch Pattern Parallelism B

Note Span	Stage 6	Stage 8	SS_{error}
10	.199	.168	1.380
9	.199	.168	1.380
8	.169	.138	1.380
7	.168	.137	1.380
6	.112	.079	1.380
5	.112	.079	1.380
4	.197	.189	1.380
3	.263	.244	1.380

Notes: * $p < .05$
** $p < .01$

In the third panel of Figure 4.16 observe that, like Pitch Pattern Parallelism A (Rule 6a), there is not a strong relationship between Rule 6b and the empirical data, especially not a relationship that is unique to Pitch Pattern Parallelism B.

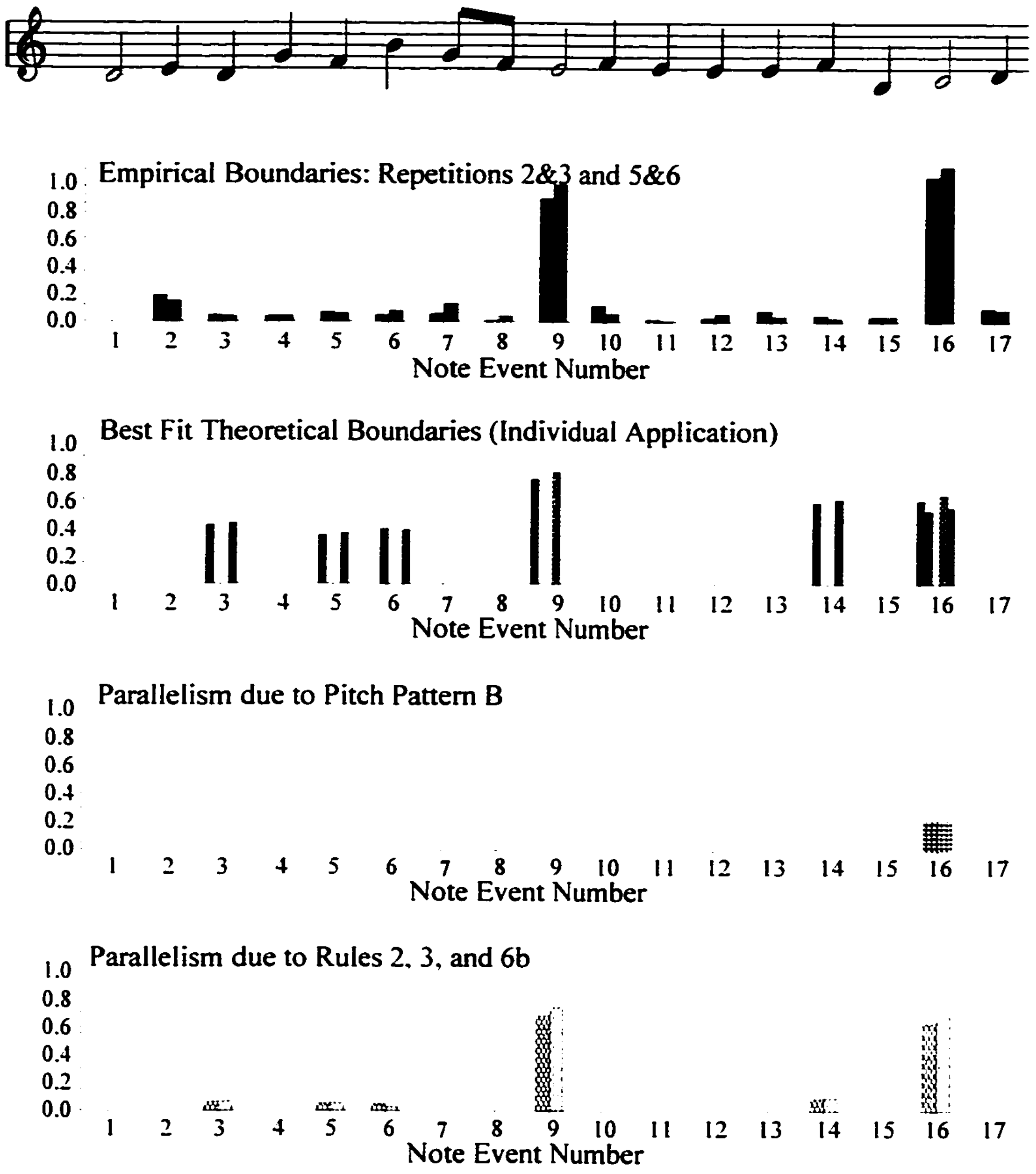


Figure 4.16a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6b. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b**
▨ **Rule 3a**
▨ **Rule 3d**
▨ **Rule 6b**
▨ **Rules 2b, 3a 4, 6 and interactions**
■ **Stage 6 (Presentations 2 & 3)**
■ **Stage 8 (Presentations 5 & 6)**

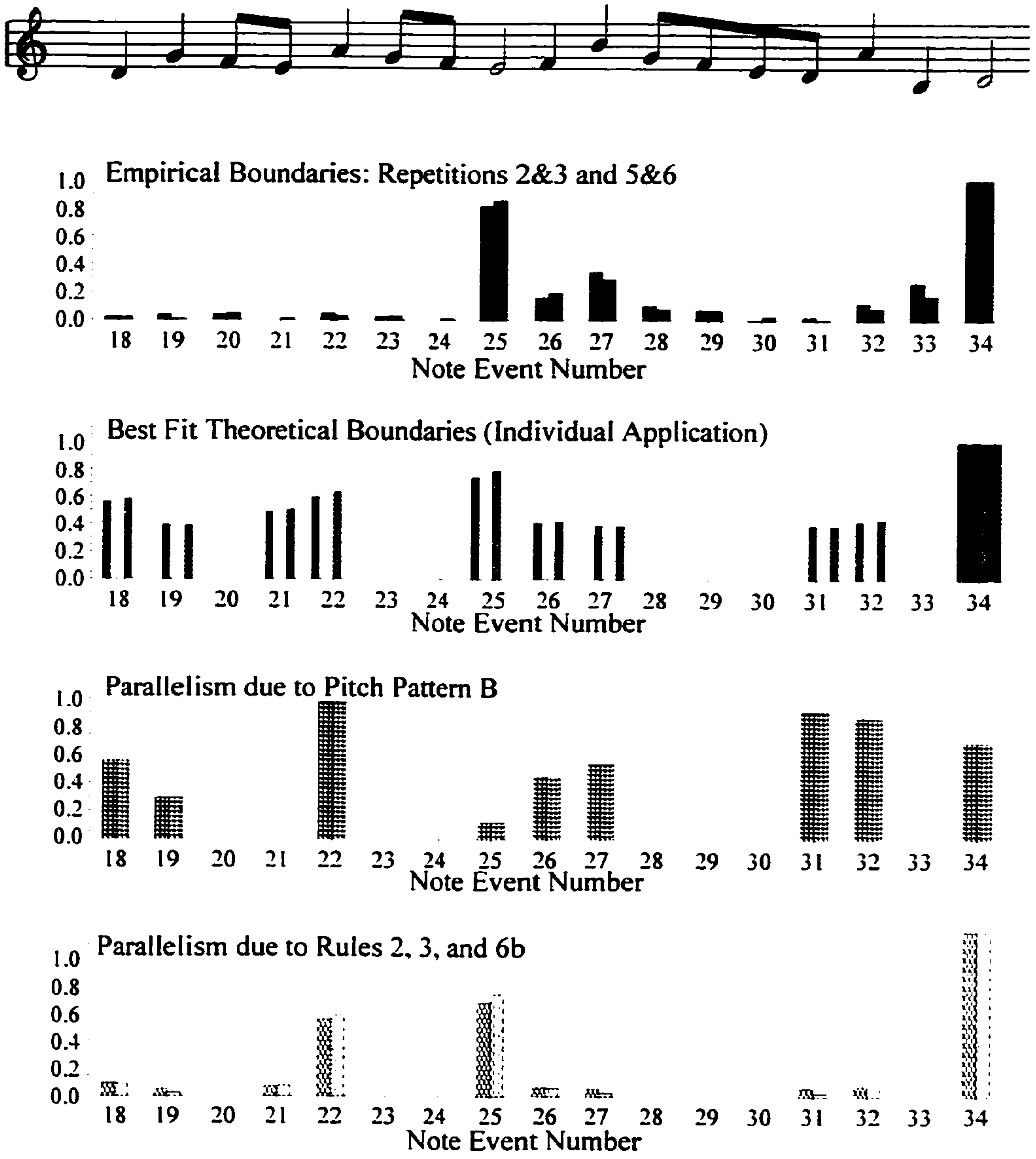


Figure 4.16b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6b. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b** ▦ **Rule 6b** ■ **Stage 6 (Presentations 2 & 3)**
▩ **Rule 3a** ▨ **Rules 2b, 3a 4, 6 and interactions** ■ **Stage 8 (Presentations 5 & 6)**
▧ **Rule 3d**

In the analysis of Pitch Pattern Parallelism B (Rule 6b), the rule was entered in conjunction with Rules 2 and 3, because this rule is built on putative boundaries identified by Rules 2 and 3. Rule 6b is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6b with Rules 2b, 3a and 3d were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a, 3d and 6b, and then all rules and then the model containing all rules and their interactions was examined. Table 4.25 provides the details for Stage 6 (Repetition 2 and 3 averaged) and Table 4.26 provides the details for the Stage 8 (Repetitions 5 and 6 averaged). Note that the combination of all rules and interactions did not provide any improvement overall (Stage 6: 0.07%; Stage 8: 0.00%) and did not provide any real change in the miss or false alarm rates. In contrast to Rule 6a (Pitch Pattern Parallelism A), Rule 6b did not help predictability at all. It is interesting that Rule 6a and 6b had similar simple correlations with the original data, but only Rule 6a had some utility. What matters is not so much the simple correlation, but rather the question of what each rule can explain that has not been explained elsewhere.

Table 4.25: The Combination of Rules 2, 3 (Rule 4) and 6b for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.093*2b+0.115*3a+0.116*3d	1.467
2b+3a+3d+6b	16	b=1.093*2b+0.109*3a+0.103*3d+0.016*6b	1.466
2b+3a+3d+6b+2b*6a	16	b=1.093*2b+0.109*3a+0.103*3d+0.016*6b + 0.000*2b*6a	1.466
2b+3a+3d+6b+3a*6a	16	b=1.093*2b+0.109*3a+0.103*3d+0.016*6b + 0.000*3a*6b	1.466
2b+3a+3d+6b+3d*6a	16	b=1.093*2b+0.109*3a+0.103*3d+0.016*6b + 0.000*3a*6b	1.466
all	16	b=1.093*2b+0.109*3a+0.103*3d+0.016*6b +0.000*2b*6b+0.000*3b*6b+0.000*3d*6b	1.466

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.500	0.841	10	0.551	1.118
2b+3a+3d+6b	7	0.491	0.826	9	0.560	1.118
2a+3a+3d+6b+2b*6b	7	0.491	0.826	9	0.560	1.118
2a+3a+3d+6b+3a*6b	7	0.491	0.826	9	0.560	1.118
2a+3a+3d+6b+3d*6b	7	0.491	0.826	9	0.560	1.118
all	7	0.491	0.826	9	0.560	1.118

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 4.26: The Combination of Rules 2, 3 (Rule 4) and 6b for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.189+2b+0.115*3a+0.060*3d	1.380
2b+3a+3d+6b	16	b=1.189*2b+0.115*3a+0.060*3d+0.000*6b	1.380
2b+3a+3d+6b+2b*6b	16	b=1.189*2b+0.115*3a+0.060*3d+0.000*6b + 0.000*2b*6b	1.380
2b+3a+3d+6b+3a*6b	16	b=1.189*2b+0.115*3a+0.060*3d+0.000*6b + 0.000*3a*6b	1.380
2b+3a+3d+6b+3d*6b	16	b=1.189*2b+0.115*3a+0.060*3d+0.000*6b + 0.000*3d*6b	1.380
all	16	b=1.189*2b+0.115*3a+0.060*3d+0.000*6b +0.000*2b*6b+0.000*3b*6b+0.000*3d*6b	1.380

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	7	0.486	0.868	9	0.650	1.214
2b+3a+3d+6b	7	0.432	0.868	9	0.650	1.214
2a+3a+3d+6b+2b*6b	7	0.432	0.868	9	0.650	1.214
2a+3a+3d+6b+3a*6b	7	0.432	0.868	9	0.650	1.214
2a+3a+3d+6b+3d*6b	7	0.432	0.868	9	0.650	1.214
all	7	0.432	0.868	9	0.650	1.214

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 4.27 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown). It is apparent from the matrix that the combination of high predictability based on Rules 2 and 3 in combination with the low predictability based on Rule 6b is what makes Rule 6b so ineffectual as an aid for prediction.

Table 4.27: Correlation Matrix for "Softly Now the Light of Day", Pitch Pattern Parallelism B.

	Rule							
	2b	3a	3d	4	6b	2b*6b	3a*6b	3d*6b
Stage 6	.867**	.313	.367*	.876**	.199	.511**	.418*	.425*
Stage 8	.883**	.310	.334*	.882**	.168	.490**	.401**	.388*
Rule 2b		.292	.386*	.988**	.338	.783**	.428*	.459**
Rule 3a			.178	.412*	.314	.335	.695**	.254
Rule 3d				.467**	.465**	.526**	.404*	.893**
Rule 4					.377*	.799**	.504**	.502**
Rule 6b						.533**	.591**	.561**
R 2b*6b							.524**	.599**
R 3a*6b								.477**

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6b is useless for predicting boundaries.

Time Pattern Parallelism A (Rule 6c)

Time Pattern Parallelism A (Rule 6c), as can be seen in Figure 4.17 and Table 4.28, shows more promise as a predictor than either of Rule 6a or 6b. The value of the best example of Time Pattern Parallelism A (Rule 6c) at each point in the melody is presented in Figure 4.17 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel; again Rules 2 and 3 for Stages 6 and 8 are presented in a single panel), which are necessary for Time Pattern Parallelism A to operate. For this melody, the best match between the assessed parallelism (this form of pitch pattern parallelism) was fairly obvious (especially in contrast to Rules 6a and 6b). The correlations and the errors for the relationship between Rule 6c and the empirical boundaries at each value of beat span are shown in Table 4.28. Since there was not much to distinguish beat spans of ten from beat spans of nine, beats spans of ten were used.

Table 4.28: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism A

Beat Span	Stage 6	Stage 8	SS_{error}
10	.487**	.483**	1.211
9	.487**	.483**	1.211
8	.489**	.484**	1.218
7	.240	.235	1.378
6	.248	.243	1.378
5	.213	.209	1.375
4	.281	.293	1.380
3	.056	.068	1.380

Notes: * $p < .05$
** $p < .01$

In the third panel of Figure 4.17 observe that there are many notes exhibiting some degree of Time Pattern Parallelism A. Also note that, more critically, every major empirical boundary is associated with some parallelisms. Given that false alarms are not weighted equally with misses, on the basis of the figure, one would expect Time Pattern Parallelism A to be helpful.

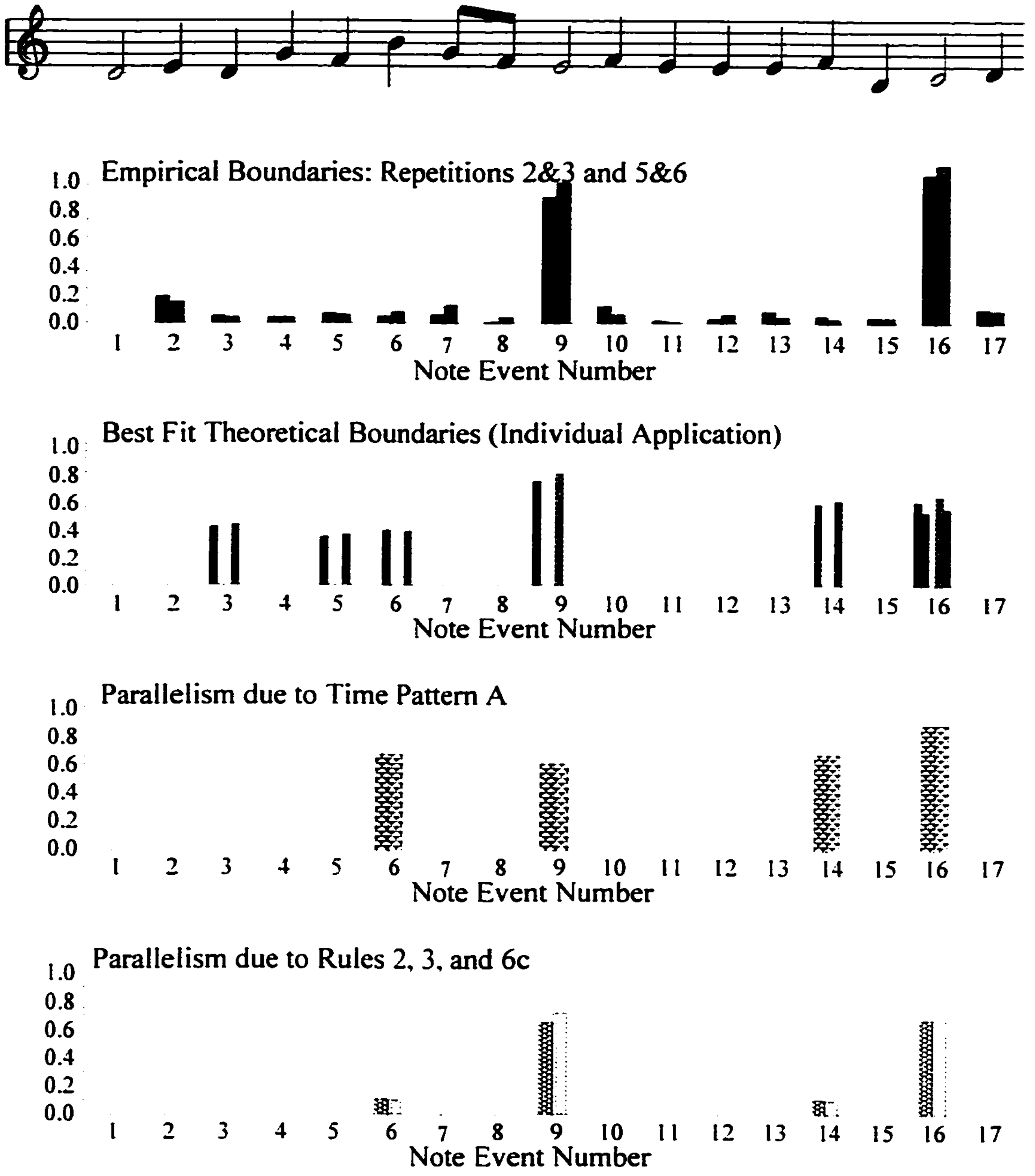


Figure 4.17a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6c. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b**
▨ **Rule 3a**
▨ **Rule 3d**
▨ **Rule 6c**
▨ **Rules 2b, 3a 4, 6 and interactions**
■ **Stage 6 (Presentations 2 & 3)**
■ **Stage 8 (Presentations 5 & 6)**

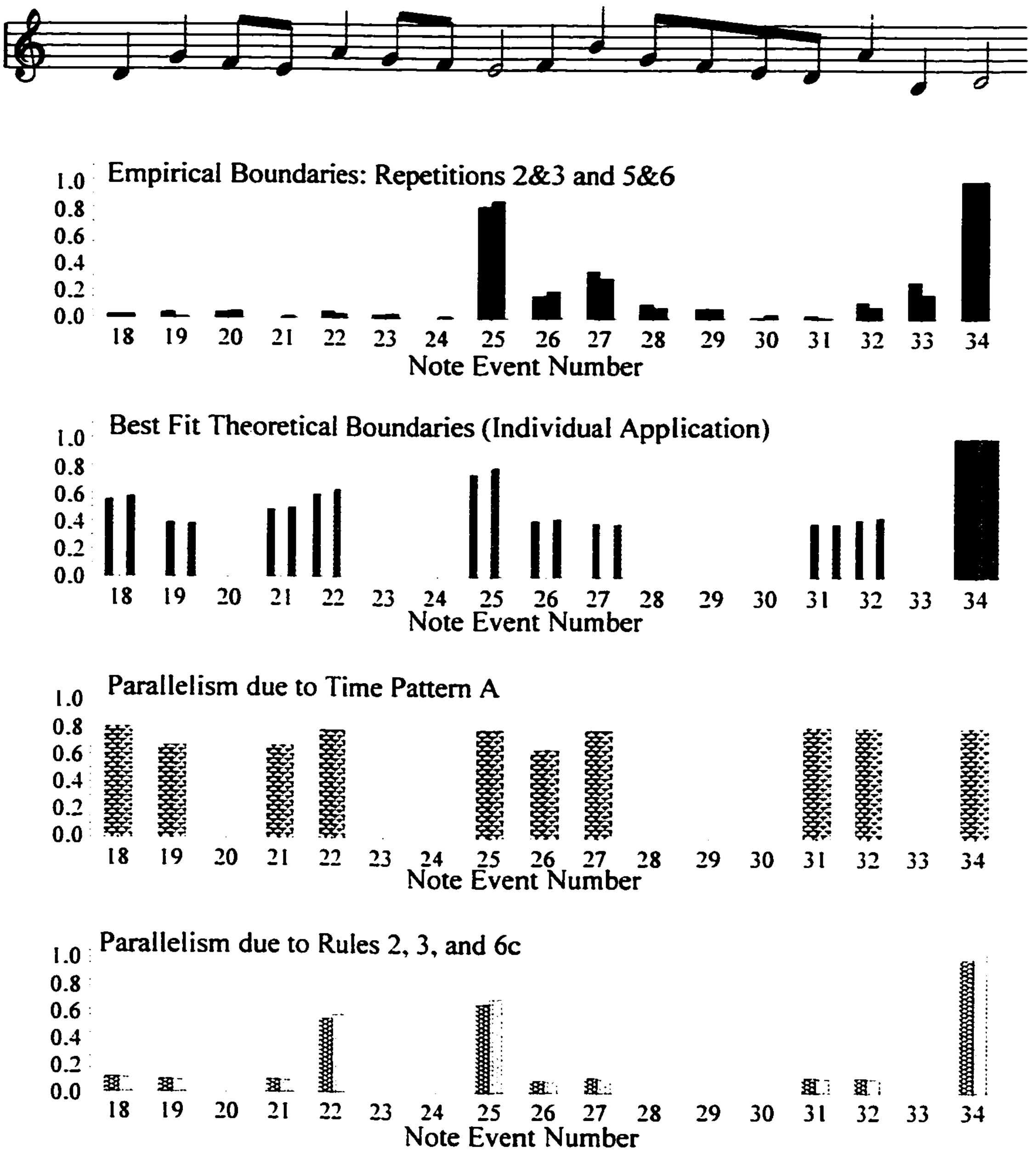


Figure 4.17b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6c. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b** ▨ **Rule 6c** ■ **Stage 6 (Presentations 2 & 3)**
▤ **Rule 3a** ▨ **Rules 2b, 3a 4, 6 and interactions** ■ **Stage 8 (Presentations 5 & 6)**
▧ **Rule 3d**

In the analysis of Time Pattern Parallelism A (Rule 6c), Rule 6c was entered in conjunction with Rules 2 and 3, because Rule 6c is conceived as something that gives a boost to particular locations by virtue of their parallel structure. The interactions of Rule 6c with Rules 2b, 3a and 3d were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a, 3d and 6c, and then all rules and then the model containing all rules and their interactions was examined. Table 4.28 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 4.29 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged). Note that the combination of all rules and interactions did provides substantial improvement overall (13.49% and 12.83%) and in the miss rate. The false alarm rate was fairly constant.

In this analysis, it is apparent that Time Pattern Parallelism A (Rule 6c) did add predictability to the model beyond that which could be explained by the low level rules. This predictability was due mainly to the rule itself and the interaction of this rule with Rule 2b. In the analysis, Rules 3a and 3d (particularly 3d) were dropped because 6c and its interaction with 2b could explain everything that the other rules and their interactions could explain (but better). One might argue that the improvement results from the fact that Rule 6c does what Rule 3d did, but better (the argument is made from a numerical perspective, not a conceptual one).

Table 4.29: The Combination of Rules 2, 3 (Rule 4) and 6c for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.093*2b+0.115*3a+0.116*3d	1.467
2b+3a+3d+6c	16	b=0.984*2b+0.008*3a+0.000*3d+0.166*6c	1.285
2b+3a+3d+6c+2b*6c	16	b=0.512*2b+0.000*3a+0.000*3d+0.166*6c + 0.571*2b*6c	1.269
2b+3a+3d+6c+3a*6c	16	b=0.984*2b+0.008*3a+0.000*3d+0.166*6c + 0.000*3a*6c	1.285
2b+3a+3d+6c+3d*6c	16	b=0.984*2b+0.008*3a+0.000*3d+0.166*6c + 0.000*3d*6c	1.393
all	16	b=0.512*2b+0.000*3a+0.000*3d+0.166*6c +0.571*2b*6c+0.000*3b*6c+0.000*3d*6c	1.269

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.500	0.841	10	0.551	1.118
2b+3a+3d+6c	7	0.307	0.657	9	0.563	1.269
2a+3a+3d+6c+2b*6c	7	0.284	0.695	9	0.570	1.265
2a+3a+3d+6c+3a*6c	7	0.307	0.657	9	0.563	1.269
2a+3a+3d+6c+3d*6c	7	0.307	0.657	9	0.563	1.269
all	7	0.284	0.695	9	0.570	1.265

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 4.30: The Combination of Rules 2, 3 (Rule 4) and 6c for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.189+2b+0.115*3a+0.060*3d	1.380
2b+3a+3d+6c	16	b=1.071*2b+0.012*3a+0.000*3d+0.152*6c	1.211
2b+3a+3d+6c+2b*6c	16	b=0.745*2b+0.000*3a+0.000*3d+0.155*6c + 0.366*2b*6c	1.203
2b+3a+3d+6c+3a*6c	16	b=1.071*2b+0.012*3a+0.000*3d+0.152*6c + 0.000*3a*6c	1.211
2b+3a+3d+6c+3d*6c	16	b=1.071*2b+0.012*3a+0.000*3d+0.152*6c + 0.000*3d*6c	1.211
all	16	b=0.745*2b+0.000*3a+0.000*3d+0.154*6c +0.366*2b*6c+0.000*3b*6c+0.000*3d*6c	1.203

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	7	0.432	0.846	9	0.650	1.214
2b+3a+3d+6c	7	0.311	0.693	9	0.656	1.396
2a+3a+3d+6c+2b*6c	7	0.295	0.721	9	0.664	1.391
2a+3a+3d+6c+3a*6c	7	0.311	0.693	9	0.656	1.396
2a+3a+3d+6c+3d*6c	7	0.311	0.693	9	0.656	1.396
all	7	0.295	0.721	9	0.664	1.391

Notes: n = number of points where that rule applied (either as a miss or falses alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630\ )^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|\text{Deviation}| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 4.31 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown). Note that in comparison with the previous melody, "Three Blind Mice" , the correlations between Rule 6c (Time Pattern Parallelism A) and Rules 2b (Attack-Point), 3a (Register) and 3d (Length) are not much higher for the second melody, and yet the rule is much more effective. This difference may arises because Rule 6c in this melody was constrained on the basis three rules (Rules 2b, 3a or 3d) and those three rules (Rules 2b, 3a and 3d) themselves are not highly correlated.

Table 4.31: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism A.

	Rule							
	2b	3a	3d	4	6c	2b*6c	3a*6c	3d*6c
Stage 6	.867**	.313	.367*	.876**	.486**	.876**	.417*	.391*
Stage 8	.883**	.310	.334*	.882**	.483**	.879**	.414*	.357*
Rule 2b		.292	.386*	.988**	.501**	.994**	.386*	.411*
Rule 3a			.178	.412*	.510**	.326	.910**	.199
Rule 3d				.467**	.479**	.406*	.247	.997**
Rule 4					.568**	.989**	.493**	.493**
Rule 6c						.507**	.618**	.474**
R 2b*6c							.424*	.433*
R 3a*6c								.269

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6c and its interaction with Rule 2b are helpful for predicting boundaries.

Time Pattern Parallelism B (Rule 6d)

For "Softly Now the Light of Day", the value of the best example of Time Pattern

Parallelism B (Rule 6d) at each point in the melody is presented in Figure 4.18 (third panel). The empirical boundaries are also shown for comparison (first panel), along with Rules 2 and 3 (second panel; Rules 2 & 3 for Stages 6 & 8 are combined in one panel), which are necessary for Time Pattern Parallelism B to operate. For this melody, like the previous Time Pattern Parallelism A, the best match between the assessed parallelism and the empirical data was fairly obvious. The correlations between Rule 6d and the empirical boundaries as a function of note span are provided in Table 4.32, along with the associated error measure. Based on these data, the note span of 10 notes was considered optimal.

Table 4.32: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism B.

Note Span	Stage 6	Stage 8	SS_{error}
10	.555**	.541**	1.176
9	.554**	.539**	1.184
8	.533**	.542**	1.294
7	.533**	.542**	1.294
6	.212	.227	1.380
5	.232	.249	1.373
4	.232	.249	1.373
3	.232	.249	1.373

Notes: * p < .05
** p < .01

In the third panel of Figure 4.18 observe that, unlike the previous Time Pattern Parallelism A (Rule 6c), Time Pattern Parallelism B (Rule 6d) does not capture all the strong empirical boundaries (i.e., Note 9), though it does capture more than the previous Pitch Pattern Parallelism A and B (Rule 6a & 6b). On this basis, one would expect Time Pattern Parallelism B to perform at an intermediate level.

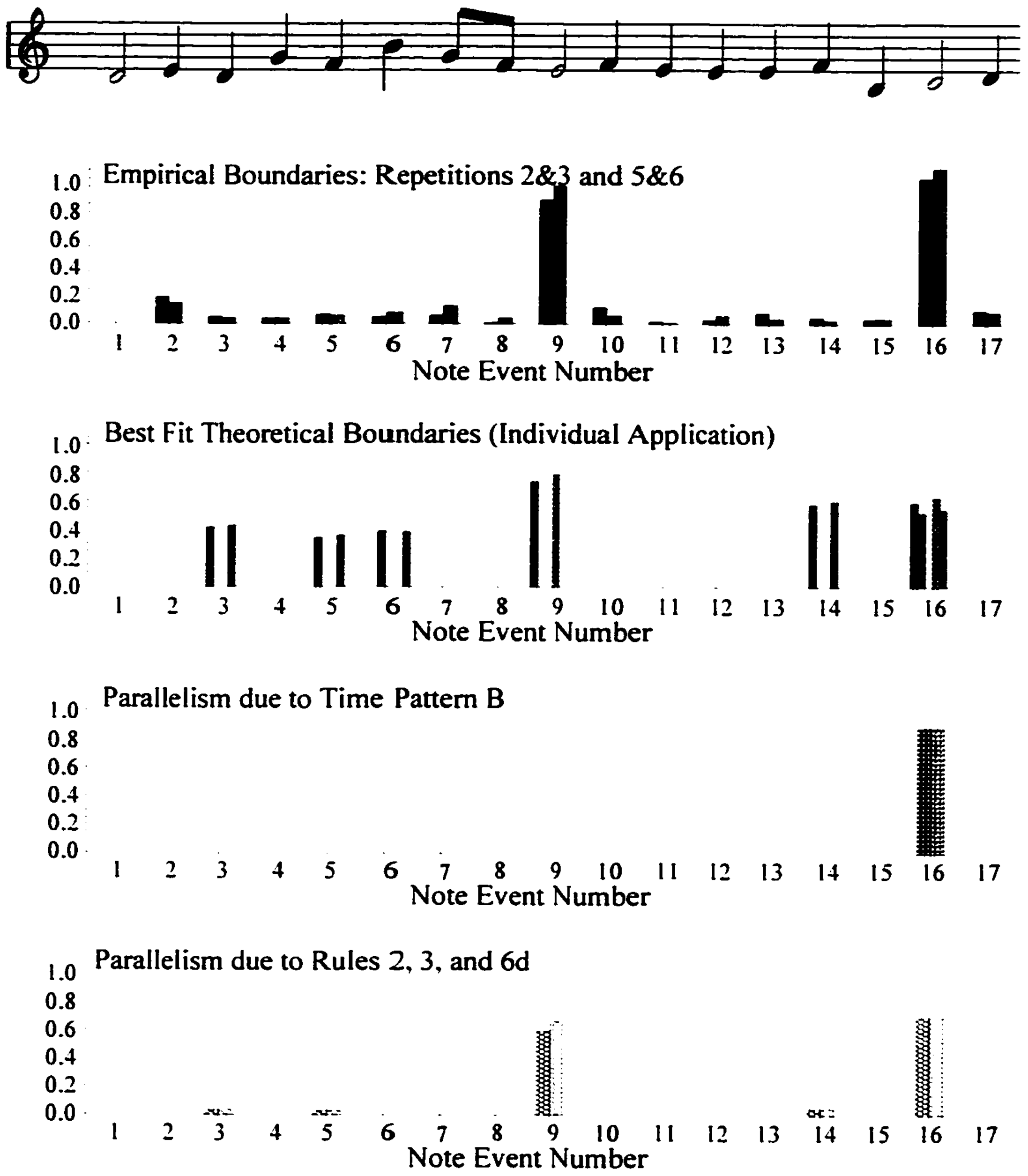


Figure 4.18a: The first part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6d. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ Rule 2b
▨ Rule 3a
▨ Rule 3d
■ Rule 6d
▨ Rules 2b, 3a 4, 6 and interactions
■ Stage 6 (Presentations 2 & 3)
■ Stage 8 (Presentations 5 & 6)

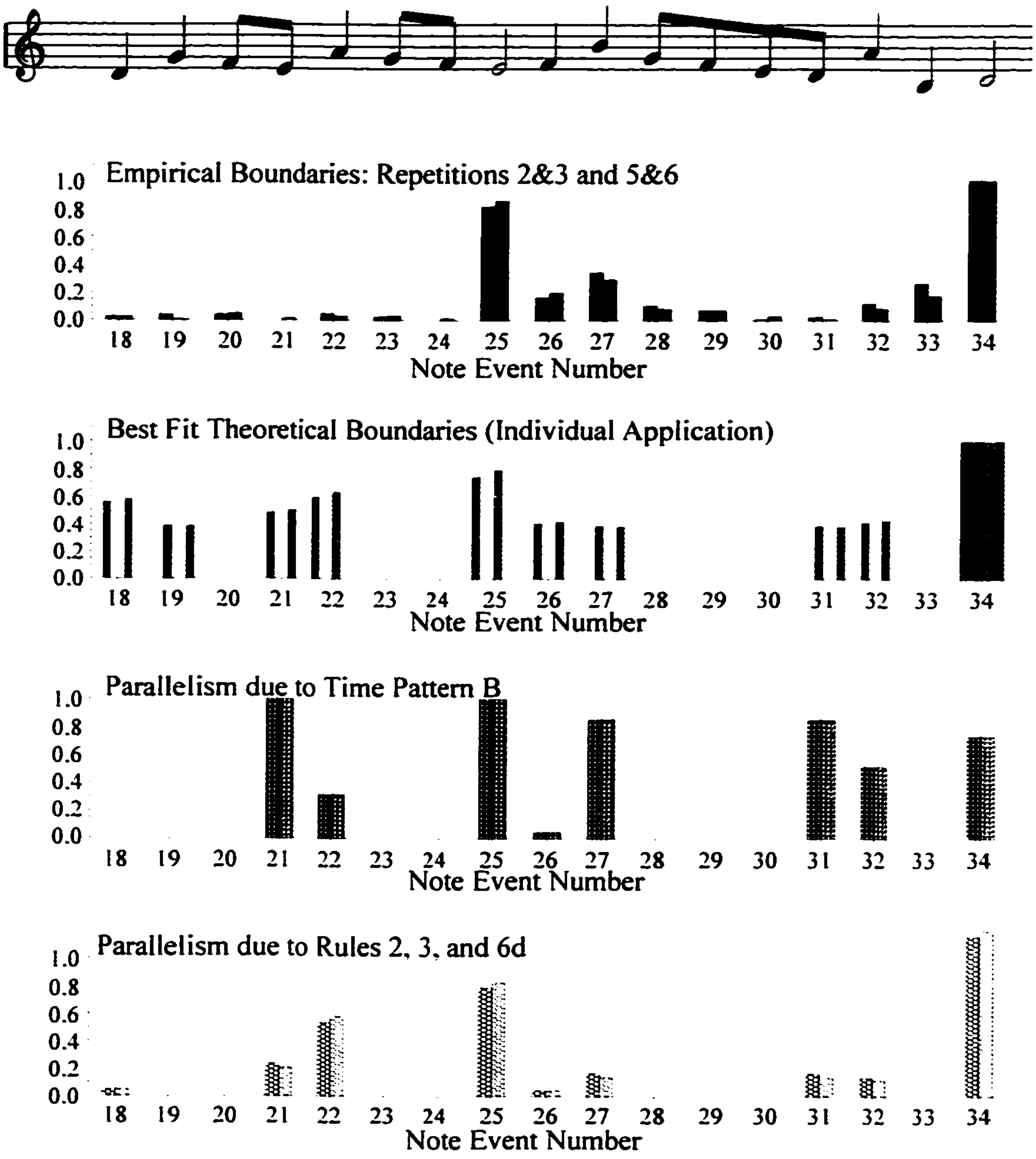


Figure 4.18b: The second part of the melody used in Stages 6 & 8, Part 1, the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules 2b, 3a and 3d and Rule 6d. This can be compared to the location of the of subjective boundaries based upon the average of Repetitions 2 and 3 and average of Repetitions 5 and 6.

▨ **Rule 2b** ■ **Rule 6d** ■ **Stage 6 (Presentations 2 & 3)**
▨ **Rule 3a** ▦ **Rules 2b, 3a 4, 6 and interactions** ■ **Stage 8 (Presentations 5 & 6)**
▧ **Rule 3d**

In the analysis of Time Pattern Parallelism B (Rule 6d), Rule 6d was entered in conjunction with Rules 2 and 3, because Rule 6d is built on putative boundaries identified by Rules 2 and 3. The interactions of Rule 6d with Rules 2b, 3a and 3d were also computed. Each interaction was considered separately as an addition to the model containing Rules 2b, 3a, 3d and 6d, and then all rules and then the model containing all rules and their interactions was examined. Table 4.33 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 4.34 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged). Note that the combination of all rules and interactions did provide substantial improvement overall (17.52% & 15.94%) that is greater than that of the previous Time Pattern Parallelism A. This improvement was due, largely it seems, to a reduction in the false alarm rate.

In this analysis, it is apparent that Time Pattern Parallelism B (Rule 6d) did add predictability to the model beyond that which could be explained by the low level rules. Unlike Time Pattern Parallelism A (Rule 6c), this predictability was due entirely to the rule itself: The interaction terms contributed nothing. As with Rule 6c, Rules 3a and 3d (particularly 3d) were dropped as the analysis proceeded because Rule 6d could explain that proportion of variance better.

Table 4.33: The Combination of Rules 2, 3 (Rule 4) and 6d for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.093*2b+0.115*3a+0.116*3d	1.467
2b+3a+3d+6d	16	b=0.953*2b+0.059*3a+0.000*3d+0.195*6d	1.210
2b+3a+3d+6d+2b*6d	16	b=0.953*2b+0.059*3a+0.000*3d+0.195*6d + 0.000*2b*6d	1.210
2b+3a+3d+6d+3a*6d	16	b=0.953*2b+0.059*3a+0.000*3d+0.195*6d + 0.000*3a*6d	1.210
2b+3a+3d+6d+3d*6d	16	b=0.953*2b+0.059*3a+0.000*3d+0.195*6d + 0.000*3d*6d	1.210
all	16	b=0.953*2b+0.059*3a+0.000*3d+0.195*6d +0.000*2b*6d+0.000*3b*6d+0.000*3d*6d	1.210

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	6	0.500	0.841	10	0.551	1.118
2b+3a+3d+6d	9	0.268	0.734	7	0.527	1.069
2a+3a+3d+6d+2b*6d	9	0.268	0.734	7	0.527	1.069
2a+3a+3d+6d+3a*6d	9	0.268	0.734	7	0.527	1.069
2a+3a+3d+6d+3d*6d	9	0.268	0.734	7	0.527	1.069
all	9	0.268	0.734	7	0.527	1.069

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 4.34: The Combination of Rules 2, 3 (Rule 4) and 6d for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.189+2b+0.115*3a+0.060*3d	1.380
2b+3a+3d+6d	16	b=1.058*2b+0.061*3a+0.000*3d+0.166*6d	1.176
2b+3a+3d+6d+2b*6d	16	b=1.058*2b+0.061*3a+0.000*3d+0.166*6d + 0.000*2b*6d	1.176
2b+3a+3d+6d+3a*6d	16	b=1.058*2b+0.061*3a+0.000*3d+0.166*6d + 0.000*3a*6d	1.176
2b+3a+3d+6d+3d*6d	16	b=1.058*2b+0.061*3a+0.000*3d+0.166*6d + 0.000*3d*6d	1.176
all	16	b=1.058*2b+0.061*3a+0.000*3d+0.166*6d +0.000*2b*6d+0.000*3b*6d+0.000*3d*6d	1.176

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses				False Alarms	
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	7	0.486	0.868	9	0.650	1.214
2b+3a+3d+6d	8	0.319	0.769	8	0.613	1.206
2a+3a+3d+6d+2b*6d	8	0.319	0.769	8	0.613	1.206
2a+3a+3d+6d+3a*6d	8	0.319	0.769	8	0.613	1.206
2a+3a+3d+6d+3d*6d	8	0.319	0.769	8	0.613	1.206
all	8	0.319	0.769	8	0.613	1.206

Notes: n = number of points where that rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 4.35 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown). Note that all the correlations are relatively high.

Table 4.35: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism B.

	Rule							
	2b	3a	3d	4	6d	2b*6d	3a*6d	3d*6d
Stage 6	.867**	.313	.367*	.876**	.555**	.838*	.526*	.454**
Stage 8	.883**	.310	.334*	.882**	.541**	.832**	.525**	.416*
Rule 2b		.292	.386*	.988**	.485**	.866**	.503**	.455**
Rule 3a			.178	.412*	.339*	.367*	.665**	.256
Rule 3d				.467**	.395*	.419*	.335	.851**
Rule 4					.534**	.878**	.577**	.527**
Rule 6d						.610**	.652**	.390*
R 2b*6d							.576**	.483**
R 3a*6d								.396*

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6d is helpful for predicting boundaries, but this is due to, as in Rule 6b, its lack of false alarms in conjunction with its overlap with Rules 2 and 3.

General Conclusions and Discussion

The objective of this chapter was to attempt to refine the quantification of the rules stated and implied by Lerdahl and Jackendoff (1983) and to attempt to extend the analysis to the higher level rules of Intensification (Rule 4) and Parallelism (Rule 6).

The first subgoal was to find or demonstrate a "new" analytic method to properly weigh the different rules in a manner that was consistent with their role in the model. This goal was generally met. The use of asymmetric error terms allows for a fair comparison between rules, and more importantly, allows for a fair method for combining rules. However, it is acknowledged that many refinements or alterations can be made to the analysis. In these reanalyses of Rules 2 and 3, Rule 2b (Attack-Point changes) was retained as the single most important rule. Rules 3a (Register changes) and 3d (Length changes) were much less important. One might be tempted to conclude that Rules 3a and 3d (particularly 3a) can be dropped from the repertoire, but this would be premature based on only two melodies.

The second goal was to attempt to quantify intensification and parallelism and to apply that quantification to the melodies used herein. Again the objective was met. Using asymmetric error terms, it was possible to find the optimal combination of Rules 2 and 3 for predicting boundaries (i.e., Rule 4: Intensification), granting more importance to misses than to false alarms. These results re-emphasize the dominant importance of Rule 2b. When used in combination with Rule 2b, Rules 3a and 3d were generally reduced in importance relative to their use in isolation. That is, even though the asymmetric error did not care about overprediction (false alarms), Rules 3a and 3d did not add predictability beyond that which could be explained by Rule 2b.

Four different versions of parallelism were analyzed, all based in the literature (see Chapter 1 of this work). Each type of parallelism had something to offer, though not necessarily in both melodies. To compare the different rules for parallelism within the different melodies, Table 4.36 provides the percent improvement for each rule (all are compared to the optimal combination of Rules 2 and 3, that is, Rule 4). Note that Rules 6a and 6d are approximately equally effective in both melodies, while Rule 6b is only useful in "Three Blind Mice" (a simple, very familiar, nursery rhyme) and Rule 6c is only useful in the extract from "Softly Now the Light of Day" (a more complex, unfamiliar melody). The flip-flop of these two rules may imply that different musical styles use different rules, or it may imply that a new version of these rules is needed.

Table 4.36: The percentage improvements within each stage for each Parallelism rule.

Rule		Three Blind Mice	Softly Now the Light of Day	
		Stage 4	Stage 6	Stage 8
Pitch Pattern Parallelism B	6a	7.63	5.86	2.97
Pitch Pattern Parallelism B	6b	9.77	0.07	0.00
Time Pattern Parallelism B	6c	1.69	13.49	12.83
Time Pattern Parallelism B	6d	17.18	17.52	15.94

Since this was a first attempt at such an analysis, many refinements and alterations are yet to be explored. Some of these will be mentioned. Firstly, in the analysis of parallelism, there were several constraints: minimum numbers of notes per subsequence; requirement for adjacent sequences, requirement that both subsequences have a potential boundaries at the end.

Of these constraints, the demand that both of the subsequences end on potential boundaries (as defined by Rules 2 and/or 3) seemed the most restrictive. Rule 6 in all its forms (6a, 6b, 6c & 6d) was so restricted because Rule 6 was conceived to act in a manner that boosted (existing) boundaries, rather than created boundaries. The rule was implemented because it seemed consistent with the construction of Lerdahl and Jackendoff (1983). Implementing the rule was trivial: Rule 6 (all forms) did not apply at any particular note of the melody unless that particular note had some potential for a boundary (based on Rules 2 and/or 3). Recall that this note of the melody marked the end of the second subsequence. Rule 6 (all forms) did not apply unless there was a potential boundary at the end of the first subsequence being compared (the position of this boundary varied depending on the beat- or note-span being used). Because this seemed to be the most restrictive, the entire set of analyses presented above were repeated with this constraint lifted. In effect, it was decided that Rule 6 (all forms) could induce a boundary at some note location even if that location did not have a potential for a boundary based on the low level rules (i.e. Rules 2 and/or 3). There still had to a be a boundary at the end of the first subsequence being compared. This restriction was retained under the assumption that subjects would not be able to retroactively impose a completely new boundary where none had existed before. That is, the fact that the current subsequence is parallel in structure to a previous subsequence can help to create a boundary at this location, but not to reinforce previous locations (this constraint might be lifted for analyses of the learning of melodies). The results were very messy. Essentially, when the constraint was lifted, so much parallelism was detected (remember that all analyses used continuous scales) that there resulted in a potential boundary on every note. This was not amenable to analysis. However, once other details are resolved, it might be possible to return to this with new insight, particularly regarding some minimum value of the correlation before any similarity is considered (i.e., a lower cutoff on the correlation).

The second major constraint required that subsequences be adjacent. This was simply to make the analysis tractable. Once a few more melodies have been analyzed, it should be possible to lift this restriction because more will be known about what constitutes parallelism (i.e., empirically driven modifications of Rule 6 in all its forms). For example, it might be desirable to capture the parallelism in "Softly Now the Light of Day" that is between sequences at Note Events 6-9 and 22-25 or sequences at 14-16 and 32-34. In "Three Blind Mice", the restriction of adjacent sequences meant that Pitch Pattern Parallelism could not detect the rather obvious parallelism of Note Events 46 with 7-10. This restriction was used in this work because it raises the complication of the number of notes (or amount of time) between the two sequences being compared. This requires more knowledge about memory for sequences over intervening melodies (or notes).

One other main concern of the analysis is the use of the correlation coefficient to assess parallelism. I did not want to use the correlation coefficient for one major reason. If two subsequences were to have constant pitch (e.g., d-d-d-d and e-e-e-e) the correlation would be zero. More importantly, if only one of the two subsequences was to be of constant pitch (e.g., d-d-d-d and d-e-f-g) the correlation would also be zero. Many would like to grant some notion of parallelism to these structures (e.g., d-d-d-d is more parallel with d-d-e-e than with d-g-c-a). Unfortunately, there is not a good pattern measure that will capture what the correlation does while also capturing the parallelism in the previous structures. The cosine was considered, but the cosine measure requires a ratio scale, and the cosine is affected by the distance of the values from the zero point. If the cosine is normalized to remove the effect of the zero point, then the cosine becomes the correlation. Other non-pattern measures of similarity were also attempted (e.g., the Euclidian, the city-block) but these measures are not about patterns and did not capture the structures of interest. The search for an appropriate measure continues[49].

A final point concerns the implementations of the rules themselves. Essentially, in this work the scaling of the rule strength to boundary formation was considered constant throughout the melody. In simple short melodies, such as were used in this work, this seems like a reasonable choice. However, it might need to be modified for more complex (orchestral) pieces. That is, a half note within a stream of sixteenth notes (Rule 2b: Attack-Point) would seem to be much more important than a half note with a stream of quarter and half notes. That is, the scaling and the rules should be altered to represent not so much an absolute scaling for an entire melody, but rather a relative scaling for the last portion of the melody. For example, Rule 2b could be recast to analyze the length of current note (or the change in attack points) relative to the previous *n* notes. The variable *n* might then be linked to training or other experience with music. Similar constructions could be attempted for each of the rules. This might serve better to model the on-line perception of music, which in term, might lead to better fits between the model and the data.

As to the results themselves, they are encouraging, but I do not think that it is appropriate to make strong claims regarding the relative utility of the different rules on the basis of only two melodies. The results do indicate that different melodies may use

---

[49] I am looking for [working on] a 3 dimensional version of the correlation that would be capable of accomplishing that task.

different structures. As such, many more simple melodies should be explored. The results are also suggestive for other tests of parallelism. That is, these results highlight the need to test the perception of parallelism within limited quantifiable structures. Different types of quantifications could be used to address issues raised herein. Similarly, to be able to test for the AA parallelism in an ABAC structure, it is necessary to set some limits on the number of notes that can be in B. At this point, I would argue that an extension to more complex music should be avoided until there is more information about the quantitative aspects within simpler musical materials. This is based on the results of various alterations to the algorithms that assessed parallelism. The chosen algorithm was restrictive because it was found that relaxing those restrictions lead to so much detected parallelism that subsequence of the melodies could be found to be parallel with some other subsequence of the melodies.

Generally, the results were informative, and also suggestive of many new directions for research.

CHAPTER 5

Experiment 2: Introduction

Overall , the goal of this thesis was the empirical assessment of the location of boundaries in a melody and the relation of those boundaries to theoretically derived notions of that location. In this regard, the previous experiment was successful (see Chapters 3 and 4 of this work). However, pertaining to the results of the previous experiment, there was one serious caveat that had to be addressed. The previous experiment assessed the efficacy of boundaries using a recognition-memory task. As such, demand characteristics must be considered (Orne, 1962). It is possible that subjects -- cognisant of the subsequent memory task (which they must have been after practice trials were completed) -- utilized a particular strategy for processing the stimulus: a strategy that imposed boundary formation even though it is not a usual aspect of musical processing.

The seriousness of this concern cannot be assessed within the previous design[50]. Hence, the present experiment was intended mainly to address that concern. The intent was to remove the incentive for a memory-based strategy by removing the memory component of the design. To accomplish this end, one could simply have removed the boundary efficacy part of the design. That is, the experiment would be retained except for the removal of Part 2 from Stages 2, 4, 6 and 8. However, instead of eliminating the boundary efficacy test, a different test was employed. That is, the basic designed was retained, but the assessment of efficacy was altered to one that did not rely on memory. Essentially, the memory-recognition test was changed to a click detection test in which the subject must detect a click while listening to a melody. This simple change completely removed any incentive to treat the melody as a sequence of notes to be remembered while retaining some measure of boundary efficacy.

Click detection experiments have been used in both music and speech studies

---

[50] On might argue that the inter-subject consistency argues against such a notion -- it seems unreasonable that all subjects opportunistically used the same strategy and created the same boundaries -- but this is not compelling.

(e.g., Garrett, Bever & Fodor, 1966; Gregory, 1978; Sloboda & Gregory, 1980; see also Clark & Clark, 1977) for the assessment of putative boundaries between units. Typically subjects must detect the presence of a click while processing the stimulus of interest (e.g., music or speech). The premise of this dual task experiment (cf. Berent & Perfetti, 1993; see also Kahneman, 1973, cited in Berent & Perfetti) is that participants must split their available processing resources between the two tasks. Subjects must listen to the music (perhaps an automatic process), and subjects must detect the click (a conscious process). As the demands on the primary task increase (processing the music), the resources available to the secondary task are decreased (click detection). If boundaries represent the closure of a unit in the music, then it is argued that in the moments prior to closure, resource utilization is maximized. That is, the processing associated with closing the unit and moving it to memory takes additional resources. Hence, at this point, few resources are available for the click detection task and performance on that task suffers momentarily. In fact, it is arguable that as information pertaining to a particular musical unit accrues, more resources are consumed by the processing of music (most easily considered in terms of short term memory constraints). Hence, performance on the secondary task (click detection), will show an increasing decrement from the beginning of a unit to the end of a unit. Although interesting, this concept should not be taken as binding at this point until more empirical data has been obtained because notions of processing load and memory constraints, particularly within music, are still not well defined. Of course, the greater the similarity of the resource demands of the two tasks, the greater the potential interference.

Gregory (1978) used a click detection task with music to demonstrate that subjects' responses tended to migrate to the regions where phrase boundaries should occur. Gregory presented subjects with six-note sequences, monophonically at either 200ms or 400ms per note (with an equal duration pause). One group of subjects was instructed to treat the sequence as two three-note subsequences and a second group of subjects was instructed to treat the same sequences as three two-notes subsequences (i.e., the only difference between the groups was the instructions). Clicks, centred on the presentation of the second, third, fourth or fifth note or on the pauses between those notes, were presented to the other ear. Of relevance to boundary location, the perception of clicks tended to migrate to the boundary between subsequences regardless of the group to which subjects were assigned. Hence, clicks prior to the boundary tended to be perceived late and clicks after the boundary tended to be perceived early. It must be remembered that early and late are defined relative to some normative value -- usually the mean of reaction times.

Sloboda and Gregory (1980) used a similar format, but with longer sequences designed to assess the reality of phrase endings. Tonal sequences of 20 or 21 notes were designed as two phrases. The phrase marker could be physical (a longer note[51]), structural (harmonic motion), both or neither. Results were similar to those of Gregory (1978) in that the clicks tended to migrate to the theoretically notated phrase boundary. Predictably, results were strongest for the structural plus physical combination and essentially non-existent for the neither condition. Another observation was that of a greater similarity between the physical only and combination profiles than between the structural only and the combination profile. This offers some indirect support for the notions of Lerdahl and Jackendoff (1983), in that the acoustic features tended to be the more important (in this case, the equivalent of Rule 2b: Attack Point Change). On the other hand, note that structural features in isolation were not ineffective. Other observations indicated that the migration effect was less pronounced after the boundary than before it (where the boundary was defined to be between the last note of the first phrase and the first note of the second phrase).

In the studies of both Gregory (1978) and Sloboda and Gregory (1980), subjects actually indicated the point when they thought the click had occurred by reference to a simple score (i.e., subjects had the musical score in view). In this regard, subjects may

---

[51] This corresponds roughly to the Group Preference Level of Lerdahl and Jackendoff (1983). However, although this might seem like an application of Lerdahl and Jackendoff's Rule 3d (Length Changes), it actually represented an application of Rule 2b (Attack Point Change) because there was only one lengthened note.

have been influenced by the visual pattern, and more importantly, by translation of the auditory memory to a spacial representation. That musically trained and untrained subjects did not differ in performance provides some confidence that the visual pattern was not major factors in performance. However, more importantly, such a design limits the stimulus to a fairly short (and possibly, not too complex) melody.

A second concern with both studies is that boundary locations were defined by theoretical notions of where those boundaries should occur. Though few would quibble with the locations as defined in either experiment, there is some concern that subjects may not have heard the melodies as intended. This is particular true for Gregory's (1978) study in which subjects were instructed to use a particular phrase structure in the complete absence of acoustic indicators. Though less important, it is also a concern in the study of Sloboda and Gregory (1980) for that condition that contained few or none of the usual cues marking a phrase boundary (i.e., the neither structural nor physical markers). In addition, both designs suffered from the artificiality of having only one phrase boundary (i.e., there were only two phrases).

Stoffer (1985, Experiment 2) actually assessed reaction times to clicks positioned at boundaries (defined on the basis of music theory) or off boundaries in binary and trinary forms. Reaction times were fastest for clicks on boundaries. Berent and Perfetti (1993) used a somewhat different approach. Three clicks were placed within fairly complex tonal pieces (details were not provided, but the single example was not a simple melody). Some pieces contained a unprepared chromatic modulation. The interesting comparison was for clicks before the modulation, clicks at the modulation and clicks after the modulation. When results were scored for reaction time, results indicated to main points: Reaction times decreased for each successive click (this was collapsed over 16 different stimuli) and reaction times at corresponding points were slower when the piece contained a modulation, but only at the modulation and after the modulation. That is, the increased cognitive load imposed by the modulation delayed detection of the click. Most importantly, this effect was consistent for both musicians and non musicians, though it was more pronounced for the musicians. The work of Stoffer, as well as that of Berent and Perfetti demonstrated that reaction times to clicks can be used within a musical stimulus. Berent and Perfetti demonstrated that multiple clicks can be used within a melody.

Furthermore, Stoffer (1985) implied that reaction times should increase off boundaries while Berent and Perfetti (1993) implied that reaction times should increase with processing demands. Together with Gregory (1978) and Sloboda and Gregory (1980), one can generally conclude that reaction times should increase with processing demands. Given the somewhat contradictory results of Gregory and Sloboda and Gregory versus Stoffer, it is difficult to say, a priori, whether such reaction times should be higher on or before boundaries. However, given the higher similarity of the stimuli of Sloboda and Gregory to those used in the current study, it would seem prudent to predict results that align with Sloboda and Gregory.

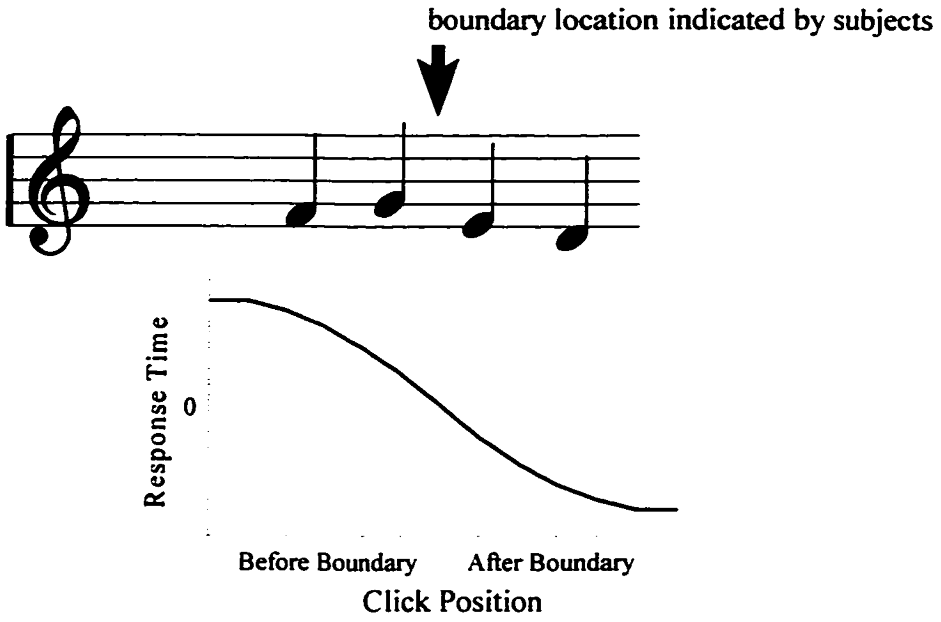
In the present experiment, a variation of these tasks was used to assess boundary efficacy. This assessment was the only change from the previous experiment (see Chapter 3 of this work). Overall, the entire experiment consisted of an eight stage experiment. Stages 1, 3, 5 and 7 used a modified probe-tone task to assess the internal representation of tonality. Stages 2, 4, 6, and 8 assessed boundary location and boundary efficacy in a two part procedure. The first part assessed boundary locations using a key-press task (hereafter: *boundary location*). The second part assessed the impact of such boundaries on subsequent processing using a click detection task (hereafter: *boundary efficacy*).

In the boundary location part, subjects heard a melody and indicated the position of boundaries by a simple key press. Subjects were requested to make their units as small as is meaningfully possible. For each subject, this resulted in a boundary profile -- a point-to-point map of the locations of boundaries -- for the melody. Each individual boundary profile is a binary representation of the location of boundaries. The underlying rationale for each aspect of this design and the associated analyses have been stated previously (see Chapter 3 of this work) and so will not be repeated. However, note that the analysis used only a selected subset of the analyses (of boundary location) used previously.

The second part (boundary efficacy) assessed the role of boundaries on processing by measuring processing demands. Subjects heard the same melody, but were asked to detect (as quickly and accurately as possible) clicks placed within the melody. In each melody, there were eight clicks, spaced randomly with constraints. By design, clicks were defined by reference to the boundaries that the subject had actually indicated. One half the clicks were supposed to fall on the boundaries that the subjects had indicated in the previous rendition of the melody. The other half were placed between boundaries. Click positions were not defined relative to theoretical notions of phrase structure. If subjects did not provide four boundaries on which to place clicks, then the ratio of clicks on and off boundaries was adjusted. In addition, a minimum distance of two note events was required between clicks.

Eight clicks were chosen as a compromise between the desire to have enough data to analyse and the desire that the clicks be distant from each other. In this regard, the results of Sloboda and Gregory (1980) were used: Particularly, in that work, it was observed that post boundary effects decayed quite quickly. Based on previous work (Gregory, 1978; Sloboda & Gregory, 1980), one would generate the basic prediction that the detection of clicks before boundaries would be delayed while detection of clicks after boundaries would be advanced. Extending the notion to more than one click per stimulus, one might reasonably expect a pattern like that shown in Figure 5.1. That is, to meet the simple predictions, one must get from the faster detection just after a boundary to the slower detection just before a boundary. There are several possibilities: A smooth linear transition, a non-linear smooth transition (i.e., a monotonic, non-linear function), a step function or something else. Since there is no data available to make more precise predictions, the analysis was designed to look for the simplest cases first, and extend to more complex analyses in a post hoc fashion.

A: From Gregory (1987)



B: a plausible extentions

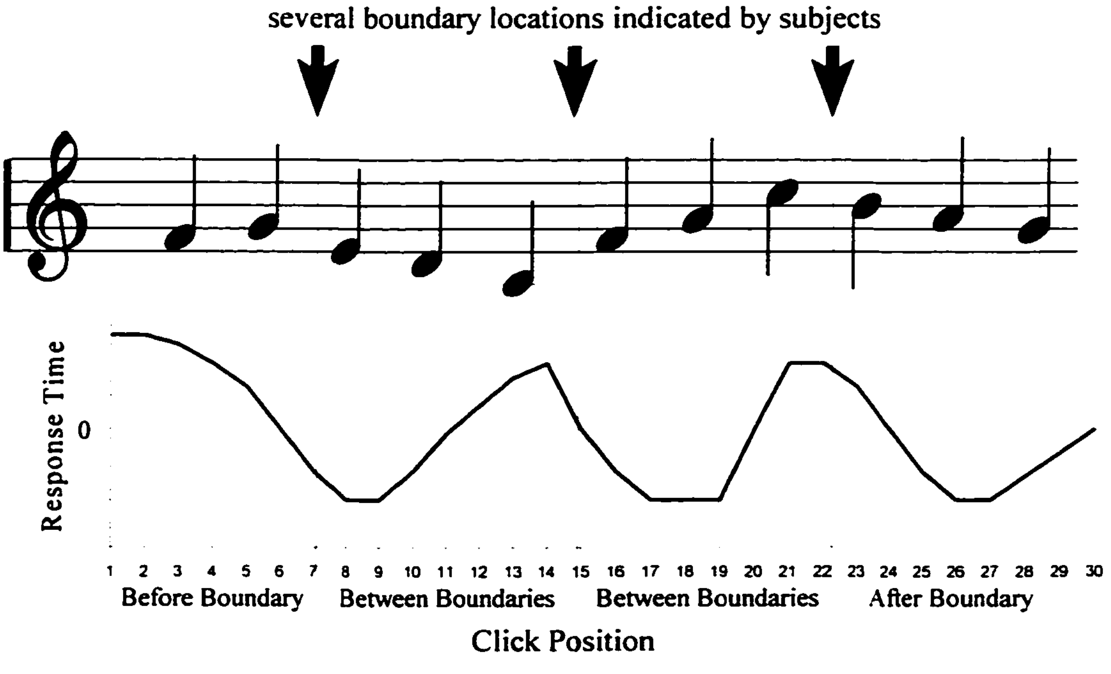


Figure 5.1: Possible expected response times for click detection given the relative positions of boundaries and clicks. Essentially clicks just after boundaries should be detected more quickly while those before boundaries should be detected more slowly. However the transistion between boundaries may take several forms and the relation between reaction times on and off boundaries is somewhat ambiguous.

Because it would not be possible to indicate the position of multiple clicks within longer melody by reference to a visual analogue, subjects indicated the detection of a click by a key press, which was timed from the onset of the previous click. That is, subject heard the melody with its embedded click and pressed a key whenever a click was heard.

The present experiment used the same stimuli as the Experiment 1. Although as a whole, the Experiment 2 would seem redundant, it was thought that the issue of demand characteristics producing artifactual results in Experiment 1 was simply too important to ignore.

In summary, the previous experiment was repeated with a change only in the boundary efficacy part. It was expected (hoped) that the results would essentially mimic those of the previous experiment thereby negating any concerns of artifactual results.

Method

Subjects

Sixty-two subjects (39 females and 23 males) having a mean age of 21.15±4.87 years (range: 17 - 42; see Figure 5.2) were recruited from the university community, primarily from the departments of psychology (i.e., the subject pool) and music (musicians). Subjects either volunteered, participated for class credit or financial compensation for the one hour experiment. No subject participated in both Experiment 1 and Experiment 2. Of the males, one was ambidextrous, 2 were left handed and 21 were right handed: Of the females, 3 were left handed and 36 were right handed. The clustering of ages in the low twenties is typical for a university based population. One subject reported absolute pitch. Because the basis for the assignment of subjects to groups was to be a cluster analysis, no rigid selection criteria were imposed, although a wide range of training was desired.

Several measures of involvement with music were obtained using the same self-report questionnaire of the previous experiment (see Table 5.1).

Table 5.1 Training, Practicing and Playing Variables Retained for Analysis.

Training	TotalInst	total hours of instruction (hours)
	MaxCatInst	maximum total hours of instruction for any single category (hours)
	RecInst	the recency of instruction (years before experiment)
	MaxSinInst	highest intensity of instruction on a single instrument (hours/week)
Practicing	TotalPrac	total hours of practice (hours)
	MaxCatPrac	maximum total hours of practice for any single category (hours)
	MaxSinInst	highest intensity of practice on a single instrument (hours per week)
Grade	Grade	highest grade achieved (on any single instrument)
Playing	TotalPlay	total hours of playing (hours)
	MaxCatPlay	maximum total hours of playing within any single category (hours)
	RecPlay	the recency of playing (years before experiment)
	MaxSinPlay	highest intensity of playing on a single instrument (hours/week)

To reiterate briefly, for each instrument formally studied, subjects reported the date instructions commenced and ceased and Royal Conservatory of Music (RCM) grade (or equivalent) as well as the hours per week of instruction and the hours per week practicing. For these same instruments, subjects also reported the hours per week playing for those periods when they were not having formal instructions. Subjects also reported the same measures for all instruments played but not formally studied ("playing" could only occur when instruments were not studied; all playing time while receiving instruction was considered practicing). For each subject, for each instrument studied ("studied" is used as an umbrella term for instructions plus practicing) and played, within each category (see the text associated with Table 3.1 for details), the variables shown in Table 5.1 were retained.

Subjects reported an average of 2.48±1.51 instruments studied (minimum: 0.00, maximum: 6.00), but subjects only reported training (lessons and practicing) an average of 2.23±1.38 instruments (minimum: 0.00, maximum: 6.00). The difference is due mainly to the number of subjects who taught themselves the guitar. Subjects reported playing an average of 0.67±0.95 instruments (however, recall that a person who was currently studying an instrument would not be considered as playing that instrument). When considering number of instruments studied (lessons and/or played), 4 were not involved with any instrument, 14 were involved with one, 16 were involved with two, 13 were involved with three, 9 were involved with three, 3 were involved with four, and 3 were involved with six. For studying, 6 subjects did not study any instrument (hence, 2 subjects had taught themselves 1 or more instruments without any formal instruction), 15 studied one, 16 studied two, 12 studied three, 11 studied four, and 1 each studied five and six. Considering playing, 36 did not play any instrument, 14 played one, 9 played two, 2 played three, and 1 played four. Note that of all the subjects, only 4 had neither studied nor played an instrument and only 6 had never studied an instrument, but 36 did not play any instrument beyond the lessons that they had taken at some point.

The actual distributions for the various measures are shown in Table 5.2 and Figure 5.2. For comparison purposes, the TotalInst was equivalent to 16.31±24.20 years at one hour per week, or 5.44±8.07 years at three hours per week. Again, for comparison purposes, the TotalPrac was equivalent to 22.02±32.74 years at one hour per week, or 7.34±10.91 years at three hours per week. Note that the total times while training and the maximum intensities while studying are based on all subjects (*n*=62), but the time since last training includes only those subjects who actually studied an instrument (*n*=56). A total time of 0.0, or a maximum intensity of 0.0 are meaningful values for subject who never trained (i.e., no training or practice time), but the time since last training is not meaningful for a subject who did not train. The categories of woodwind (recorder, flute, clarinet and saxophone; n = 37), keyboard (mainly piano; n=34), fretted strings (mainly guitar; n=16) and fretless strings (mainly violin; n=11) were the most cited. Other categories were brass (n=8,) tuned percussion (n=1), untuned percussion (n=4) synthesiser (n=0), harp (n=0) and voice (actually 2 categories; n=12 in total)

Most subjects (43) did not report any RCM equivalent grade (though many of these had substantial training), three reported Grade 2, four reported Grade 3, one each reported Grades 4 and 5, two reported Grade 6, one each reported Grades 7 and 8, and three each reported Grades 9 and 10. Four subjects were in the first year of the undergraduate music program at Dalhousie: One subject had completed the undergraduate music program. Exclusive of those in the music program, four subjects reported that they were currently performing in public (whether or not the public would pay to hear was not discussed).

For comparison purposes, the mean playing was equivalent to 7.10±19.70 years at one hour per week, or 2.37±6.56 years at three hours per week. Note again, that the total time playing, and maximum intensity of playing include all subjects (*n*=62), while the time since last playing includes only those subjects who actually played some instrument (*n*=26: these are those who had played but never studied any instrument). The categories of keyboard (n=19), woodwind (n=9), fretted strings (n=6) were the most cited. Other categories were fretless strings (n=2), brass (n=3), tuned percussion (n=0), untuned percussion (n=2) synthesiser (n=0), harp (n=0) and voice (actually 2 categories; n=1 in total).

The subjects of Experiment 2 tended to have somewhat more training, and more playing and had played more recently. However, the distributions overlapped considerably (compare Figure 3.1 with Figure 5.2).

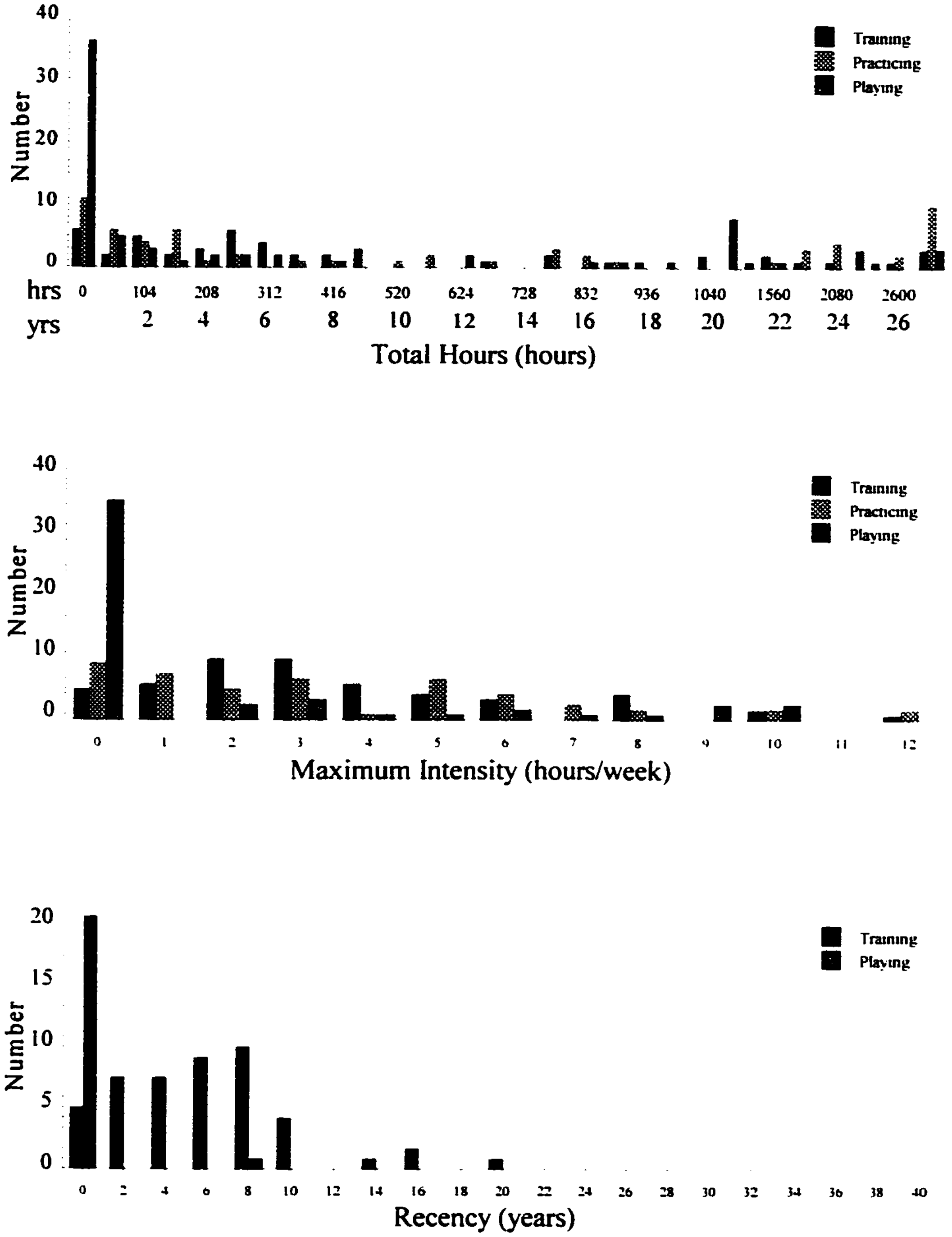


Figure 5.2: Hours of training, practicing and playing, maximum intensity of training, practicing and playing and recency of training (practicing) and playing. For hours of involvement, a second scale on the abscissa indicates the equivalent years of involvement at one hour per week.

Table 5.2: Training, Practicing and Playing over all subjects.

		Mean±SD	Min	Max	Units
Training	TotalInst	848.16±1258.14	0.00	8705.63	hours
	MaxCatInst	643.22±914.84	0.00	6015.32	hours
	RecInst	5.07±4.01	0.00	16.75	years
	MaxSinInst	3.69±3.05	0.00	15.00	hours/ week
Practicing	TotalPrac	1144.84±1702.71	0.00	8843.31	hours
	MaxCatPrac	975.43±1545.37	0.00	8497.23	hours
	MaxSinPrac	4.57±5.04	0.00	23.00	hours/ week
Grade	Grade	1.79±3.18	0.00	10.00	
Playing	TotalPlay	369.24±1023.31	0.00	6166.89	hours
	MaxCatPlay	306.47±871.03	0.00	3460.78	hours
	RecPlay	0.30±1.54	0.00	7.83	years
	MaxSinPlay	1.29±3.00	0.00	15.00	hours/ week

As in Experiment 1, several measures are correlated with each other (see Table 5.2). Firstly, total instruction (TotalInst) was highly correlated with the maximum instruction within any single category (MaxCatIns; r=0.74, p<0.01), total time practicing (TotalPrac) was highly correlated with maximum time practicing within any single category (MaxCatPrac; r=0.91, p<0.01) and total time playing (TotalPlay) was highly correlated with maximum time playing within a single category (MaxCatPlay; r=0.83, p<0.01). Obviously, as was concluded in Experiment 1, most people focus on a single instrument category (or instrument).

Table 5.3 presents the correlations between training and practicing, along with the correlations of those measures with number of instruments studied or played. The important point is that the different measures are related but not redundant. When interpreting these simple correlations, it must be remembered that age is an uncontrolled factor.

Table 5.3 The correlations between the various attributes of training, including the number of instruments trained and played.

	Training (Lessons)			Practice	
	TotalInst	MaxSinInst	RecInst	TotalPrac	MaxSinPrac
# Studied	0.651**	0.601**	-0.600**	0.501**	0.454**
# Played	0.553**	0.509**	-0.611**	0.526**	0.499**
TotalInst		0.978**	-0.357**	0.364**	0.326**
MaxSinInst			-0.375**	0.331**	0.316**
TotalPrac			-0.382**		0.981**
MaxSinPrac			-0.381**		

Notes: All correlations that involve Recency of Training and Recency of Playing are on 56 pairs: Other correlations are based on 62 pairs.
* $p<0.05$
** $p<0.01$

Table 5.4 presents the correlations between playing and other measures. Again, the different measures are related but not identical. Generally the pattern is the same as in Experiment 1.

Table 5.4: The correlations between the various attributes of playing and other measures of music study.

	Playing		
	TotalPlay	MaxSinPlay	RecPlay
# Studied	-0.180	-0.028	-0.350**
# Played	0.142	0.128	-0.478**
TotalInst	0.038	0.029	-0.290*
MaxSinInst	0.010	0.006	0.278 *
RecInst	0.101	0.112	0.093
TotalPrac	0.168	0.155	-0.237
MaxSinPrac	0.316**	0.112	-0.215
TotalPlay		0.994**	-0.447**
MaxSinPlay			-0.459**

Notes: All correlations that involve Recency of Training and Recency of Playing are on 26 pairs: Other correlations are based on 62 pairs.
* $p<0.05$
** $p<0.01$

These measures of training and playing were correlated against various demographic factors (Age, Sex and Handedness; see Table 5.5). The only correlations were between age and recency of training or playing.

Table 5.5: The correlations between the various attributes of music study and demographic variables.

	Age	Sex	Hand
Studied	-0.385**	0.199	0.006
Played	-0.248	0.047	-0.181
TotalInst	-0.209	0.047	0.101
MaxSinInst	-0.215	0.022	0.160
RecInst	0.611**	0.049	0.081
TotalPrac	-0.208	0.009	-0.022
MaxSinPrac	-0.208	0.012	0.039
TotalPlay	0.408**	-0.374**	0.025
MaxSinPlay	0.372**	-0.369**	0.218
RecPlay	-0.077	0.002	0.040
Age		-0.010	0.109
Sex			0.104

Notes: All correlations that involve Recency of Training and Recency of Playing are on 45 pairs: Other correlations are based on 61 pairs.
* $p<0.05$
** $p<0.01$

All the correlations are fairly similar to Experiment 1. The only difference is that for this group of subjects, Age was correlated with TotalPlay and MaxSinPlay.

Two supplementary measures of involvement with music were also obtained. The first was time spent listening to music "by choice" (Figure 5.3), which was 17.53±13.59 hours per week (minimum: 0.00, maximum 84.00! -- this number was checked, several times, for accuracy). The second was time spent listening to music "by default" (Figure 5.4) which was 9.83±12.82 hours per week (minimum: 0.0, maximum 56.00 -- a reasonable number when one considers that many commercial establishments play music all day, and many students work in such places). Again these measures compare favourably with Experiment 1.

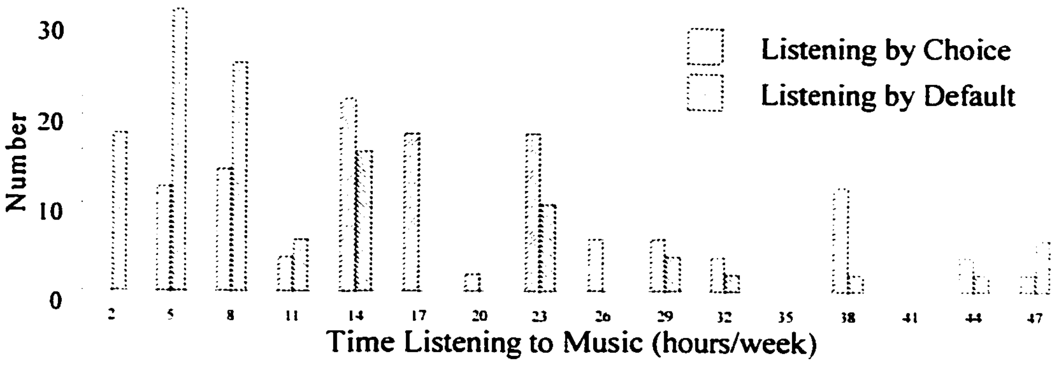


Figure 5.3: Hours per week spent listening to music.

Overall, the subjects in this experiment were very similar to the subjects in Experiment 1.

Procedure

Overall the experiment was presented as an eight stage experiment (see Figures 5.4 & 5.5). The basic design was essentially the same as that of the previous experiment (see Chapter 3 of this work) using the same design and stimuli, except in Part 2 of Stages 2, 4, 6 and 8. Stages 1, 3, 5 and 7 were designed to assess the internal representation of tonality by the subject using a modified probe tone task and as such were essentially the same, with only minor changes to the stimuli. Stage 1 was designated as practice for Stages 3, 5 and 7. The design is reiterated briefly.

Stages 2, 4, 6 and 8 were designed to assess the boundary locations of the subjects within particular melodies. Stages 2, 4, 6 and 8 were the same, with no changes to the stimuli and Stage 2 was designated as practice for Stages 4, 6 and 8.

Successively abbreviated instructions (the initial were quite detailed) were presented to the subject at the beginning of each stage, at a pace determined by the subject, via the computer. Additional abbreviated instructions remained on screen during the each stage. The experimenter remained with the subject during the practice stages (Stages 1 and 2) to insure, by direct observation of the subject, that the instructions were understood and to provide clarification if needed. The instructions provided for each and every Stage are included in Appendix: Instructions.

For all Stages of the experiment (approximately 45 minutes, exclusive of debriefing), subjects were seated comfortably in front of an IBM style computer in an Industrial Acoustics single-walled sound-attenuating room. To minimize noise, only the monitor and keyboard were kept in the room with the subject; the actual CPU and case was isolated from the subject by being outside the acoustics chamber.

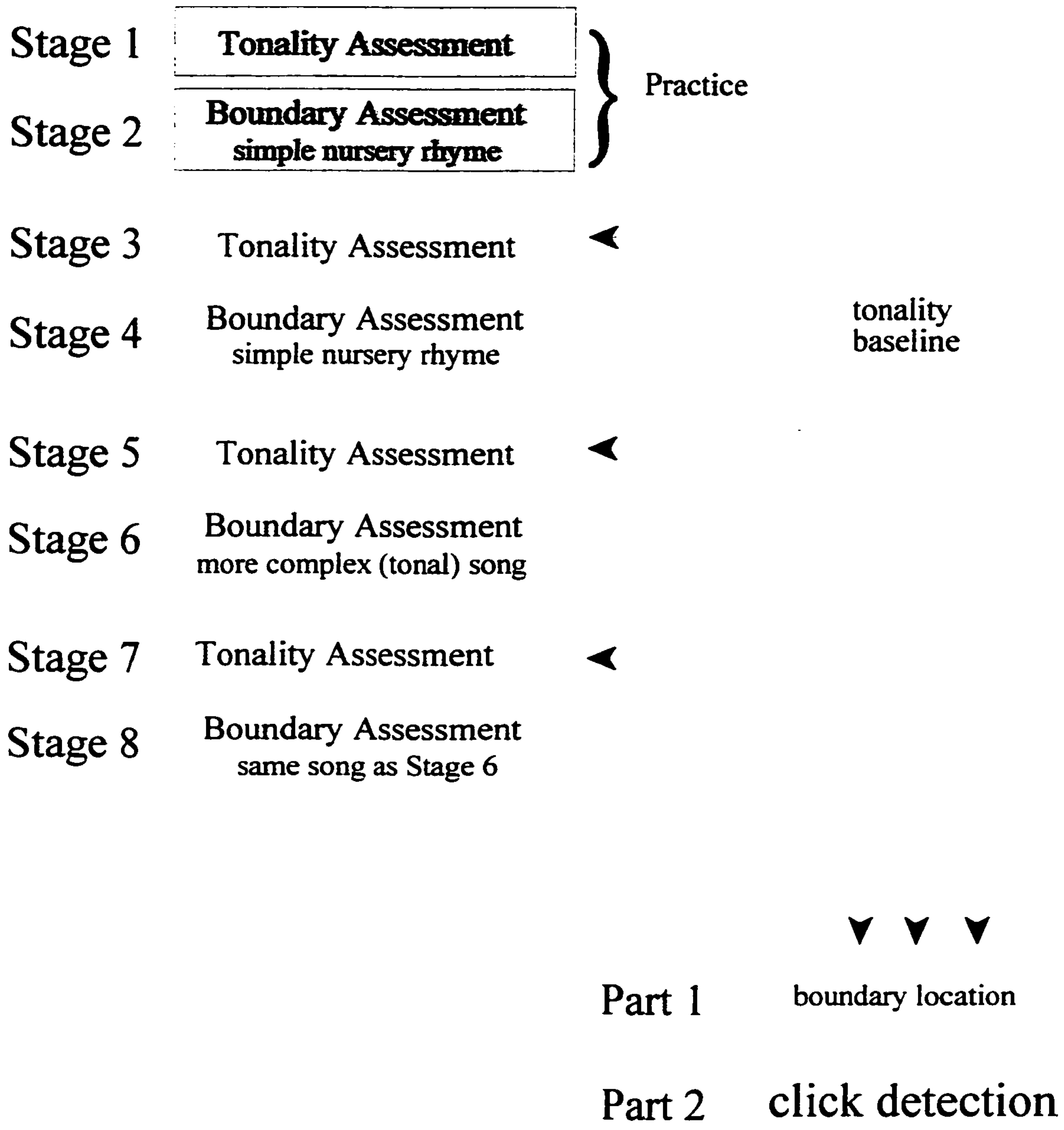


Figure 5.4: The overview of the eight-stage design. Note that stages devoted to boundary assessment are further subdivided into two phases. Click detection was the main change from the previous.

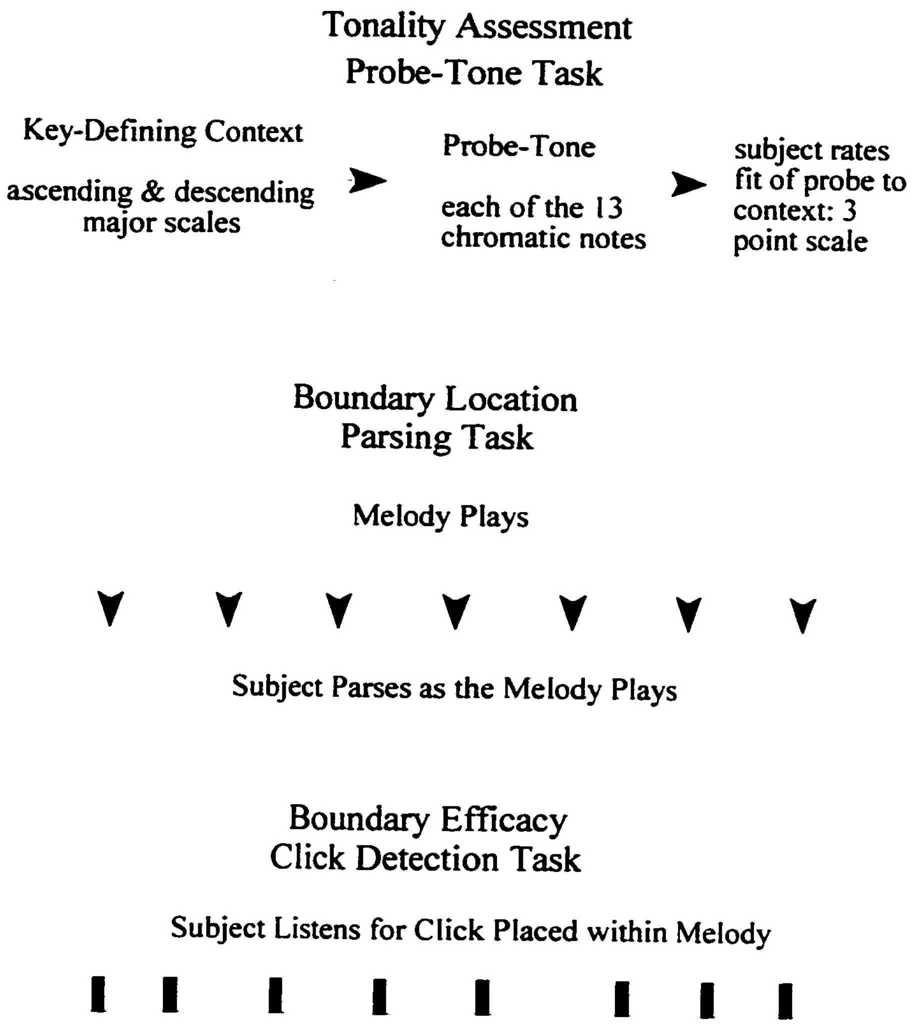


Figure 5.5: Basic design within each block. Note that the Tonality Assessment and the Boundary Assessment components are relatively independent, but Boundary Efficacy depends on Boundary Location. Hence, Tonality Assessment are considered separate stages, while Boundary Location and Boundary Efficacy are considered two phases of one stage.

Tonality Assessment

Stage 1 was designated as Tonality Assessment Practice while Stages 3, 5 and 7 were designated as Tonality Assessment 1, 2 and 3 respectively. In each of these, a modified probe tone task (Krumhansl, 1990; Frankland & Cohen, 1990) was used to determine the tonality profile of each subject. Subjects were presented with an ascending or descending major scale followed by a single probe tone drawn from the octave bounded by the scale. The scale defined a context and the subject rated the fit of the probe tone within that context on a three point scale (poor, indeterminant/uncertain, and good).

In a block of 26 trials, each of the 13 chromatic tones was presented once as a probe tone for both the ascending and descending contexts (i.e., each context -- probe-tone combination was presented only once). Within a given stage, all scale contexts were in the same key, but the presentation of the ascending and descending forms was a unique randomization for each subject, for each stage and for each block. The presentation of probe tones was also a unique randomization for each subject, for each stage, and for each block. Stage 1 (Practice) had only one block of trials while Stages 3, 5, and 7 each had two blocks of trials. Key was varied across Stages such that the key of the Tonality Assessment Stage matched the key of the melody presented in the subsequent Boundary Assessment Stage.

Responses were indicated by the use of the 1, 2 or 3 of the keypad or by the use of the left, center and down arrow keys; for left-handed subjects, the use of the 1, 2 and 3 of the top row of the standard QWERTY keyboard was also provided.

Boundary Assessment

Stage 2 was designated Boundary Assessment Practice while Stages 4, 6, and 8 were designated as Boundary Assessment 1, 2 and 3 respectively. In each Stage, there were two parts. The first part was designed to assess the location of boundaries within a melody: Subjects simply listened to the melody and indicated the location of boundaries. The second part was designed to test the efficacy of those boundaries as they pertain to the formation of informational units: Subjects were presented with a fourth repetition of the melody and asked to listen for (and to respond to as quickly as possible) clicks placed within the same auditory stream.

In the first part of each stage (boundary location), a melody was presented. Subjects were instructed to press the space bar (or any key of their own choosing) any time they felt that one section of the melody had ended and a new section of the melody had begun. Subjects were informed that there would be three trials using the same melody, and that each subsequent repetition was intended to allow them to refine their answers.

At the end of the first presentation of the melody, subjects were asked to rate the familiarity of the melody on a 10 point scale, with 1 representing completely unfamiliar and 10 representing completely familiar. After rating the familiarity, subjects were asked to provide the name of the melody, if known (subjects were also informed verbally that a line from the melody could be provided). Subjects then heard the melody and indicated boundaries for the second and third repetitions. Subjects initiated the presentation of each repetition of the melody at the time of their own choosing, but the timing within the melody was fixed for all subjects.

After the third repetition, subjects moved onto the click detection task, again at the time of their own choosing (instructions were presented before this part began). In this part of each stage (boundary test: click detection), subjects were presented with the same melody, but asked to listen for clicks within the melody. Eight clicks were placed within each melody, subject to constraints noted below. Subjects indicated the detection of a click by a simple key press as rapidly as possible.

Upon completion on the eighth stage, subjects completed the short questionnaire pertaining to musical background (see Appendix: Instructions), and were debriefed. The entire experiment last no more that 1 hour and generally no more than 55 minutes.

Apparatus and Stimuli

All tones within the entire experiment were created using the internal MIDI driver of a Creative Sound Blaster 16, housed within an IBM AT (80286, 12MHz) compatible computer. All tones were created using the default instrument 0, mode 0, corresponding to the sound of an acoustic piano. The same computer provided instruction via a B/W monitor (Amber) and recorded responses. Programs for the presentation of stimuli and recording of responses were written in-house, using Borland's Turbo C/C++, Version 3.0, aided by the Creative's Sound Blaster Developers Kit, Version 2.0. Tones were presented binaurally through a pair a Realistic LV 10 headphones connected directly to the audio output of the Sound Blaster 16 at a level considered comfortable by the subject. The monitor and keyboard were housed within the Industrial Acoustics single-walled sound-attenuating room, but the main computer was external to the sound-attenuating room in order to minimize noise.

Tonality Assessment

For Stage 1 (Tonality Assessment Practice), as shown previously in Figure 3.6, the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave C_4 to C_5. The context consisted of eight notes creating the ascending or descending C major scale, with each note having a duration of 400ms (approximating a tempo of 150mm). Probe tones consisted of each of the 13 chromatic tones, again with a duration of 400ms. In addition, there was a 400ms rest (gap) between the final note of the context and the probe tone. All tones were presented at the same volume.

For Stage 3 (Tonality Assessment 1), the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave G_3 to G_4 (see Figure 3.7). All other features remained as in Stage 1

For Stages 5 and 7 (Tonality Assessments 2 and 3), the basic stimuli consisted of 13 tones representing the equal-tempered chromatic scale within the octave C_3 to C_4 (see Figure 3.8). All other features remained the same as in Stage 1.

Boundary Assessment

In Stage 2 (Boundary Assessment Practice), the melody presented to subjects was "The Mulberry Bush", presented in the notated key of C (see Figure 3.9), with notes in the range G_4 to G_5, at a tempo of 135mm, in 3/4 time. All notes were presented at the same volume with precise computer controlled timing: No indications of beat or meter were manifest as departures from absolute timing or intensity of the notes. This melody had a moderate familiarity for subjects, achieving a mean rating (scaled from 0 to 10) of

5.98±3.16 (median: 7.00, mode: 8.00, minimum: 0.00, maximum: 10.00). Three subjects correctly labeled the song, one by name and two by citing a line of the melody. One other cited the line "this is the way we wash" and another cited the line "early in the morning" was are words to the second verse (one said "something from Haydn"). This melody could be said to be tonal.

During the assessment of boundaries, each response of a subject was tagged to the currently sounding note. That is, boundaries were associated with particular notes of the melody; boundaries were not associated with the gaps between notes (see Figure 3.10).

Clicks were placed within the melody online after the subject had indicated the placement of boundaries within the melody. Eight clicks were placed within the melody, and the intent was that four clicks be placed on randomly selected boundaries (if the subject had indicated 4 or more boundaries) and four clicks were placed randomly between boundaries. If the subject did not place four or more boundaries, or if those boundaries failed to be sufficiently distant in number of note events, then clicks were randomly placed within the melody. Clicks were dispersed throughout the melody, such that there was at least two note events between subsequent clicks. Responses to clicks were timed to the nearest 10 ms (a limitation of the equipment available, but a pretest with four pilot subjects, one being the experimenter, indicated that this would produce sufficient effects). Clicks were generated by the sounding of a percussion (program 10, mode 75) click. The click was initiated on a separate channel immediately after the onset of the corresponding note (i.e., the click was sounded while the note played). Clicks were played with the same volume setting as the melody. No subject complained of an inordinate difficulty in detecting clicks.

In Stage 4 (Boundary Assessment 1), the melody presented to subjects was "Three Blind Mice" (see Figure 3.11, or see Figure 5.8). The tempo at presentation was 145MM in 3/4 time. This melody had a high familiarity for subjects, achieving a mean rating (scaled from 0 to 10) of 9.18±2.36 (median: 10.00, mode: 10.00, minimum: 0.00, maximum: 10.00). Fifty-three subjects correctly identified the melody as "Three Blind Mice", two labeled it as "Hot Cross Buns", and one labeled it as "Frere Jacques". This melody could be said to be unambiguous in its tonal center. The click detection part was conducted in the same manner as Stage 2.

In Stages 6 and 8 (Boundary Assessments 2 and 3), the melody presented was "Softly Now the Light of Day" (see Figure 3.12 or Figure 5.12). In Stage 6, this melody was very unfamiliar to subjects, achieving a mean rating (scaled from 0 to 10) of 1.41±2.09 (median: 0.88, mode: 0.00, minimum: 0.00, maximum: 10.00). In Stage 8, the ratings improved to 3.77±3.50 (median: 3.00, mode: 0.00, minimum: 0.00, maximum: 10.00). No one identified the tune, although one subject though it to be "baroque". In Stage 8, three subjects explicitly stated that was the same melody as in the previous stage. Although the melody could be said to be tonal, it is more ambiguous with respect to its tonal center. The previous method of Stage 2 was used to test boundary efficacy.

Annotated Results

Of the eight stages of this experiment, Stages 1, 3, 5, 7 were concerned with the assessment of tonality, while Stages 2, 4, 6 and 8 were concerned with the assessment of subjective boundaries in music. Analyses and discussion were limited as much as possible to focus on the main purpose of this experiment which was the confirmation and extension of the previous work.

The main focus of the boundary analysis was the assessment of the placement of boundaries, and the relationship between empirically determined boundary locations and the model of Lerdahl and Jackendoff (1983). Replication of the previous work would imply that the conclusions of the previous experiment were not an artifact of task design. As will be detailed momentarily, the results did support the previous experiment.

As before, the main focus of the tonality analysis was the creation of groups of subjects homogeneous in tonality (as it is assessed by the probe-tone task). The results of this analysis were expected to, and did replicate the previous findings. In both this experiment, the tonality analysis found one large triadic group (2 were found in Experiment 1), one smaller proximity group (as in Experiment 1), and several smaller groups were found. After creating groups homogeneous for tonality, different patterns for boundary formation exploration were explored within each tonality group. Stated succinctly, the relationships found mimicked those of the previous experiment: That is, results indicated only minimal relationships between tonality group and boundary formation. As such, it was felt that reporting the details would only detract from the central focus so those results are not cited here.

General Considerations for Boundary Assessment

General Considerations for Boundary Location (Part 1 of each Stage)

In Part 1 of Stages 2, 4, 6 and 8, subjects heard a monophonic melody, repeated three times. Subjects indicated the location of boundaries within the melody, by the pressing of a key (usually the space bar). The key press was associated with the currently sounding note, creating a boundary profile for each melody (see Figure 5.6). Because subjects heard three repetitions of each melody within one stage, subjects actually produced three boundary profiles for each melody. Each boundary profile consisted of a note-by-note map of the location of those subjective boundaries, with boundaries being coded as a 1 and the lack of a boundary being coded as a 0. Because Stage 2 was practice, it was not analysed.

As before, the most important question concerns the locations of the boundaries, which becomes three subquestions: Where does each individual subject locate the boundaries?, are those locations stable within a single subject?, and are those locations stable across subjects? The first subquestion is answered by boundary profiles produced by each subject in response to each presentation of the melody. The answer to the second question is the same consistency analysis as in the previous experiment. The answer to the third subquestion is the same similarity analysis as in the previous experiment. However, in this experiment, only the phi similarity measure (the binary version of the Pearson correlation statistic) was used for the consistency and similarity analyses: Previously, it was shown that the five previous measures did not differ sufficiently to warrant retaining all (see Chapter 3 of this work; see Table 3.8).

The first analysis was the consistency analysis in which the correlations between the first, second and third repetitions were computed for each subject individually, for each melody. If subjects performed consistently, then there should be high similarities between the repetitions; in particular, because subjects were explicitly instructed to use the first and second repetitions to refine their answers, one would expect a highest similarity between the second and third repetitions. This was the general finding of Experiment 1.

The second analysis was the similarity analysis in which the correlations between all subjects on the third repetition were computed. Only the third repetition was used for this comparison in virtue of the aforementioned instructions. The examination of between-subject similarity was augmented by a cluster analysis on the similarity ratings: A cluster analysis indicated whether or not there are any homogeneous subgroups within the complete set of 62 subjects. As before (see Chapter 3 of this work), several variations of the basic cluster analysis were used (average between-groups linkage, average within-groups linkage, single linkage and complete linkage), but only the average between-groups linkage will be reported since no strong differences emerged. Before comparisons across subjects could be made (similarity analysis), the same pretest as before (see Chapter 3 of this work) was performed. It was possible that individual subjects scored boundaries differently: Some subjects could place boundaries at the beginning of units, some could place boundaries at the end of units (see Chapter 3 of this work, particularly, Figure 3.13) or some subjects could have had uniformly longer response times (for the actual key press, of the duration of note events). These would all result in relatively shifted boundaries although there would be no real difference in the unitization of the melody. As before, the mean profile was created based on all subjects and each individual subject was then correlated against this mean profile using relative shifts of -1, 0, and +1. The best shift for that individual was considered to be the value that maximized this correlation. As in Experiment 1, all subjects were maximized for a shift of 0, indicating a remarkable degree of intersubject consistency, so only those results are reported.

The third analysis examined the predictability of boundaries based on the Group Preference Rules (GPR) of Lerdahl and Jackendoff (1983). This analysis was essentially the same as that of the previous analyses (particularly, Chapter 4 of this work).

Essentially, the theoretical rules were fitted to the empirical data using non-linear regression and asymmetrical error terms (weighing false alarms much less than misses). The rules of Lerdahl and Jackendoff rules were shifted to align with the empirical data using the same best fit criteria. Only the best match is presented.

General Considerations for Boundary Efficacy: (Part 2 of each Stage)

The fourth analysis examined the utility of the boundary locations. If boundary construction is more than an artifact of the experiment (regardless of any intersubject consistency), then boundary formation should have some impact on a secondary click-detection task. Stated succinctly, click detection should be delayed for clicks preceding boundaries and click detection should be advanced for clicks following boundaries.

To assess this, the response times for each subjects were normalized to a mean of 0 and a variance of 1.0 (i.e., scores for each subject were converted to z-scores to remove individual differences in overall reaction times). These z-scores were then analysed using a hierarchical regression as a function of distance from the click to the preceding boundary, the distance from the click to the subsequence boundary and the distance of the click into the melody (i.e., melody position). The use of z-scores puts all subjects on an equal footing with respect to their mean reaction times and their range of scores (this type of analysis is, in fact, analogous within-subjects analysis[52]): The goal of the analysis is the determination of which positions, if any, result in relatively longer or short reaction times consistently for all subjects.

<u>Boundary Location: Stage 4, Part 1</u>

In Part 1 of Stage 4, (see Table 5.6), the mean consistency values for subject-by-subject repetitions of the melody were 0.74±0.45 (phi measure only). All values indicate a high degree of similarity within individual subjects. The modal value for adjacent repetitions was 1.0. Furthermore, by both the standard deviations and the minima, only a few subjects fell to the level of chance indicated in Table 3.8 and even

---

[52] The within-subjects analysis actually equates subjects for their mean scores. The variance within a subject becomes a part of the within-cell error term.

then, it was only on the first to second repetition comparison[53]. However, recall that these chance levels assumed equal numbers of boundaries per boundary profile and a particular probability for boundaries within a boundary profile.

Table 5.6: The average consistency of subjects over different repetitions of the same melody within Stage 4, Part 1, given the phi similarity measure of Table 3.8.

	Consistency	Mean & SD	Median	Mode	Minimum	Maximum
Phi Similarity	1 to 2	0.739±0.370	0.742	1.000	0.216	1.000
	1 to 3	0.717±0.176	0.726	0.863	0.281	1.000
	2 to 3	0.773±0.174	0.756	1.000	0.390	1.000

To further examine the consistency ratings, the correlation between consistencies were computed (Table 5.7). That is, each subject produced three consistency values (phi measures) and the correlations between these consistency values was assessed. The consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 1 and 3, the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3 and the consistency between Repetitions 1 and 2 was correlated against the consistency between Repetitions 2 and 3. The resulting correlations were all high (and significant at a type 1 error rate of $\alpha=0.05$), indicating that the subjects who were the most consistent in the first two repetitions remained the most consistent in the third repetition. Note however, that the lowest rating occurs when the first to second repetition consistency is correlated against the second to third repetition consistency. This implies that there is some drift in the responses of subjects.

---

[53] Throughout this work, original untransformed scores are presented where possible. In a number of analyses, log-transformed and/or Fisher-transformed scores were examined: Those results are only reported when they are different from the untransformed. Mention is made of both in the first few examples as a reminder.

Table 5.7: The correlations between consistency ratings over the different repetitions of the same melody within Stage 4, Part 1, given the phi similarity measure of Table 3.8.

Correlations	Consistency Between Repetitions	Consistency Between Repetitions		
		1 to 2	1 to 3	2 to 3
Phi	1 to 3	1.000	0.627**	0.313**
	2 to 3			0.571**

Notes: * $p<0.05$
** $p<0.01$
Significance levels did not change when log-transformed or Fisher-transformed scores were used.

As shown in the within-subjects ANOVA[54] of Table 5.8, consistency ratings did change slightly as a function of the repetitions being compared but the effect was not particularly strong (note the η^2 and ω^2 values: since this is a within-subjects design, one would expect these to be much higher; see comments in the "Notes on the Communication and Interpretation of Results", Chapter 3 of this work).

---

[54] For most analyses it is the contrasts (single df) analyses that are of interest. In many cases, the omnibus within-subjects ANOVA is not the most appropriate analysis (i.e., assumptions were violated), but for consistency, the within-subjects ANOVA is always reported. Multivariate results are included if needed. Note that contrast analysis using specific error terms (as done herein) are not affected by the appropriateness of the overall ANOVA.

Table 5.8: The ws-analysis of the change in consistency ratings for the different repetitions of the same melody within Stage 4, Part 1, given the phi similarity measure of Table 3.8.

WS-ANOVA		SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Phi	Overall	0.097	2	1.820	122	3.251	0.042	0.051	0.035
	Ψ_1	0.036	1	1.243	61	1.767	0.189	0.028	
	Ψ_2	0.061	1	0.577	61	6.457	0.014	0.095	

Notes: Ψ_1 compares Consistency from 1 to 2 with Consistency from 2 to 3 (i.e., nearest neighbours in time)
Ψ_2 compares the average of Consistency from 1 to 2 and Consistency from 2 to 3 with the Consistency from 1 to 3 (close in time to far in time)
Although the Mauchy sphericity test indicated a violation of sphericity, the multivariate results were the same.
Results (including multivariate results) were the same using transformed scores.

When comparing across subjects (Table 5.9), average similarity ratings were lower, as before (see Chapter 3 of this work) with a mean of 0.65±0.18. This value is still above the critical and chance levels specified in Table 3.8, and in fact, only the minimum dips below the critical values cited. As can be seen in the subsequent cluster analysis, this only occurred for a few subjects (3 or 4 depending upon what one wishes to call chance).

Table 5.9: The average similarity ratings between different subjects for the same melody, within Stage 4, Part 1, Repetition 3 only, given the phi similarity measure of Table 3.8.

Similarity Measure	Mean & SD	Median	Mode	Minimum	Maximum
Phi	0.645±0.180	0.655	0.787	0.005	1.000

Overall, the central tendencies in the similarity analyses (mean, mode, median) indicate that subjects perform the same. However, the minimums in both the consistency and the similarity analyses imply that a few subjects may be somewhat unique.

To further explore the relationship between different subjects, a cluster analysis was performed on the bases of the phi similarity measure determined from the third repetition of the melody. As can be seen in Figure 5.6, the analysis is indicative of one large group to which individuals gradually accrue. Using the previous critical or chance values one can see that the clustering would consist of one large group with three or four stragglers. As such, the cluster profiles indicate that all subjects should be treated as one homogeneous group (one could argue that 3 or 4 subjects be dropped, but relative to 62, this in not an important factor).

Given that all subjects can be treated as one group (as in the previous experiment), it is possible to create the average boundary profile. The average boundary profile for each repetition is shown in Figures 5.7 (as before, the melody is split for ease of presentation). Subjects placed strong boundaries (i.e., many subjects placed boundaries at these points) at note events 4, 7, 11, and 16, with weaker boundaries at note events 25 and 36 (tied notes represent one event because they represent a single sound). As before, a boundary was placed at the end of the melody for all subjects. Generally the agreement between subjects is high at the beginning but lower near the end. Figure 5.7 also provides the boundary profile that represents the average of Repetitions 2 and 3; that is, the average profile for each subject (called the individual profile as in Chapter 3 of this work) was created and then these average profiles were averaged.

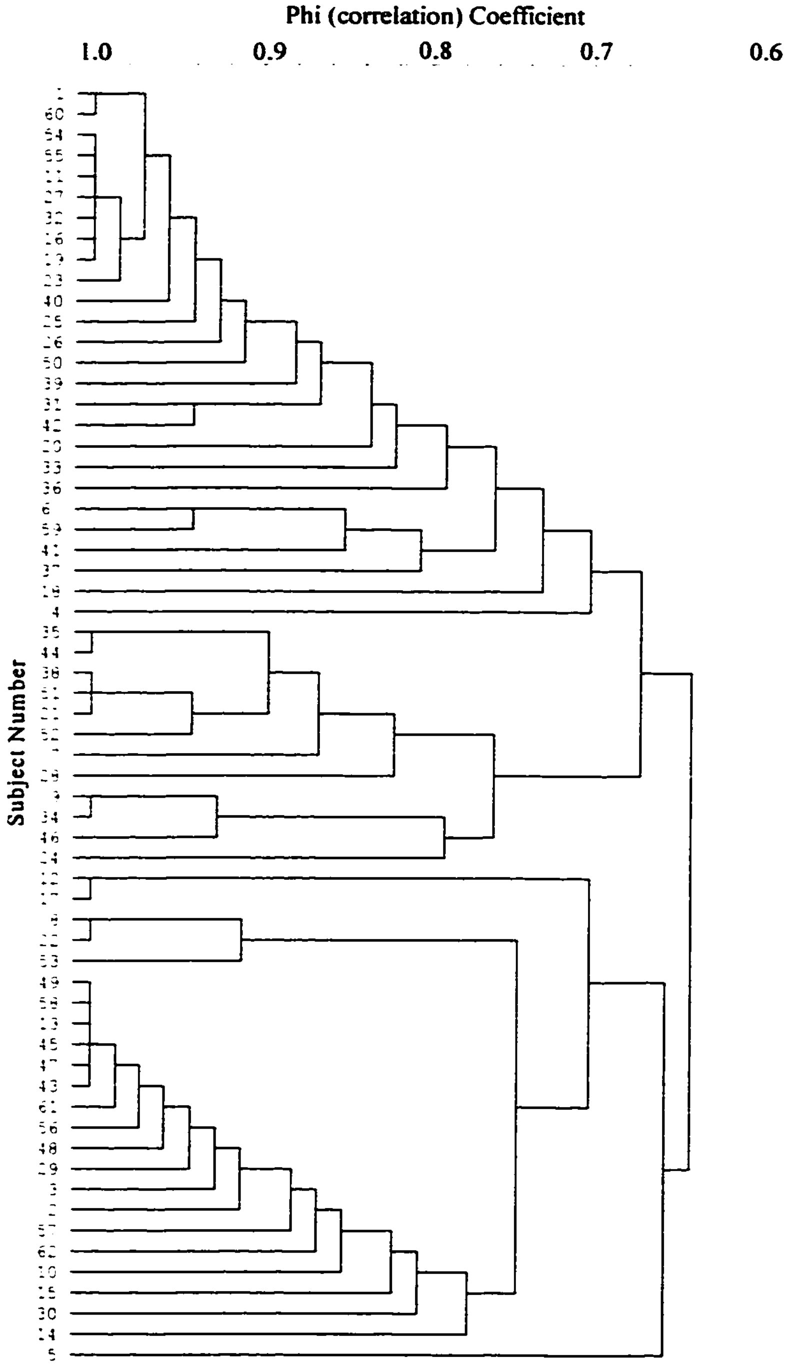


Figure 5.6: Cluster analysis of boundary profiles of individual subjects using the Phi measure (the binary form of the correlation coefficient) for the third repetition of the melody used in Stage 4 ("Three Blind Mice"). Note that the maximum correlation was 1.0 and the minimum value was 0.64 indicating that all subjects were fairly similar. Subject numbers are arbitrary.

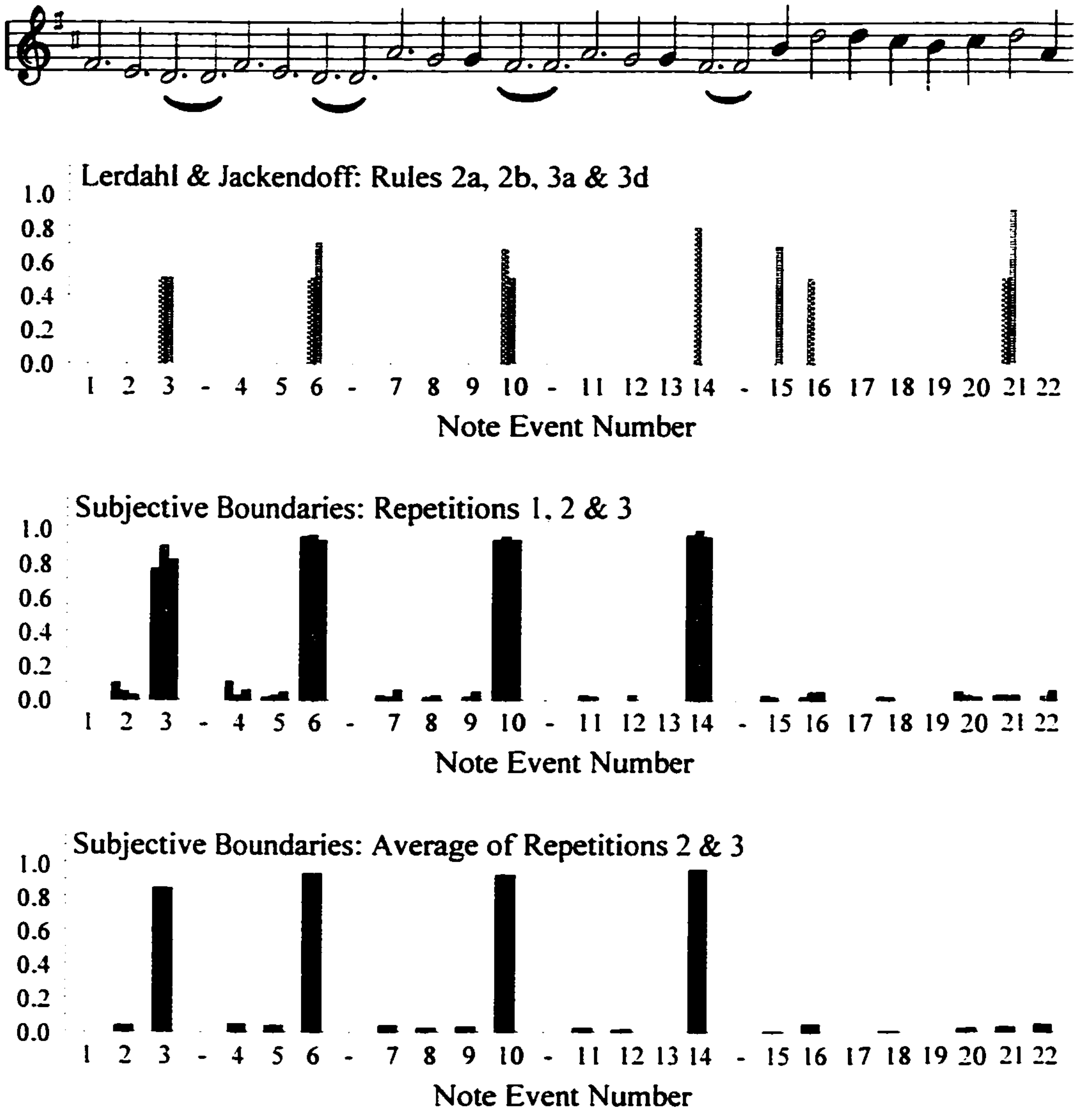


Figure 5.7a: The first part of the song used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries for each repetition of the song, and the average of the second and third repetitions (the average is used in subsequent analyses). These profiles are based on all 62 subjects.

▨ Rule 2a	■ First Presentation
▩ Rule 2b	■ Second Presentation
▦ Rule 3a	■ Third Presentation
▧ Rule 3d	

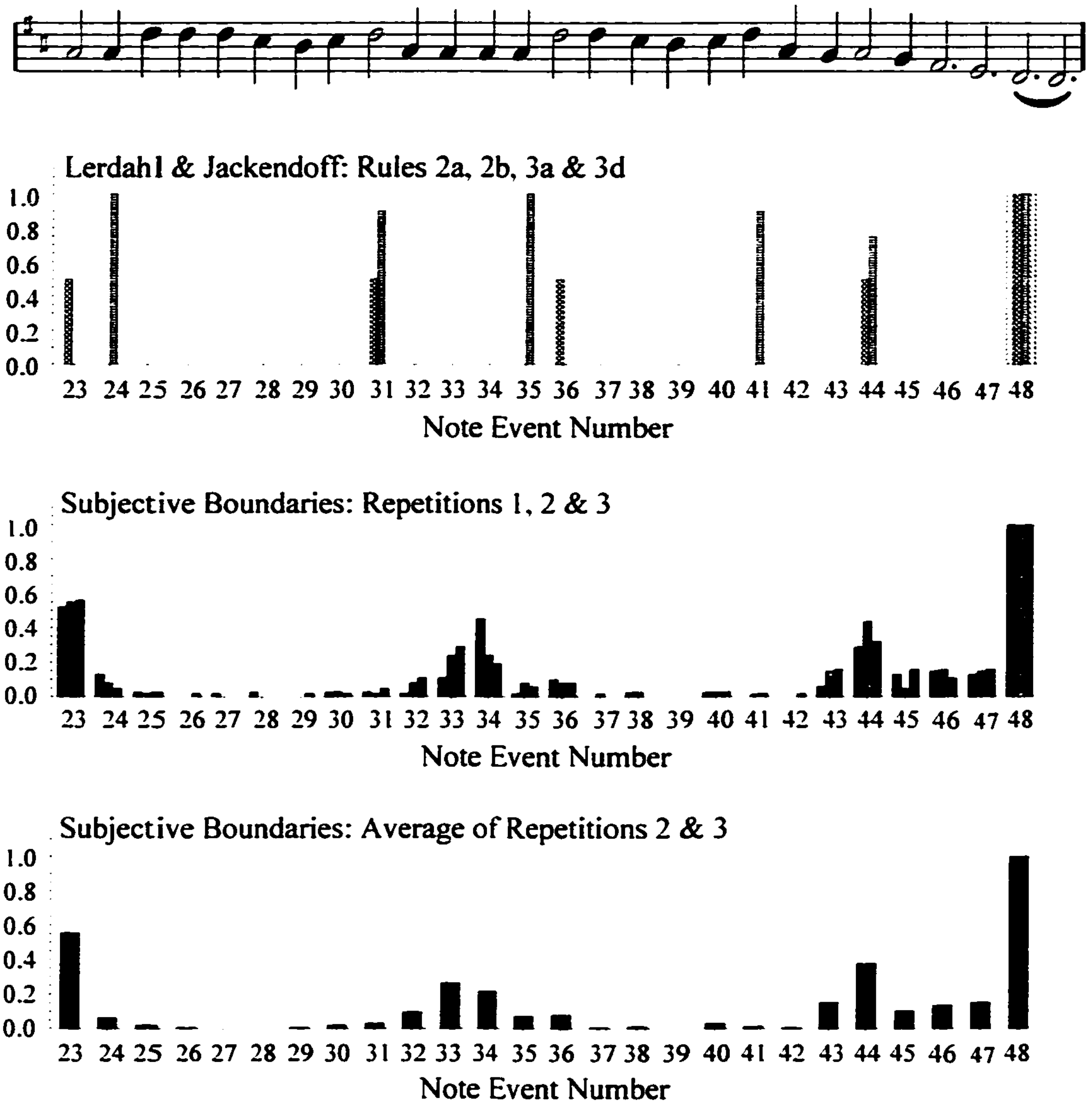


Figure 5.7b: The second part of the song used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries based on the average of all subjects for each repetition of the song and the average of the second and third repetition for all subjects (the average is used in subsequent analyses).

▨	Rule 2a	■	First Presentation
▩	Rule 2b	■	Second Presentation
▦	Rule 3a	■	Third Presentation
▧	Rule 3d		

The detailed descriptive statistics for this profile are provided in Tables 5.10 through 5.12 (various measures pertaining to the locations of boundaries) and Table 5.13 through 5.15 (correlations between those measures of boundary location) for each repetition of the melody.

Table 5.10: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 1 (the first presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	1.532±3.007	1.000	1.000	0.000	21.000
Maximum	16.790±9.180	12.000	10.000	4.000	33.000
Mean	6.250±3.127	5.714	4.875	1.611	22.500
Number with 0	0.500±1.225	0.000	0.000	0.000	7.000
Number with 1	0.935±0.787	1.000	1.000	0.000	5.000
Number with 2	1.048±0.612	1.000	1.000	0.000	3.000
Number with 3	2.097±0.762	2.000	2.000	0.000	4.000
Number with 4	0.258±0.745	0.000	0.000	0.000	5.000
Number with 5	0.097±0.433	0.000	0.000	0.000	3.000
Number with 6	0.032±0.178	0.000	0.000	0.000	1.000
Number with 7	0.065±0.248	0.000	0.000	0.000	1.000
Number with 8	0.548±0.592	0.500	0.000	0.000	2.000
Number with 9	0.500±0.763	0.000	0.000	0.000	3.000
Number with ≥10	1.194±0.623	1.000	1.000	0.000	3.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).
Number with "xx" is the mean number of times subjects created units containing xx note events (exclusive of the actual boundaries).

Table 5.11: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 2 (the second presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	1.065±1.158	1.000	1.000	0.000	8.000
Maximum	16.629±8.906	11.000	10.000	8.000	33.000
Mean	5.791±2.176	4.875	4.875	1.611	14.667
Number with 0	0.613±1.692	0.000	0.000	0.000	9.000
Number with 1	1.081±0.581	1.000	1.000	0.000	3.000
Number with 2	1.000±0.573	1.000	1.000	0.000	3.000
Number with 3	2.161±0.706	2.000	2.000	0.000	3.000
Number with 4	0.210±0.410	0.000	0.000	0.000	1.000
Number with 5	0.081±0.329	0.000	0.000	0.000	2.000
Number with 6	0.097±0.349	0.000	0.000	0.000	2.000
Number with 7	0.129±0.338	0.000	0.000	0.000	1.000
Number with 8	0.645±0.603	1.000	1.000	0.000	2.000
Number with 9	0.516±0.593	0.000	0.000	0.000	2.000
Number with ≥10	1.048±0.493	1.000	1.000	0.000	3.000

Notes: see Table 5.10

Table 5.12: The important statistics for the location of boundaries within Stage 4, Part 1, Repetition 3 (the third presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.500±8.051	1.000	1.000	0.000	46.000
Maximum	16.177±10.269	10.000	10.000	5.000	46.000
Mean	6.968±7.510	4.875	4.875	1.611	46.000
Number with 0	0.710±1.928	0.000	0.000	0.000	11.000
Number with 1	0.952±0.734	1.000	1.000	0.000	5.000
Number with 2	0.919±0.609	1.000	1.000	0.000	2.000
Number with 3	2.048±0.838	2.000	2.000	0.000	5.000
Number with 4	0.339±0.651	0.000	0.000	0.000	3.000
Number with 5	0.097±0.349	0.000	0.000	0.000	2.000
Number with 6	0.097±0.349	0.000	0.000	0.000	2.000
Number with 7	0.129±0.338	0.000	0.000	0.000	1.000
Number with 8	0.726±0.605	1.000	1.000	0.000	2.000
Number with 9	0.516±0.593	0.000	0.000	0.000	2.000
Number with ≥10	1.016±0.528	1.000	1.000	0.000	2.000

Notes: see Table 5.10

Over all repetitions, the mean number of boundaries[55] was fairly constant at 8.47±0.17 (in this case, the standard error is cited). The modal number of boundaries was either 8.00 or 9.00. This number of boundaries represents a mean number of note events *between* boundaries equal to 6.34±0.59; the mean number of note events within a unit should therefore be 7.34 because each boundary should be associated with either the beginning or the end of a unit (i.e., boundaries are located on particular note events). The size of individual units ranged from 0.00 note events to 46.00 (effectively the whole melody) note events. However, on a subject by subject basis, the most common size was 3.00 note events (implying 4.00 note events when the boundary is included), and the least common

---

[55] Analyses of number of boundaries are not included. Number of boundaries is somewhat redundant with mean distance between boundaries, but mean distance between boundaries is more useful for comparing different melodies.

sizes were 5.00, 6.00, 7.00 and to a lesser extent 4.00 (implying 6.00, 7.00, 8.00, and 4.00 when boundaries are included).

From the correlation matrices, one can see that the minimum number of notes between boundaries (the smallest grouping size for each subject) was correlated with the maximum number of notes between boundaries (the largest grouping size for each subject) only in the third repetition. This positive correlation implies those subjects who used the largest maxima, also used the largest minima; that is, some subject tended to use larger units overall. Note that the minima and maxima were also correlated with the means. In addition, larger minima were associated with fewer small units and larger maxima were associated with fewer large units (that may be a reasonable restriction: one cannot have many large units if one has one very large unit).

Table 5.13: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 1 (the first presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.124	1.000	
Mean	.837**	.590**	1.000
Number with 0	-.212	-.078	-.304*
Number with 1	-.311*	-.193	-.428**
Number with 2	-.433**	-.278*	-.525**
Number with 3	-.524**	-.424**	-.667**
Number with 4	-.004	-.222	-.122
Number with 5	-.002	-.164	-.085
Number with 6	-.002	-.146	-.076
Number with 7	.019	.114	.119*
Number with 8	-.176	-.446**	-.328**
Number with 9	-.082	-.425**	-.252*
Number with ≥10	.320*	-.001	.274*

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 5.14: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 2 (the second presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.163	1.000	
Mean	.675**	.735**	1.000
Number with 0	-.339**	-.194	-.470**
Number with 1	-.520**	-.121	-.538**
Number with 2	-.495**	-.161	-.485**
Number with 3	-.434**	-.293*	-.522**
Number with 4	.213	-.086	.084
Number with 5	-.057	-.158	-.106
Number with 6	-.138	-.236	-.269*
Number with 7	.146	-.000	.142
Number with 8	-.084	-.672**	-.414**
Number with 9	-.073	-.491**	-.327**
Number with ≥10	.396**	.071	.304*

Notes: * Significant (one tailed) at p<0.05
** Significant (one tailed) at p<0.01

Table 5.15: The correlations between the various statistics for the locations of boundaries within Stage 4, Part 1, Repetition 3 (the third presentation of the melody).

Correlations	Minimum	Maximum	Mean
Minimum	1.000		
Maximum	.557**	1.000	
Mean	.972**	.716**	1.000
Number with 0	-.116	-.194	-.201
Number with 1	-.296*	-.308*	-.373**
Number with 2	-.310*	-.294*	-.325*
Number with 3	-.465**	-.506**	-.527**
Number with 4	-.061	-.249	-.113
Number with 5	-.029	-.129	-.074
Number with 6	-.029	-.202	-.104
Number with 7	-.036	-.073	-.031
Number with 8	-.244	-.615**	-.352**
Number with 9	-.192	-.465**	-.271*
Number with ≥10	.006	.151	.077

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

A within-subjects ANOVA was performed on the minimum, maximum and mean distances between boundaries to see if there were any changes as a function of repetition (see Table 5.16). There were none. The associated correlation matrix (Table 5.17) implied that individuals were more consistent over the first two repetitions (as can be seen more directly in the previous tables).

Table 5.16: The analysis of the changes in the boundary statistics across repetitions of the same melody, for Stage 4, Part 1.

WS-ANALYSIS		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Minimum Distance Between Boundaries	Overall	66.462	2	2949.538	122	1.375	0.257	0.022	0.006
	Ψ_1	29.032	1	2150.968	61	0.823	0.368	0.013	
	Ψ_2	37.430	1	798.570	61	2.859	0.096	0.045	
Maximum Distance Between Boundaries	Overall	12.516	2	4478.817	122	0.170	0.843	0.003	0.000
	Ψ_1	11.645	1	3027.355	61	0.235	0.630	0.004	
	Ψ_2	0.871	1	1451.462	61	0.036	0.849	0.001	
Mean Distance Between Boundaries	Overall	43.683	2	2337.089	122	1.140	0.323	0.018	0.002
	Ψ_1	15.993	1	1782.913	61	0.547	0.462	0.009	
	Ψ_2	27.690	1	554.176	61	3.048	0.086	0.048	

Notes: All tests violated the Mauchly sphericity test, but in no case did any corresponding multivariate test, Pillais, Hotellings Wilks or Roys, indicate a different conclusion.

Table 5.17: The correlations for the boundary statistics across repetitions of the same melody, for Stage 4, Part 1.

Correlation Within Measures			Repetition Number		
			1	2	3
Minimum Distance Between Boundaries	Repetition Number	1	1.000	0.216	0.069
		2		1.000	0.026
Maximum Distance Between Boundaries	Repetition Number	1	1.000	0.619**	0.479**
		2		1.000	0.690**
Mean Distance Between Boundaries	Repetition Number	1	1.000	0.730**	0.164
		2		1.000	0.283*

Notes * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Given that there exists an average boundary profile that can be said to represent the parsing by individual subjects, the question shifts to one of the prediction of those boundaries based upon the structure of music. In this analysis, the method of non-linear regression with asymmetric error terms (see Chapter 4 of this work) was applied to the average profile of Repetitions 2 and 3. That is, for each subject, the average profile was created for Repetitions 2 and 3 and then these average profiles were averaged (the order actually does not matter). This averaging seems valid given the high within-subject consistencies and the high between-subjects similarity (compare the lower panels of Figure 5.7).

As before (see Chapter 3 of this work), the analysis proceeded using the best alignment between the empirical data and the theoretical predictions. As before, the best alignment implied that subjects were placing their boundaries at the ends of units.

Given the asymmetric error function (Chapter 4 of this work), the results for "Three Blind Mice" (Stage 4, Part 1), for each rule are presented in Table 5.18 and Figure 5.8 (Rule 4 is included with this figure: see later analysis). Figure 5.8 presents the original scaling of Rules 2b and 3a in the first panel and the best individual fits for Rules 2b and 3a in the second panel. In this analysis, for each rule, an overall error for the entire melody was computed. This error value represents the ability of each rule, in isolation, to account for empirical boundaries in the general sense. Rule 2b (Attack-Point Change) was much more effective than Rule 3a (Register Change): The overall error for Rule 2b was much less than the overall error for Rule 3a. This difference exists even though both rules were applicable at the same total number of positions (11 each). Rule 2b (Attack-point) was generally successful at predicting boundaries while Rule 3a (Register change) was much less so.

Table 5.18: Rules 2 and 3 for "Three Blind Mice".

Overall Results				
Rule	n	*r*	Equation	SS_{error}
2b	11	0.704**	b = 1.168*2b	8.910
3a	11	0.286**	b = 0.892*3a	18.668

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b	4	0.687	0.788	7	1.651	2.543
3a	4	1.778	1.290	7	3.793	4.878

Notes: *r* = simple correlation

n = number of point that the rule applied (either as a miss or false alarm)

$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$

$|\text{Deviation}| = \Sigma(X_i - X'_i)$

In this analysis of specific errors for misses and false alarms, the only points of the melody to be used were those for which the rule had a non-zero value (i.e., there was some non-zero prediction at that point; see Chapter 4 of this work). This allowed for a direct comparison of the different rules at those points where they were expected to work. It is important to remember that the point was to minimize misses and let false alarms rise to whatever value was necessary. In this analysis, Rule 2b was still much superior than Rule 3a. The error for misses of Rule 2b was half the value of the error for misses of Rule 3a, even though both applied to an equal number of locations (misses are note events where the best-fit theoretical rule was greater than the empirical data). The same was true of errors for false alarms. Rule 2b was much better than Rule 3a, producing about one-half the error value on the same number of occurrences (false alarms are note events where the best-fit theoretical rule was less than the empirical data).

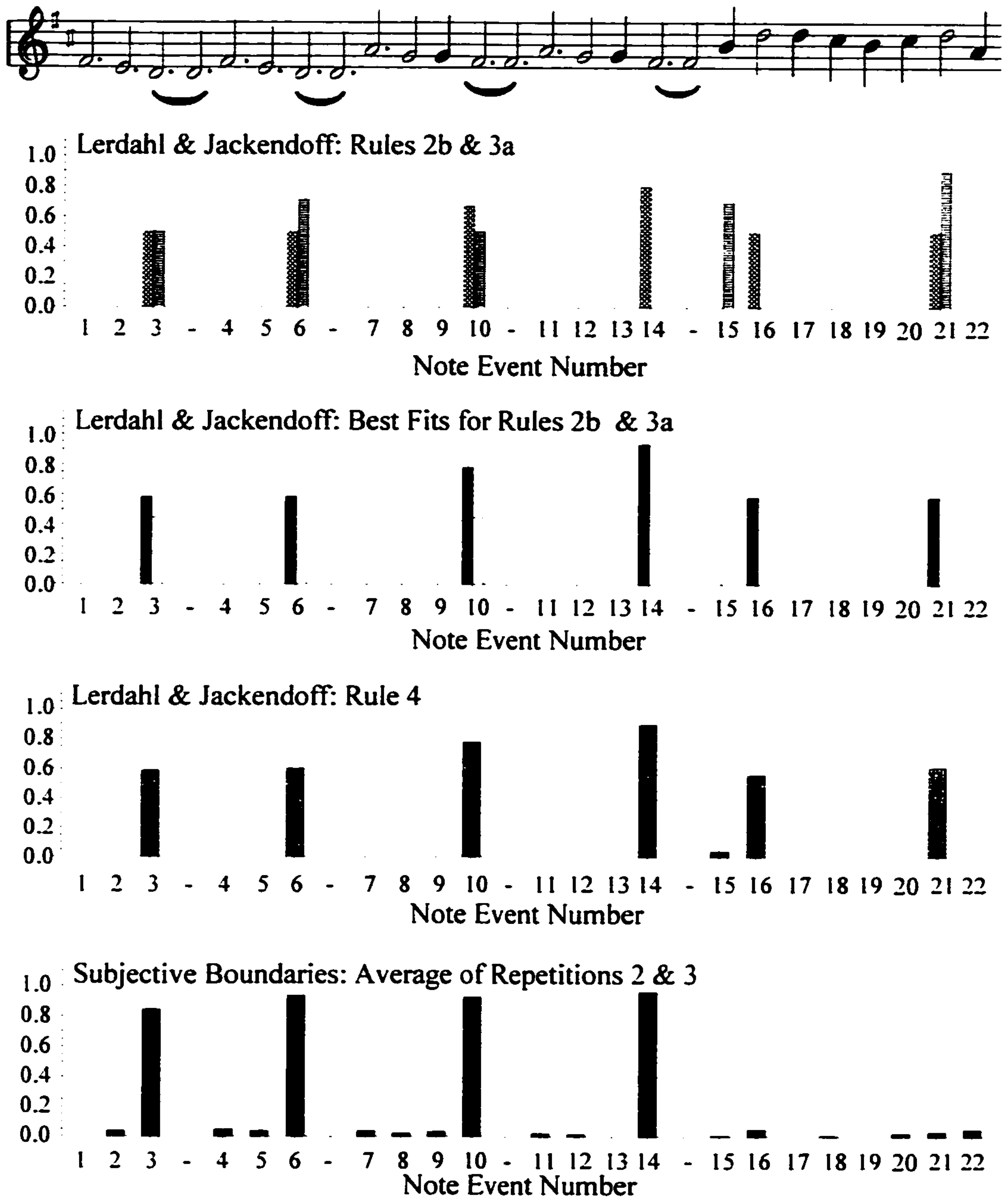


Figure 5.8a: The first part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a alone and in combination as Rule 4). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

▩ **Rule 2b** ■ **Fitted 2b** ■ **Rule 4 (Best combination of Rules 2b and 3a)**
▨ **Rule 3a** ☐ **Fitted 3a** ■ **Average Profile (Repetitions 2 & 3)**

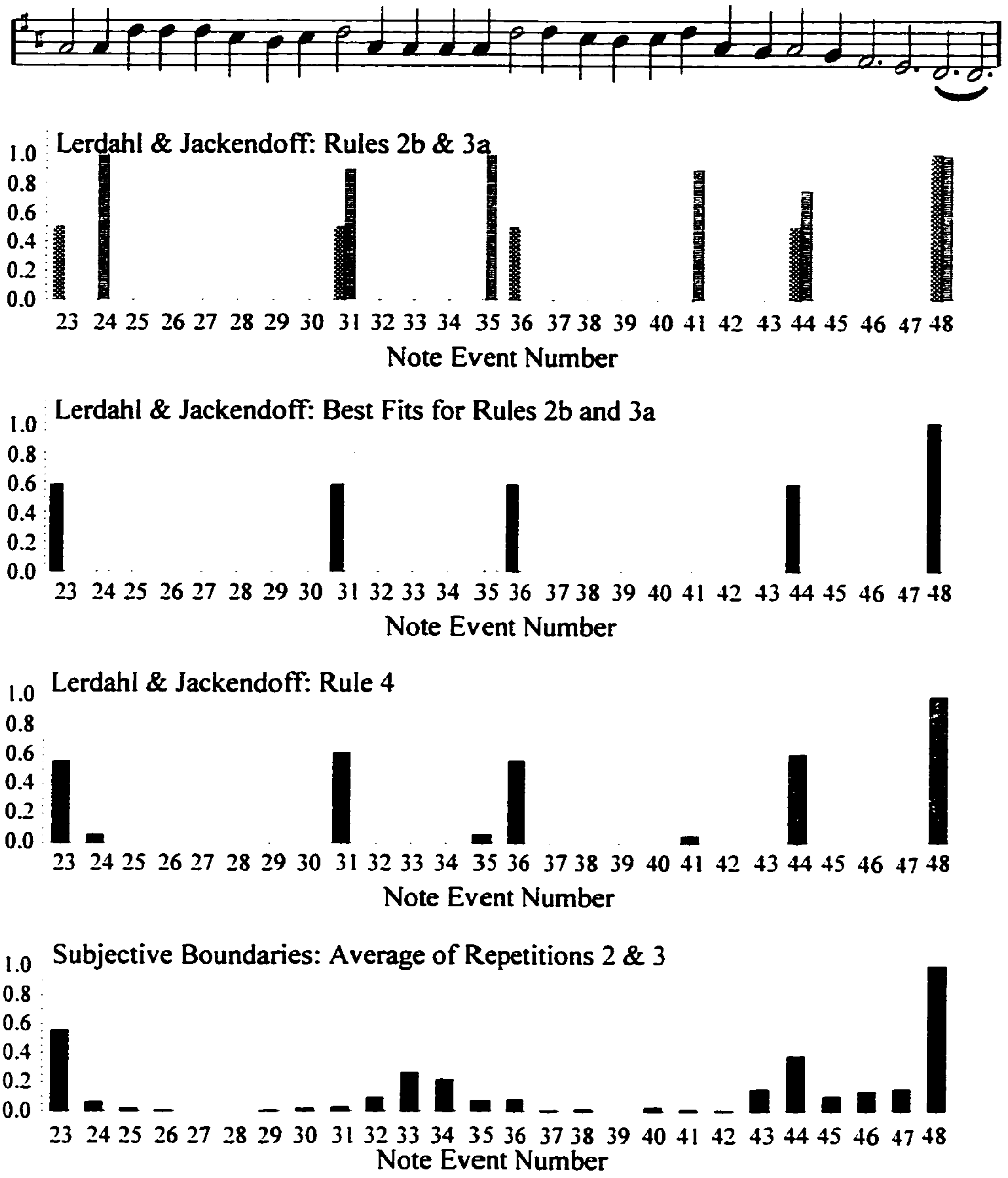


Figure 5.8b: The second part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a alone and in combination as Rule 4). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

▨ Rule 2b	■ Fitted 2b	■ Rule 4 (Best combination of Rules 2b and 3a)
▩ Rule 3a	☐ Fitted 3a	■ Average Profile (Repetitions 2 & 3)

Consistently, Rule 2b is a better predictor of the data overall, than Rule 3a. Caution is urged when extrapolating to melody not tested.

Note that the overall error need not be the sum of the error for misses and false alarms. Misses and false alarms were calculated only at those points where the prediction had some non-zero value. The overall was calculated at all points in the melody. Hence, the overall error for each rule includes all the points where that rule was never intended to be applied and as such the overall error includes all empirical boundaries that could be (would be) predicted by other rules: These points would contribute large values (as misses) to the overall error term.

The next analysis examined the ability of the combination of Rules 2b and 3a, including their interaction, to predict empirical boundaries. This has been labelled as Intensification (Rule 4) though it may not correspond perfectly with the notion of intensification at implied by Lerdahl and Jackendoff (1983; cf., Chapter 4 of this work). The basic summation of the two rules (see Table 5.19 & Figure 5.8) increased the total number of points predicted to 15 (from 11 for Rule 2b alone, or from 11 for Rule 3a alone). The use of both rules simultaneously improved the overall fit slightly from its best value for a single rule (8.910 for Rule 2b down to 8.874) as well as the misses (0.687 for Rule 2b down to 0.639) while false alarms rose slightly (1.651 for Rule 2b up to 1.698). From the equation, one can see that the best fit of the combination emphasized Rule 2b and restricts Rule 3a to a very subsidiary rule. The slope for Rule 2b was almost unchanged from its value when used as the only predictor (previously 1.168), while the slope for Rule 3a was greatly reduced (previously: 0.892). By comparison of the second panel of Figure 5.8 (Rules 2b & 3a in isolation) with the third panel of 5.8 (Rule 4), one can seen the Rule 4 closely resembled Rule 2b and not Rule 3a.

It important to realize that the improvement in fit is the result of two separate factors. Firstly, by use of the combination, more empirical boundaries were accounted for than by any one rule alone. Secondly, the use of the addition of two rules allowed the slope of each rule to be reduced. This had a major effect on false alarms (it reduces the overprediction when there is no empirical boundary). The analysis clearly indicated that most of the effects that were previously granted to Rule 3a were better explained by Rule 2b. That is, Rule 3a and 2b overlapped on seven points, and most of that overlap was best explained by Rule 2b alone. In the area where Rule 3a acted alone (i.e., at points where Rule 2b did not apply; see Note Events 15, 24, 35, a& 41), the algorithm did not find any evidence for predictive efficacy. The evidence for this is that the slope for Rule 3a declined dramatically when the rules were used in conjunction. Figure 5.8 also allows one to compare this combined effect with the individual effects and the empirical data. Note visually, that the combined data is a much better match to the empirical data, particularly in regard to false alarms. Finally, as can be seen in Table 5.19, the interaction of Rules 2b and 3a did not contribute to the prediction (as in the previous experiment).

Table 5.19: The Combination of Rules 2 and 3 (Rule 4) for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.124*2b + 0.059*3a	8.874
2b+3a+2b*3a	15	b = 1.124*2b + 0.059*3a + 0.000*2b*3a	8.874

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	6	0.639	0.817	9	1.698	2.645
2b+3a+2b*3a	6	0.639	0.817	9	1.698	2.645

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

The "final" analysis examined parallelism (symmetry was not analysed for reasons cited in Chapter 4 of this work). The quotes are used around the term final because the analysis of parallelism actually involved four separate aspects of parallelism. The first was parallelism due to pitch contour (Rule 6a, Pitch Pattern Parallelism A), analysed in a manner that included note durations. The second was parallelism due to pitch contour (Rule 6b, Pitch Pattern Parallelism B), analysed in a manner that ignored note durations (i.e., note onset to note onset). The third was parallelism due to time pattern (Rule 6c, Time Pattern Parallelism A), analysed in a manner that notated the timings of breaks between note events. The fourth was parallelism due to time pattern (Rule 6d, Time Pattern Parallelism B), analysed in a manner that compared the sequential note event durations. Details of the analysis have been provided elsewhere (see Chapter 4 of this work).

In all analysis, parallelism was treated as an addition to the lower level rules (Rules 2b and 3a, in particular). Hence, in all reports, the low level rules are reproduced first and then the parallelism is presented as an additional layer. This is consistent with the theory of Lerdahl and Jackendoff (1983), and allows one to easily see any additional benefits of parallelism. Finally, for each type of parallelism, the interaction of parallelism with each low level rule was computed and entered, followed by all rules and their interactions as the final step. The interaction of Rules 2b and 3a was not included. All the results reported are for the best case beat- or note-span analysis (see Chapter 4 of this work).

In the first of the final analyses, Pitch Pattern Parallelism A (Rule 6a) added predictability to the model (Table 5.20; Figure 5.9). That is, the different terms, Rule 6a or the interactions of Rule 6a with Rules 2b and 3a, reduced the error below the value of 8.874 that was achieved by Rule 4 (Intensification; the optimal combination of Rules 2a and 3d, Rule 4, is repeated in Table 5.20 for comparison) to the best combination of 8.692 for a change in error ΔSS_{error}=0.28 or 3.1%. Parallelism helped by allowing the constants preceding Rules 2a and 3a to be reduced in magnitude, which helps to decrease false alarms.

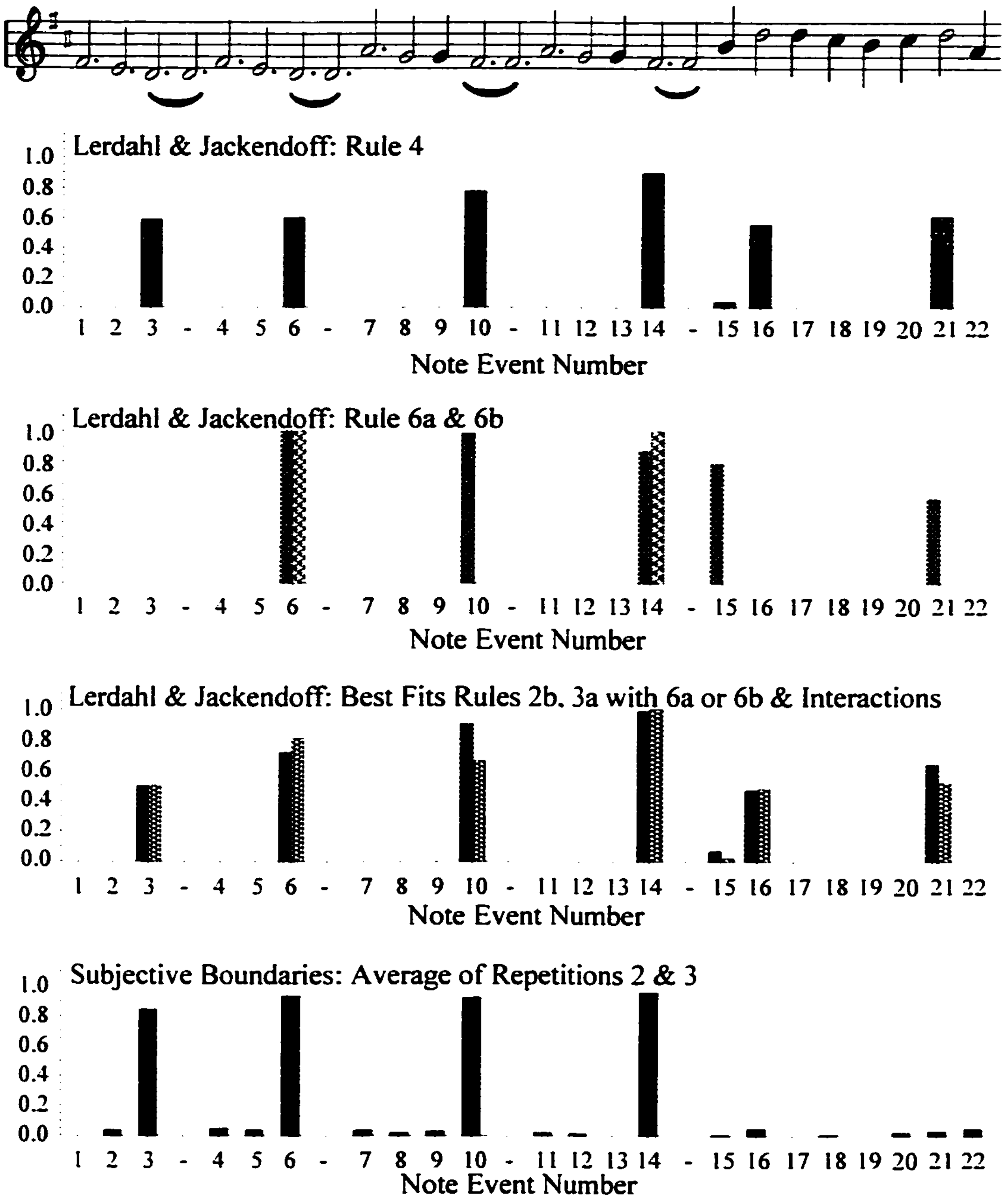


Figure 5.9a: The first part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a alone and in combination with 6a or 6b). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

■	**Rule 4**	■	**Best Fit for Rules 2b, 3a, 6a & their interactions**
▣	**Rules 2b, 3a & 6a**	▦	**Best Fit for Rules 2b, 3a, 6b & their interactions**
▨	**Rules 2b, 3a & 6b**	■	**Average Profile (Repetitions 2 & 3)**

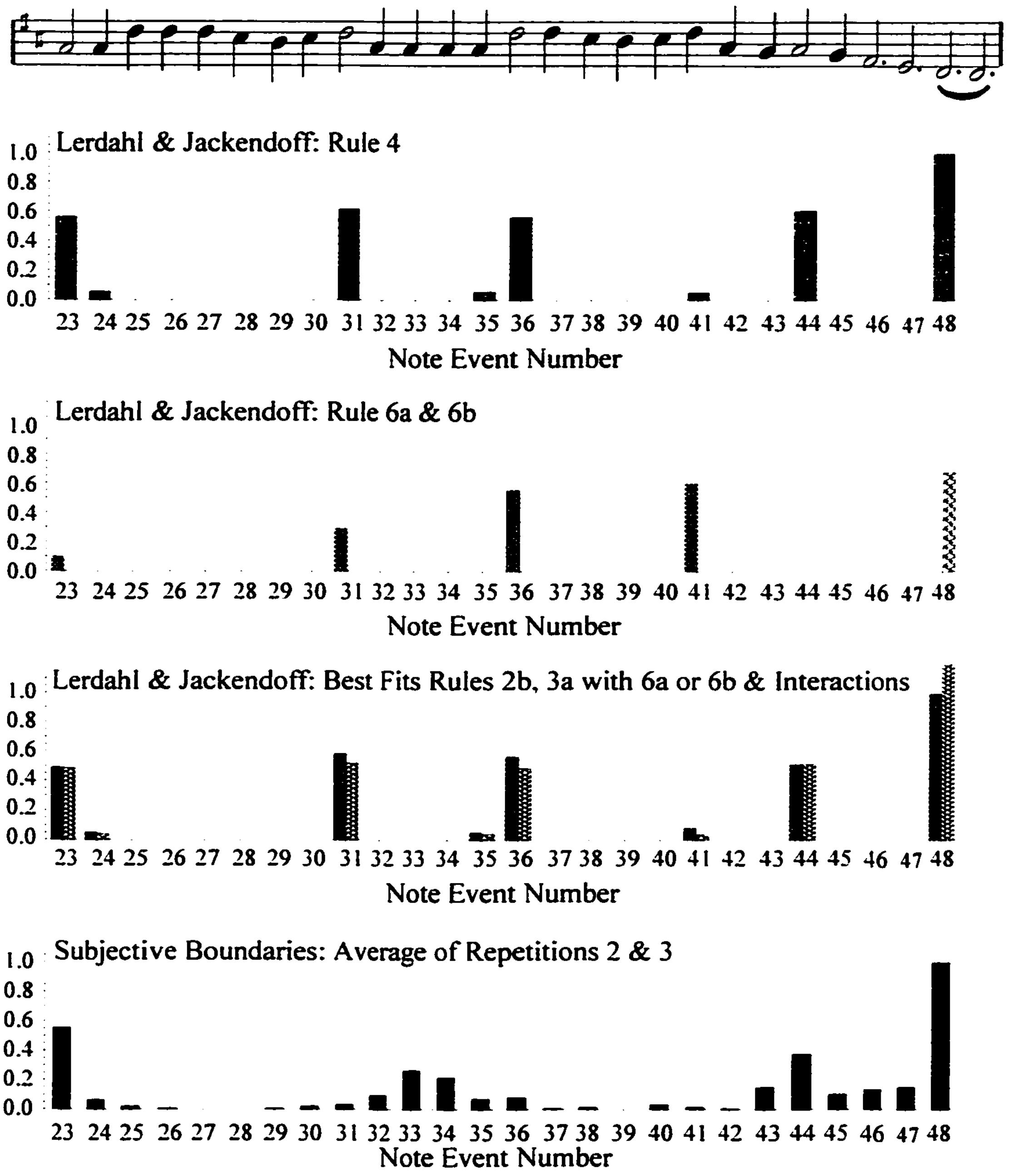


Figure 5.9b: The second part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a alone and in combination with 6a or 6b). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

■	**Rule 4**	■	**Best Fit for Rules 2b, 3a, 6a & their interactions**
▤	**Rules 2b, 3a & 6a**	▦	**Best Fit for Rules 2b, 3a, 6b & their interactions**
▩	**Rules 2b, 3a & 6b**	■	**Average Profile (Repetitions 2 & 3)**

Table 5.20: The Combination of Rules 2, 3 (Rule 4) and 6a for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.124*2b+0.059*3a	8.874
2b+3a+6a	15	b = 1.016*2b+0.022*3a+0.169*6a	8.697
2b+3a+6a+2b*6a	15	b = 0.927*2b+0.070*3a+0.000*6a+0.327*2b*6a	8.604
2b+3a+6a+3a*6a	15	b = 1.038*2b+0.007*3a+0.112*6a+1.229*3a*6a	8.692
all	15	b = 0.946*2b+0.051*3a+0.000*6a + 0.283*2b*6a+0.776*3a*6a	8.598

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	6	0.639	0.816	9	1.698	2.645
2b+3a+6a	7	0.564	0.770	8	1.595	2.541
2b+3a+6a+2b*6a	6	0.581	0.675	9	1.484	2.341
2b+3a+6a+3a*6a	7	0.542	0.807	8	1.612	2.567
all	7	0.560	0.688	8	1.500	2.359

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

About half the improvement was due to Rule 6a alone (ΔSS_{error}=0.18 or 2.0%), but the interactions of Rule 6a with Rules 2b and 3a were also somewhat helpful. The higher order interactions were also not included by design. As before (Experiment 1), there is no clear indication as to which combination of rules is most useful. The model that is plotted in the third panel of Figure 5.9 is the model that includes all terms. Note that regardless of this analysis, the application of the theory still misses some relatively minor boundaries near Notes 33/34 while generating false alarms at the surrounding points.

In addition, the simple correlations between the rules and the data are shown in Table 5.21. The intercorrelations between the rules are also included even though they are redundant with previous tables simply to make comparisons easy. Note that there were high correlations between Rule 6a (Pitch Pattern Parallelism A) and Rules 2b (Attack-Point) and 3a (Register). This was simply a consequence of the analysis since Rule 6a was constrained to generate boundaries at only those locations that had low level rule applications. Note the high correlation between Rule 6a and the interaction of Rule 6a and 2b. This is why both terms are not necessary in the final analysis.

Table 5.21: Correlation Matrix for "Three Blind Mice". Pitch Pattern Parallelism A.

	Rule						
	2b	3a	4	6a	2b*6a	3a*6a	2b*3a
data	.704**	.286*	.698**	.432**	.575**	.239	.534**
Rule 2b		.466*	.997**	.529**	.645**	.312*	.756**
Rule 3a			.533**	.428**	.222	.600**	.662**
Rule 4				.543**	.637**	.341*	.780**
Rule 6a					.849**	.831**	.277
Rule 2b*6a						.522**	.296*
Rule 3a*6a							.361*

Notes: * $p < .05$
** $p < .01$

Generally, Rule 6a was helpful for predicting boundaries. However, the effect was not a dramatic improvement on the predictions afforded by the best combination of Rules

2b and 3a (particularly 2b). Where it fails most strikingly was at Note 15 (as in Experiment 1).

In the second of the final analyses, Pitch Pattern Parallelism B (Rule 6b) also adds predictability to the model (Table 5.22 & Figure 5.8). Rules 6a & 6b are presented in the same figure to allow a direct comparison of the two versions of Pitch Pattern Parallelism. The different terms, Rule 6b or the interactions of Rule 6b with Rules 2b and 3a, reduced the error below the value of 8.874 that was achieved by Rule 4 (Intensification; the optimal combination of Rules 2a and 3d: Rule 4 is repeated in Table 5.22 for comparison) to a value of 8.536 (for ΔSS_{error}=0.34) which was slightly better the Pitch Pattern Parallelism A. In this case, as in the previous, parallelism helps by allowing the constants preceding Rules 2a and 3a to be reduced in magnitude, which helps to decrease false alarms.

However, the interaction of Rule 6b with Rule 2b was not useful and the interaction of Rule 6b with 3a was marginally at best (there was less than 0.1% improvement over Rule 6b alone). The higher order interactions were also not included by design. Though it is not completely unambiguous, it would seem that Rule 6b alone was the most useful addition to the model: It produced a ΔSS_{error}=0.33 or 3.7% improvement while the combination of Rule 6b and its interaction with 3a produced a ΔSS_{error}=0.34 or 3.8%. The model plotted in Figure 5.8 in the third panel includes all terms (for consistency with other Rules).

Table 5.22: The Combination of Rules 2, 3 (Rule 4) and 6b for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.124*2b+0.059*3a	8.874
2b+3a+6b	15	b = 0.961*2b+0.048*3a+0.280*6b	8.543
2b+3a+6b+2b*6b	15	b = 0.961*2b+0.048*3a+0.280*6b+0.000*2b*6b	8.543
2b+3a+6b+3a*6b	15	b = 0.968*2b+0.040*3a+0.231*6b+1.224*3a*6b	8.536
all	15	b = 0.968*2b+0.040*3a+0.231*6b + 0.000*2b*6b+1.224*3a*6b	8.536

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	6	0.639	0.816	9	1.698	2.645
2b+3a+6b	6	0.679	0.871	9	1.326	2.295
2b+3a+6b+2b*6b	6	0.679	0.871	9	1.326	2.295
2b+3a+6b+3a*6b	6	0.667	0.858	9	1.331	2.264
all	6	0.667	0.858	9	1.331	2.264

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

In addition, the simple correlations between the rules and the data are shown in Table 5.23 (along with the redundant intercorrelations between the rules). Note the high correlation between Rule 6b and the interaction of 6b and 2b: This was the reason both were not needed in the equation. Generally, Rule 6b was helpful for predicting boundaries. However, the effect was not a dramatic improvement on the predictions afforded by the best combination of Rules 2b and 3a (particularly 2b). Rule 6b was very similar to Rule 6a in its effects (as should be).

Table 5.23: Correlation Matrix for "Three Blind Mice", Pitch Pattern Parallelism B.

	Rule						
	2b	3a	4	6b	2b*6b	3a*6b	2b*3a
data	.704**	.286*	.698**	.647**	.650**	.534**	.534**
Rule 2b		.466*	.997**	.529**	.645**	.312*	.756**
Rule 3a			.533**	.428**	.222	.600**	.662**
Rule 4				.543**	.637**	.341*	.780**
Rule 6a					.964**	.748**	.440**
Rule 2b*6b						.699**	.503**
Rule 3a*6b							.602**

Notes: * $p < .05$
** $p < .01$

In the third of the final analyses, Time Pattern Parallelism A (Rule 6c) did not add predictability to the model (Table 5.24; Figure 5.10). Of all the terms, the best improvement was only ΔSS_{error}=0.03 or 0.3% achieved by the interaction of Rule 6c and 2b.

Table 5.24: The Combination of Rules 2, 3 (Rule 4) and 6c for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.124*2b+0.059*3a	8.874
2b+3a+6c	15	b = 1.111*2b+0.045*3a+0.023*6c	8.871
2b+3a+6c+2b*6c	15	b = 1.034*2b+0.066*3a+0.000*6c+0.122*2b*6c	8.843
2b+3a+6c+3a*6c	15	b = 1.133*2b+0.006*3a+0.000*6c+0.965*3a*6c	8.865
all	15	b = 1.034*2b+0.066*3a+0.000*6c + 0.122*2b*6c+0.000*3a*6c	8.843

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	6	0.639	0.816	9	1.698	2.645
2b+3a+6c	6	0.639	0.804	9	1.694	2.652
2b+3a+6c+2b*6c	5	0.622	0.753	10	1.683	2.604
2b+3a+6c+3a*6c	6	0.617	0.809	9	1.710	2.659
all	6	0.622	0.753	9	1.683	2.604

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

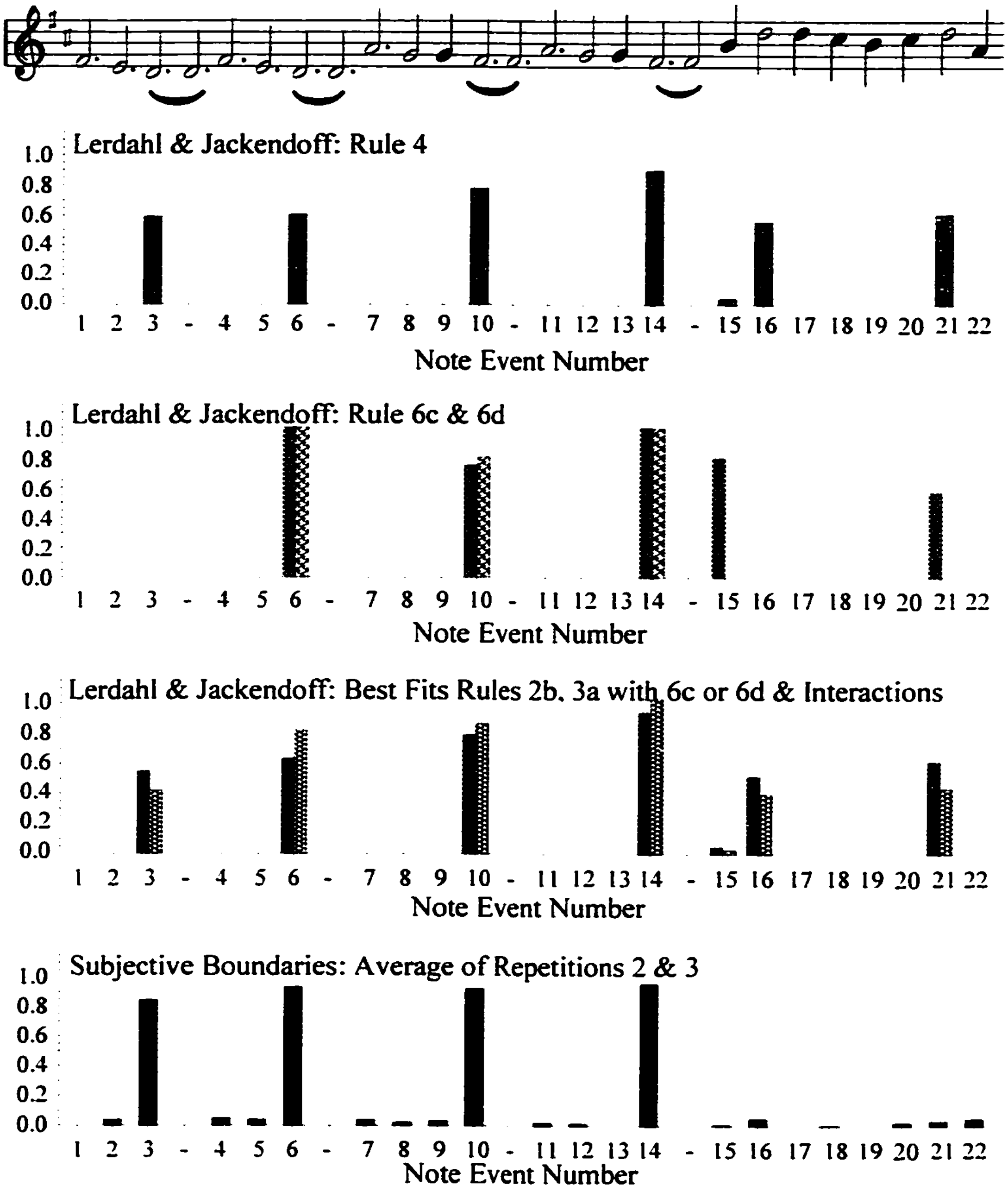


Figure 5.10a: The first part of the melodyused in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a alone and in combination with 6c or 6d). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

■ **Rule 4**
■ **Rules 2b, 3a & 6c**
▨ **Rules 2b, 3a & 6d**
■ **Best Fit for Rules 2b, 3a, 6c & their interactions**
▨ **Best Fit for Rules 2b, 3a, 6d & their interactions**
■ **Average Profile (Repetitions 2 & 3)**

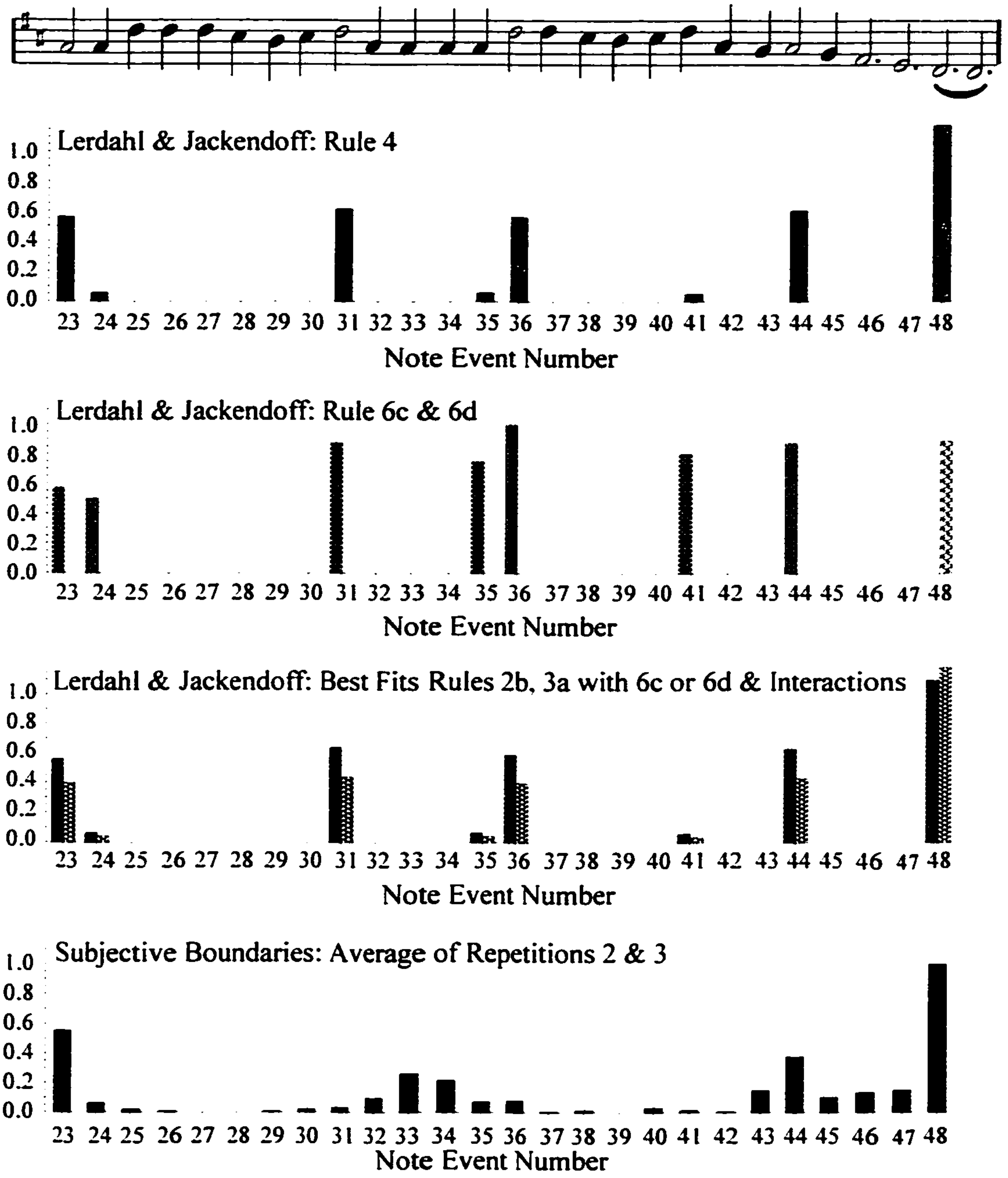


Figure 5.10b: The second part of the melody used in Stage 4, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a together & in combination with 6a or 6b). This can be compared with the location of subjective boundaries based on the average of the second and third repetition for all subjects.

■ **Rule 4**
■ **Rules 2b, 3a & 6c**
▨ **Rules 2b, 3a & 6d**
■ **Best Fit for Rules 2b, 3a, 6a & their interactions**
▨ **Best Fit for Rules 2b, 3a, 6d & their interactions**
■ **Average Profile (Repetitions 2 & 3)**

The simple correlations between the rules and the data are shown in Table 5.25 (along with the redundant intercorrelations between the rules). Note that Rule 6c and the interaction of 6c with 3a did not have high simple correlations with the empirical data.

Table 5.25: Correlation Matrix for "Three Blind Mice", Time Pattern Parallelism A.

	Rule						
	2b	3a	4	6c	2b*6c	3a*6c	2b*3a
data	.704**	.286*	.698**	.342*	.519**	.113	.534**
Rule 2b		.466*	.997**	.529**	.645**	.312*	.756**
Rule 3a			.533**	.428**	.222	.600**	.662**
Rule 4				.543**	.637**	.341*	.780**
Rule 6c					.807**	.782**	.305*
Rule 2b*6c						.402**	.357*
Rule 3a*6c							.381**

Notes: * p < .05
** p < .01

In the fourth and last of the final analyses, Time Pattern Parallelism B (Rule 6d) did add predictability to the model (Table 5.26; Figure 5.10; Rules 6c and 6d are presented together to facilitate comparisons). The best improvement was $\Delta SS_{error}=0.60$ or 6.8% achieved by the Rule 6d alone. None of the interactions contributed anything.

Table 5.26: The Combination of Rules 2, 3 (Rule 4) and 6d for "Three Blind Mice".

Overall Results			
Rule	n	Equation	SS_{error}
2b+3a	15	b = 1.124*2b+0.059*3a	8.874
2b+3a+6d	15	b = 0.797*2b+0.043*3a+0.394*6d	8.274
2b+3a+6d+2b*6d	15	b = 0.797*2b+0.043*3a+0.394*6d+0.000*2b*6d	8.274
2b+3a+6d+3a*6d	15	b = 0.797*2b+0.043*3a+0.394*6d+0.000*3a*6d	8.274
all	15	b = 0.797*2b+0.043*3a+0.394*6d + 0.000*2b*6d+0.000*3a*6d	8.274

Breakdown for Misses and False Alarms						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a	6	0.639	0.816	9	1.698	2.645
2b+3a+6d	6	0.778	0.806	9	0.958	1.835
2b+3a+6d+2b*6d	6	0.778	0.806	9	0.958	1.835
2b+3a+6d+3a*6d	6	0.778	0.806	9	0.958	1.835
all	6	0.778	0.806	9	0.958	1.835

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

In addition, the simple correlations between the rules and the data are shown in Table 5.27 (along with the now very redundant intercorrelations between the rules). Note that Rule 6d was highly correlated with its interactions and that the two interactions involving 6d were highly intercorrelated. Also note that, of Rule 6d and the interactions of 6d with 2b and 3a, Rule 6d had the highest correlation with the empirical data. This was the reason that only 6d itself was necessary.

Table 5.27: Correlation Matrix for "Three Blind Mice", Time Pattern Parallelism B.

	Rule						
	2b	3a	4	6d	2b*6d	3a*6d	2b*3a
data	.704**	.286*	.698**	.759**	.749**	.628**	.534**
Rule 2b		.466*	.997**	.529**	.645**	.312*	.756**
Rule 3a			.533**	.428**	.222	.600**	.662**
Rule 4				.543**	.637**	.341*	.780**
Rule 6d					.969**	.804**	.560**
Rule 2b*6d						.792**	.620**
Rule 3a*6d							.763**

Notes: * $p < .05$
** $p < .01$

Boundary Efficacy: Stage 4, Part 2

To assess whether or not boundaries affect subsequent processing, a secondary detection task (click detection) task was used. In this task, subjects listened to the melody, and simultaneously listened for clicks. Subjects indicated the detection of a click as rapidly (and as accurately) as possible with a keypress. The time of detection was measured from the onset of the click, in milliseconds.

Given the basic hypothesis that click detection will be relatively delayed before a boundary and relatively advanced after a boundary, the analysis focussed on the comparison of delays for clicks that preceded boundaries versus clicks that followed boundaries versus clicks that straddled boundaries.

However, before the main analysis, subjects were equated by converting reaction times (RTs) to z-scores within each subject. This means that all subjects were converted so that all had the same mean and standard deviation for RTs (0.00±1.00). This analysis allows one to see which points in the melody produce longer RTs without regard to individual differences in RTs (both differences in means and in ranges).

After the conversion, RTs were correlated against boundary location, both previous and subsequent. The predicted relationships would include faster RTs just after a boundary and slower RTs just before a boundary. From Figure 5.11a, one can see this general pattern in the first few groups of notes (4 though 10) . However the pattern begins to degrade after Note 10. For the strong boundary at Note 14, click detection is actually faster before than after the boundary. Since this is the region with the highest degree of parallelism, it is possible that the information processing requirements are not so great (i.e., subjects "know" what is coming so they can attend to other stimuli). RTs get longer after Note 14, in the first region of the melody that depicts a change from the previous parallel structure (regardless of whether or not one accepts the previous analysis of parallelism, it is obvious that there is more parallelism before Note 14 than after Note 14). The reasonably strong boundary at Note 23 has a relatively longer RT than its immediate neighbours, as does the relatively weak boundary in the area of Notes 33 or 34. Generally, the visual analysis implies some consistency with prediction.

Analytically, the situation is much more complex. Ideally, one would like to see lower RTs after a boundary, higher RTs before a boundary. Hence, one might represent this most simply as a linearly increasing (or, at least, a monotonically increasing) RT between boundaries, with boundaries themselves representing some point between the two extremes (see Figure 5.1). However, this basic approach fails to consider what click detection is about (cf., the introduction of this chapter). That is, changes in the processing demands (or information content) of the melody are assumed to affect the ability to process a second concurrent task. In this experiment, the demands for processing the melody affect the ability to detect and respond to clicks. One, therefore, should consider changes in RT that would be associated with melodic complexity. For example, the opening of this melody is very slow, corresponding to low information density per unit time, while the middle sections are much quicker corresponding to higher information density per unit time. By informal inspection, this does seem to be a factor in this melody. RTs in the opening part are generally much faster than RTs in the later sections. Hence, the analysis of RT must proceed at the local level; that is, the analysis must operate at a level that considers position within the melody.

In this analysis, click detection was related to the boundary profile created in the third repetition of the melody (not the average of the third and second repetitions). This is because the experimental design coded all information that related clicks to boundary positions by references to those boundary positions notated in the third repetition of the melody (see the previous Methods section). Spoiled trials were dropped from the analysis, so that all analyses were based only on the valid responses (i.e., the conversion to z-scores). In fact, there were very few spoiled trials (possibly indicating that the task was too easy): Only 5 of 496 trials (62 subjects, each having 8 trials) were lost.

As an initial assessment, one could ignore much of the information contained within the data and examine RTs for clicks placed on notes before boundaries, after boundaries and on boundaries. In the actual analysis, the RTs were analysed for the following conditions:

> clicks that fell on a note that was two notes in front of a boundary and at least three notes after the previous boundary
>
> clicks that fell on a note that was one note in front of a boundary and at least three notes after the previous boundary
>
> clicks that fell on a note that was a boundary
>
> clicks that fell on a note that was one note after a boundary and at least three notes before the next boundary
>
> clicks that fell on a note that was two notes in after a boundary and at least three notes before the next boundary.

The selection criteria reduced the number of boundaries from a total of 491 to 330, with 223 of those being on boundaries (hence, 161 clicks were placed between boundaries, but more than 2 notes from any boundary).

There was a tendency to be faster after a boundary and slowest on a boundary (see

Table 5.28). The overall analysis was not significant, although the comparison of clicks on boundaries to click on notes adjacent to boundaries was marginal.

Table 5.28: RT (z-scores) for clicks placed just before, on, or just after boundaries and the analysis of those RTs in "Three Blind Mice".

Boundary Position	n	v	Mean ± SD	Median	Mode	Min	Max
Two Before	45	1	-0.189 ± 1.001	-0.361	-0.829	-1.560	2.449
Before	16	0	-0.068 ± 0.720	-0.272	-1.122	-1.122	1.438
On	223	4	0.080 ± 1.000	-0.214	-0.688	-2.312	2.470
Just After	14	0	-0.353 ± 0.794	-0.401	-1.462	-1.462	1.830
Two After	32	0	0.062 ± 0.954	-0.095	1.061	-1.473	2.469
Overall	330	5	0.016 ± 0.979	-0.270	-0.688	-2.312	2.470

WS-ANOVA	SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Phi Overall	4.75	4	314.41	330	1.28	0.278	0.016	0.003
Ψ_1	1.06	1			1.11	0.292	0.004	
Ψ_2	0.92	1			0.96	0.327	0.003	
Ψ_3	0.00	1			0.00	0.950	0.000	
Ψ_4	2.90	1			3.04	0.082	0.008	

where: v refers to the number of invalid trials in that cell
Ψ_1 compares the two Before conditions
Ψ_2 compares the two After conditions
Ψ_3 compares Before to After.
Ψ_4 compares Before and After to On.

A correlational analysis was also used to tease out the effects. In this analysis, variables coding for position of click within the melody (to the third order), boundary type (on or off a boundary), distance from the previous boundary (0 if on a boundary) and distance to the next boundary (0 if on a boundary) were used. Note that previous and next boundaries were really a finer delineation of one half of the on/off boundaries (i.e., the off was subdivided into specific distances). Variables were entered hierarchically, with melody position first, boundary type second and the distances third. As can be gleaned from Figure 5.11 and Table 4.29, results did not indicate a particular strong relationship. The type of click (on a boundary or not) did add predictability to the equation even after position within the melody had been controlled for[56]. This is similar to the corresponding contrast in Table 3.28; the inclusion of the covariates helped to achieve significance. Clicks on boundaries were slower than clicks not on boundaries. Also note that the combination of previous and post distances did not add predictability. Interestingly, these variables did add predictability when the previous boundary and next boundary distances were included for clicks on boundaries (i.e., when a click occurred on a boundary, the previous distances was the number of notes between the current boundary and the previous boundary and the next distance was the number of notes between the current boundary and the next boundary). Regardless of significance, the effect size was only 2%.

Table 5.29: The correlational analysis of the RT to clicks within the melody "Three Blind Mice".

Variable	simple r	part r	partial r	ΔR^2	$F(\Delta R^2)$	$p(\Delta R^2)$
Position	0.066	0.199	0.203	0.043	7.29	0.000
Position^2	0.122	0.093	0.096			
Position^3	-0.001	-0.167	-0.170			
Boundary	0.078	0.146	0.149	0.018	9.27	0.003
Previous Boundary	0.052	0.070	0.072	0.006	1.661	0.191
Next Boundary	-0.043	0.017	0.018			
Total Effect				0.068	5.82	0.000

---

[56] Properly, the analysis should account for the fact that z-scores have been used as the dependent variable by reducing the df for the error term by 61. However, this does not change the results.

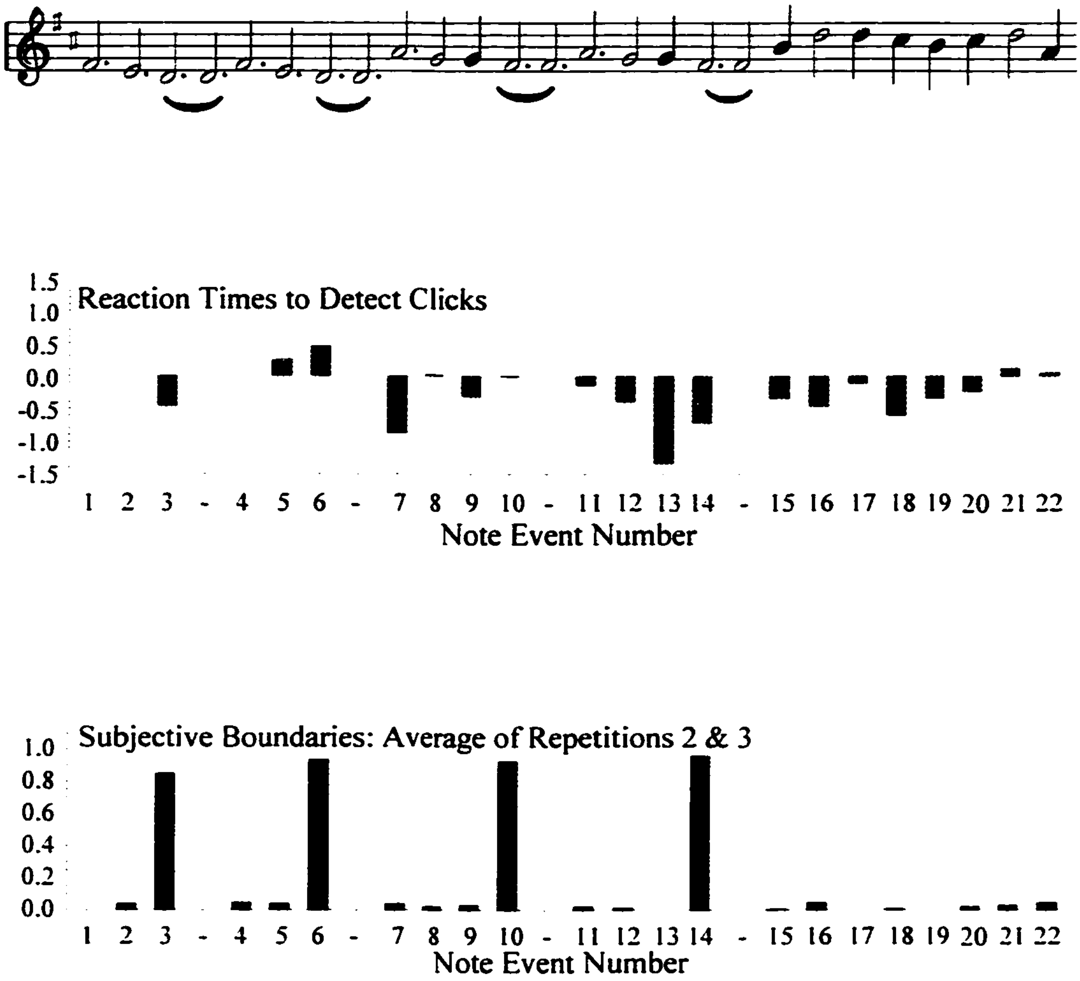
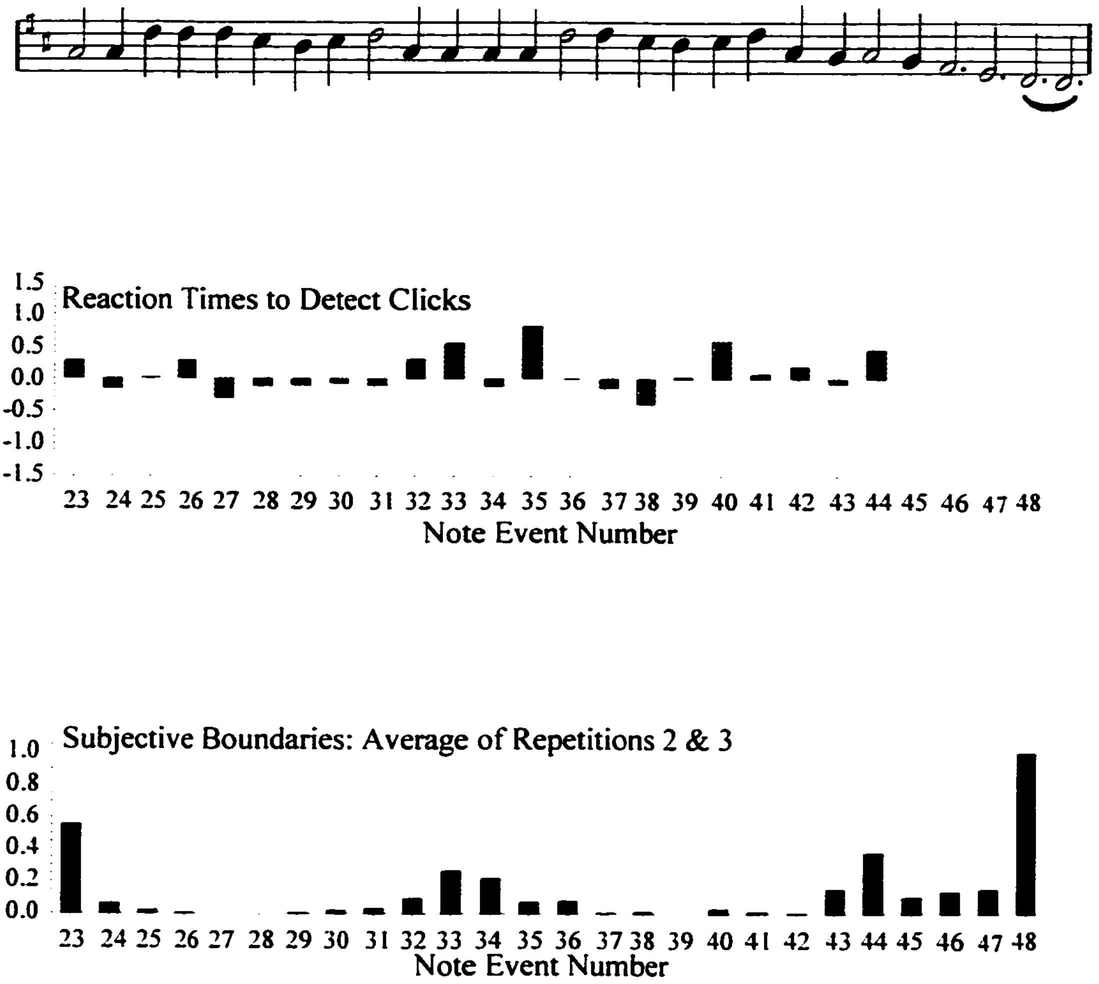


Figure 5.11a: The first part of the song used in Stage 4, Part 1, the location of subjective boundaries based on the average of the second and third repetition for all subjects, and the z-score of the time to detect clicks placed at various locations.

■ Average Profile (Repetitions 2 & 3)

■ Z-score of time to detect click



Reaction Times to Detect Clicks

Subjective Boundaries: Average of Repetitions 2 & 3

Figure 5.11b: The second part of the song used in Stage 4, Part 1, the location of subjective boundaries based on the average of the second and third repetition for all subjects, and the z-score of the time to detect clicks placed at various locations.

■ Average Profile (Repetitions 2 & 3)

■ Z-score of time to detect click

Generally, the results for the click analysis were minimal. Several other analyses were considered and a number conducted (various permutations on the relationships between boundary location and click position), but none were fruitful. The reason for this can be seen in Figure 5.11. Since the task was intended to be secondary to the main purpose of this thesis, these alternatives are not discussed.

Boundary Location: Stages 6 and 8, Part 1

In Part 1 of Stages 6 and 8, (see Table 5.30), the mean consistency values for subject-by-subject repetitions of the extract from the melody "Softly Now the Light of Day" were 0.62±0.04 and 0.65±0.01 respectively (phi measure only: the error is the standard error). All values indicated a high degree of similarity within individual subjects. The modal value for adjacent repetitions was 1.0 dropping when the repetitions most distant in time. In general, later repetitions (e.g., Stage 8) showed higher consistency values with their nearest neighbours in time. By the standard deviations only a few subjects fell to the level of chance indicated in Table 3.8 (i.e., given a mean of .60, a standard deviation of .25, and a cutoff of .30, then 11% of the scores are below cutoff). Recall that these chance levels assumed equal numbers of boundaries per boundary profile and a particular probability for boundaries within a boundary profile.

Table 5.30: The average consistency of subjects over different repetitions of the same melody within Stages 6 & 8, Part 1, given the Phi similarity measures of Table 3.8.

	Consistency	Mean & SD	Median	Mode	Minimum	Maximum
Phi Similarity	1 to 2	0.606±0.238	0.599	1.000	0.139	1.000
	1 to 3	0.579±0.229	0.602	0.897	0.101	1.000
	2 to 3	0.662±0.249	0.717	1.000	0.106	1.000
	1 to 4	0.628±0.245	0.653	1.000	0.114	1.000
	1 to 5	0.614±0.237	0.622	0.461	-0.054	1.000
	1 to 6	0.588±0.265	0.599	0.897	-0.176	1.000
	2 to 4	0.625±0.243	0.622	1.000	0.175	1.000
	2 to 5	0.642±0.267	0.679	1.000	0.060	1.000
	2 to 6	0.623±0.299	0.595	1.000	-0.278	1.000
	3 to 4	0.613±0.254	0.612	0.897	0.040	1.000
	3 to 5	0.629±0.266	0.646	1.000	0.146	1.000
	3 to 6	0.626±0.274	0.679	1.000	0.052	1.000
	4 to 5	0.646±0.239	0.676	1.000	-0.130	1.000
	4 to 6	0.641±0.244	0.622	0.897	0.093	1.000
	5 to 6	0.659±0.291	0.679	1.000	-0.095	1.000

To further examine the consistency ratings, the correlation between consistencies were computed across subjects. That is, each subject produced 15 consistency values (phi measures: 1 to 2, 1 to 3, 2 to 3, 1 to 4, 1 to 5, 1 to 6, etc.) and the correlations between these consistency values was assessed. The entire 15x15 matrix is not reproduced (see Table 5.31). Rather, only those correlations between the first three repetitions (Stage 6), the last three repetitions (Stage 8) and the average of the first three to the last three repetitions (Stage 6 to 8) are shown. The resulting correlations were all high (and significant at a type 1 error rate of α=0.05), indicating that the subjects who were the most consistent in the first two repetitions remained the most consistent in the third repetition. Note however, that the nearest neighbours in time had the highest intercorrelations, implying that there was some drift in the responses of subjects (though it was not too great).

Table 5.31: The correlations between consistency ratings over the different repetitions of the same melody within Stages 6 and 8, Part 1, given the phi similarity measure of Table 3.8.

		Stage 6 Correlations			Stage 8 Correlations		
		1 to 2	1 to 3	2 to 3	4 to 5	4 to 6	5 to 6
Stage 6 Correlations	1 to 2	1.000	0.609**	0.513**	0.482**	0.434**	0.387**
	1 to 3		1.000	0.491**	0.439**	0.537**	0.534**
	2 to 3			1.000	0.594**	0.559**	0.522**
Stage 8 Correlations	4 to 5				1.000	0.688**	0.653**
	4 to 6					1.000	0.680**
	5 to 6						1.000

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

As shown in the within-subjects ANOVA of Table 5.32, consistency ratings changed slightly as a function of the repetitions being compared but the effect was not particularly strong (note the η^2 and ω^2 values: this is a within-subjects design and one would expect these to be much higher).

Table 5.32: The ws-analysis of the change in consistency ratings for the different repetitions of the same melody within Stages 6 & 8, Part 1, given the phi similarity measure of Table 3.8. All computations were made on log transformed scores.

WS-ANOVA		SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Overall		0.473	14	24.500	854	1.180	0.287	0.019	0.003
Stage 6	Ψ_1	0.098	1	1.762	61	3.407	0.070	0.053	
	Ψ_2	0.124	1	1.468	61	5.137	0.027	0.078	
Stage 8	Ψ_3	0.005	1	1.550	61	0.212	0.647	0.003	
	Ψ_4	0.005	1	1.193	61	0.247	0.621	0.004	
Stage 6 to 8	Ψ_5	0.101	1	2.691	61	2.279	0.136	0.036	
	Ψ_6	0.028	1	0.956	61	1.793	0.186	0.029	

Notes: Ψ_1 & Ψ_3 compare Consistency 1 to 2 (4 to 5) with 2 to 3 (5 to 6).

Ψ_2 & Ψ_4 compare the average Consistency of 1 to 2 and 2 to 3 (4 to 5 and 5 to 6) with Consistency 1 to 3 (4 to 6).

Ψ_5 compares the averages within Stage 6 to those within Stage 8.

Ψ_6 compares the averages of Stages 6 and 8 to the comparisons across Stages 6 and 8 (i.e., 1 to 4 through 3 to 6).

When comparing across subjects (Table 5.33), average similarity ratings were lower, as before (see Chapter 3 of this work) with a mean of 0.55±0.24 and 0.56±0.25 for Stages 6 and 8 respectively. This value was still above the critical and chance levels specified in Table 3.8, and in fact, only the minimum dipped below the critical values cited. As can be seen in the subsequent cluster analysis, this only occurred for a few subjects (3 or 4 depending upon what one wishes to call chance).

Table 5.33: The average similarity ratings between different subjects for the same melody, within Stages 6 & 8, Part 1, Repetition 3 only, given the phi similarity measure of Table 3.8.

Stage	Mean & SD	Median	Mode	Minimum	Maximum
Stage 6	0.547±0.243	0.528	1.000	-0.113	1.000
Stage 8	0.564±0.253	0.595	0.897	-0.209	1.000

Overall, the central tendencies in the similarity analyses (mean, mode, median) indicate that subjects perform more-or-less the same. However the minimums in both the consistency and the similarity analyses imply that a few subjects may be somewhat different.

To further explore the relationship between different subjects, a cluster analysis was performed on the basis of the phi similarity measure determined from the third (the last of Stage 6) and sixth (the last of Stage 8) repetitions of the melody. These results did not differ very much (see Figure 5.12). As can be seen, these analyses were indicative of one large group to which individuals are gradually accrued. Using the previous critical or chance values one can see that the clustering would consist of one large group with three or four stragglers. As such, the cluster profiles indicate that all subjects should be treated as one homogeneous group (one could argue that 3 or 4 subjects be dropped, but relative to 62, this in not an important factor).

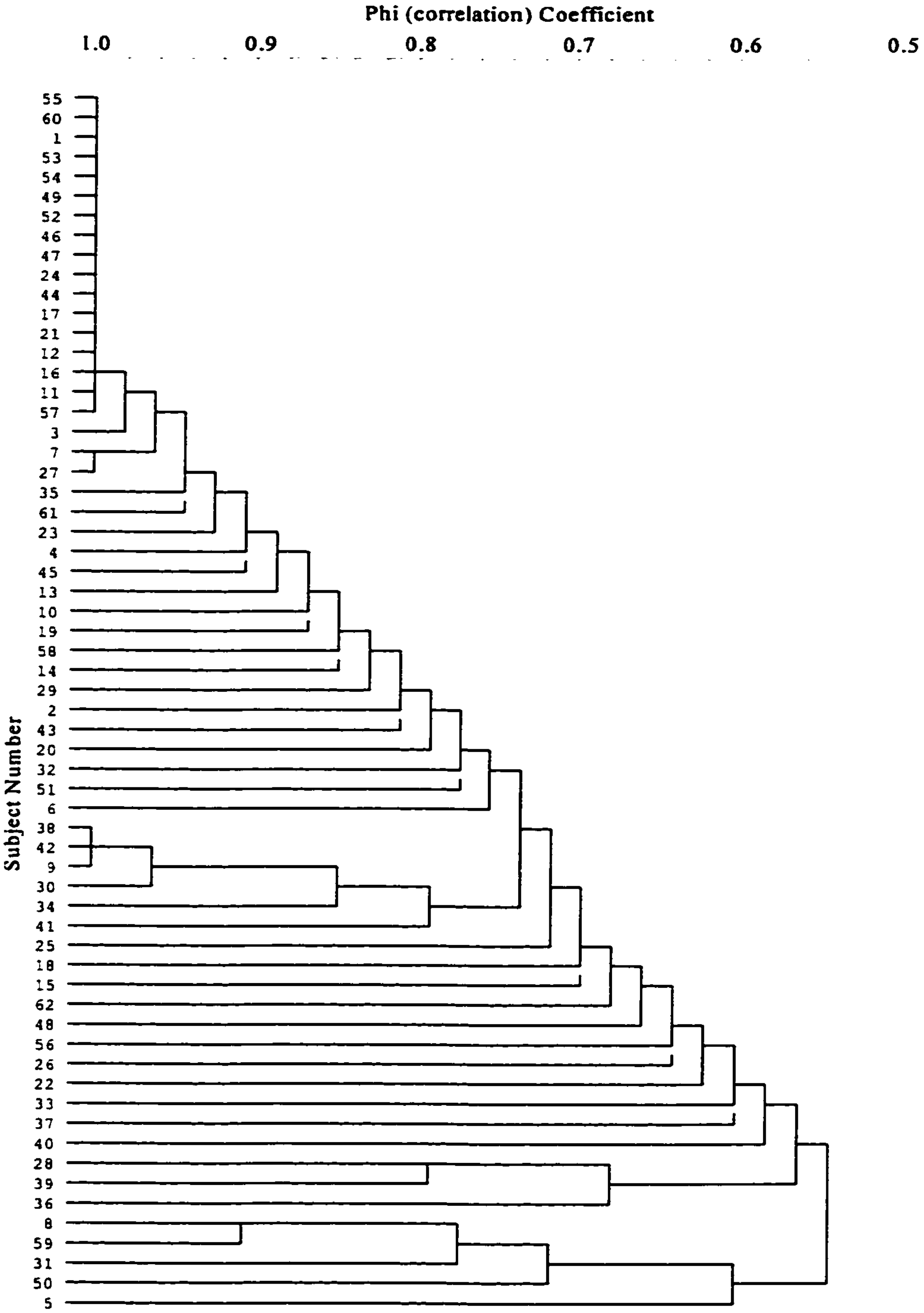


Figure 5.12a: Cluster analysis of boundary profiles of individual subjects using the Phi measure (the binary form of the correlation coefficient) for the third repetition of the melody used in Stage 6. Note that the maximum correlation was 1.0 and the minimum value was 0.54 indicating that all subjects were fairly similar. Subject numbers are arbitrary.

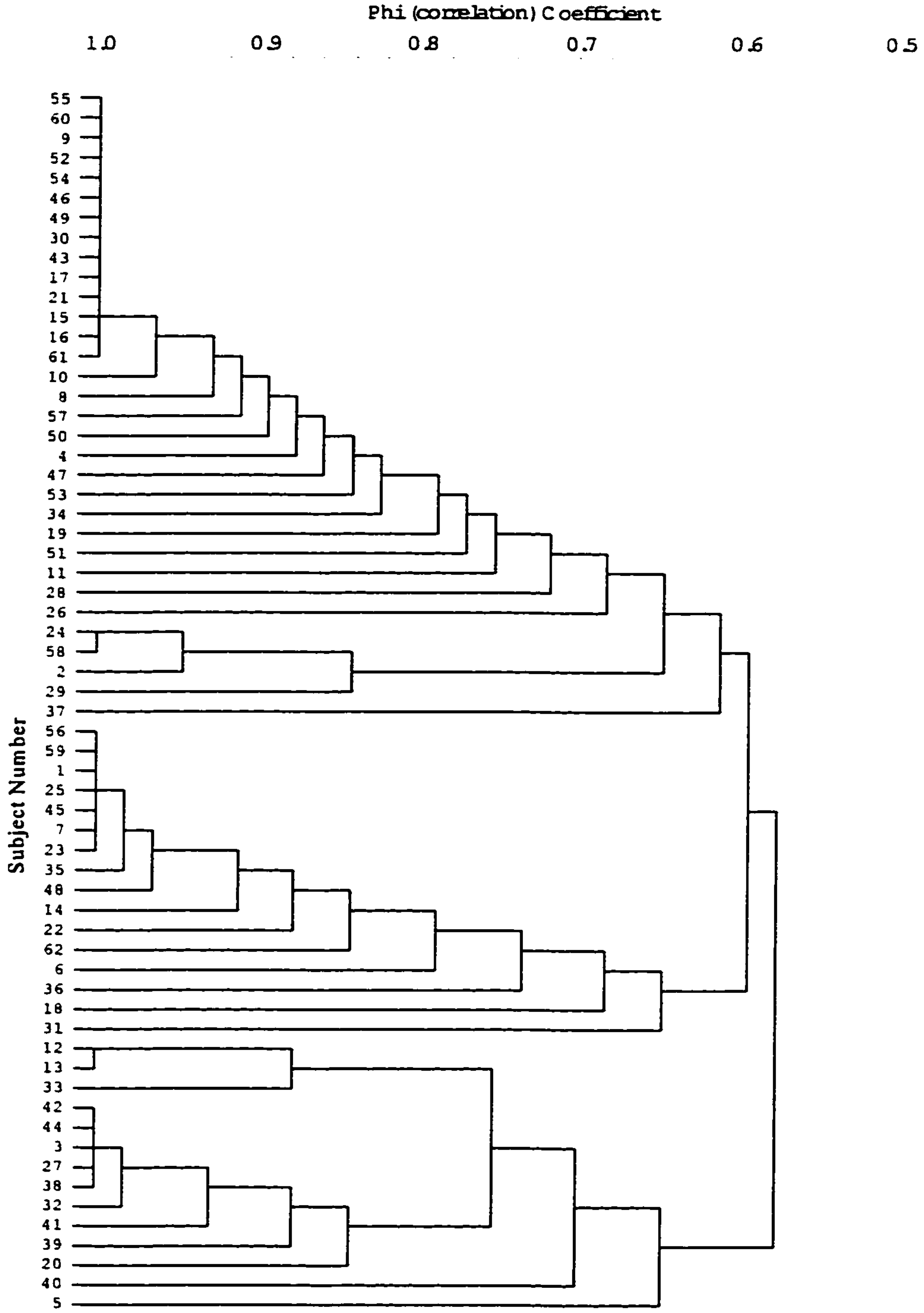


Figure 5.12b: Cluster analysis of boundary profiles of individual subjects using the Phi measure (the binary form of the correlation coefficient) for the third repetition of the melody used in Stage 8. Note that the maximum correlation was 1.0 and the minimum value was 0.56 indicating that all subjects were fairly similar. Subject numbers are arbitrary but the same as in Figure 5.12a

The detailed descriptive statistics for this profile are provided in Tables 5.34 through 5.39 (various measures pertaining to the locations of boundaries) and Tables 5.40 5.45 (correlations between those measures of boundary location) for each repetition of the melody. Over all repetitions, the mean number of boundaries was fairly constant at 6.42±0.42 (in this case, the standard error is cited). The modal number of boundaries was constant at 5.00. This number of boundaries represents a mean number of note events *between* boundaries equal to 6.14±1.08; the mean number of note events within a unit should therefore be 7.14 because each boundary should be associated with either the beginning or the end of a unit (i.e., boundaries are located on particular note events). The size of individual units ranged from 0.00 note events to 32.00 (effectively the whole melody) note events. However, on a subject by subject basis, the most common size was 8.00, 6.00 and 7.00 note events (implying 9.00, 5.00 and 8.00 note events when the boundary is included), with 8.00 being the most common in the early repetitions and 6.00 being the most common in the latter repetitions. The least common size was 9.00 (implying 10 when boundaries are included).

Overall, the pattern of the first repetition in Stage 6 was more similar to the pattern in the first repetition in Stage 8 than it is to the other repetitions in Stage 6. The same was true in Stage 8. This implies the same developmental course for parsing in Stages 6 and 8. There was a trend to slightly larger mean units, both within each stage and over both stages, and an increasing focus on a more limited set of lengths. From the correlation matrices, one can see that the minimum number of notes between boundaries (the smallest grouping size for each subject) was correlated with the maximum number of notes between boundaries (the largest grouping size for each subject) in all repetitions except the second and third. This positive correlation implies those subjects who used the largest maxima, also used the largest minima; that is, some subject tended to use larger units overall. In addition, the correlation matrices indicate that more boundaries was generally associated with the use of more smaller units (number of times units of size 0, 1, 2, 3, and 4 are used), but more boundaries was only associated with the decrease in the number of large units in the later repetitions within a stage (i.e., Repetitions 2, 3, 5 and 6).

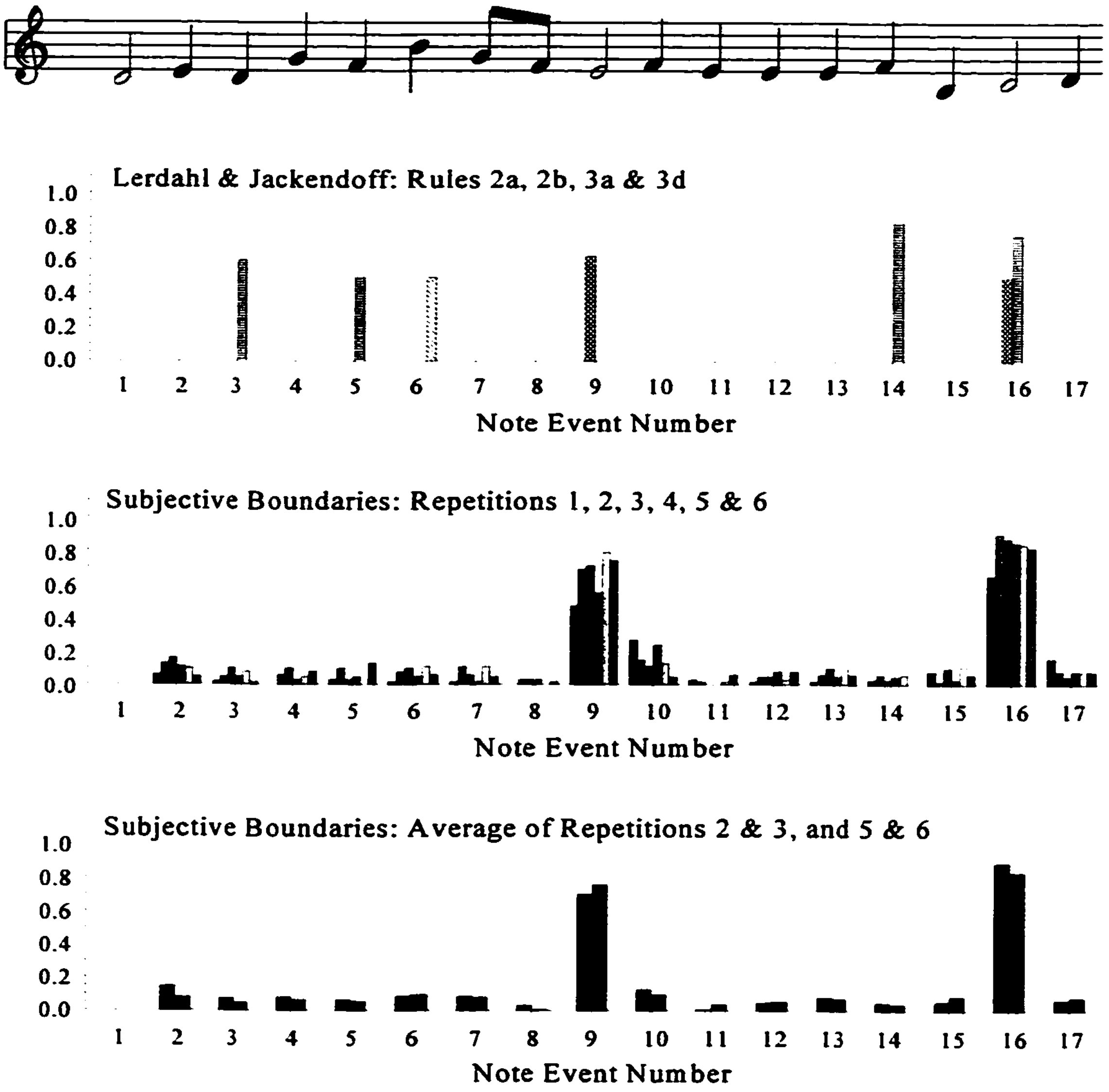


Figure 5.13a: The first part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries for each repetition of the song, and the average of the second and third repetitions and the averages of the fifth and sixth presentations (these averages is used in subsequent analyses). These profiles are based on all 62 subjects.

▨ Rule 2a	▨ Rule 3a	■ First Presentation	■ First Presentation
▩ Rule 2b	▨ Rule 3d	■ Second Presentation	☐ Second Presentation
		■ Third Presentation	■ Third Presentation

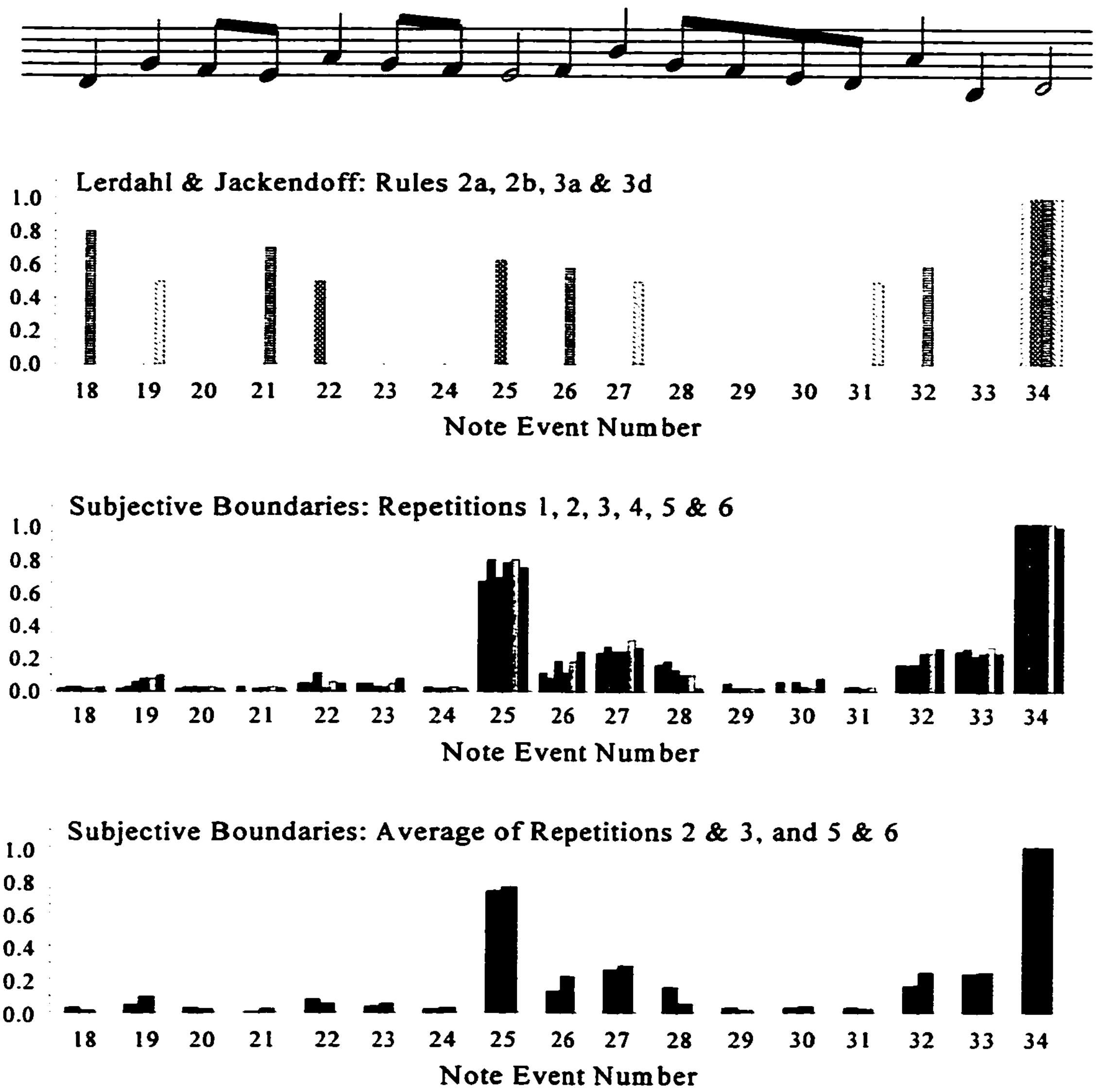


Figure 5.13b: The second part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2a, 2b, 3a & 3d). This can be compared with the location of subjective boundaries for each repetition of the song, and the average of the second and third repetitions and the averages of the fifth and sixth presentations (these averages is used in subsequent analyses). These profiles are based on all 62 subjects.

▨ Rule 2a
▩ Rule 2b
■ Rule 3a
▨ Rule 3d
■ First Presentation
■ Second Presentation
■ Third Presentation
■ First Presentation
▣ Second Presentation
■ Third Presentation

Table 5.34: The important statistics for the location of boundaries within Stage 6. Part 1, Repetition 1 (the first presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	4.806±8.073	1.000	0.000	0.000	32.000
Maximum	11.339±6.616	8.000	8.000	4.000	32.000
Mean	8.152±6.894	5.600	7.250	2.300	32.000
Number with 0	0.387±0.610	0.000	0.000	0.000	2.000
Number with 1	0.468±0.718	0.000	0.000	0.000	3.000
Number with 2	0.177±0.385	0.000	0.000	0.000	1.000
Number with 3	0.258±0.571	0.000	0.000	0.000	2.000
Number with 4	0.258±0.599	0.000	0.000	0.000	3.000
Number with 5	0.371±0.579	0.000	0.000	0.000	2.000
Number with 6	0.677±0.696	1.000	1.000	0.000	3.000
Number with 7	0.710±0.755	1.000	0.000	0.000	3.000
Number with 8	0.935±0.787	1.000	1.000	0.000	3.000
Number with 9	0.048±0.216	0.000	0.000	0.000	1.000
Number with ≥10	0.387±0.610	0.000	0.000	0.000	2.000

Notes: Mean, median, and mode refer to the number of note events between boundary locations (exclusive of the actual boundaries).

Number with "xx" is the mean number of times subjects created units containing xx note events (exclusive of the actual boundaries).

Table 5.35: The important statistics for the location of boundaries within Stage 6, Part 1, Repetition 2 (the second presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.371±2.600	1.000	0.000	0.000	7.000
Maximum	8.419±2.199	8.000	8.000	3.000	17.000
Mean	5.488±1.893	5.600	7.250	1.539	10.000
Number with 0	0.581±0.821	0.000	0.000	0.000	3.000
Number with 1	0.613±0.964	0.000	0.000	0.000	4.000
Number with 2	0.435±1.018	0.000	0.000	0.000	5.000
Number with 3	0.403±0.735	0.000	0.000	0.000	2.000
Number with 4	0.387±0.710	0.000	0.000	0.000	3.000
Number with 5	0.339±0.571	0.000	0.000	0.000	2.000
Number with 6	0.887±0.727	1.000	1.000	0.000	3.000
Number with 7	0.726±0.728	1.000	1.000	0.000	3.000
Number with 8	1.032±0.789	1.000	1.000	0.000	3.000
Number with 9	0.097±0.298	0.000	0.000	0.000	1.000
Number with ≥10	0.145±0.399	0.000	0.000	0.000	2.000

Notes: see Table 5.34

Table 5.36: The important statistics for the location of boundaries within Stage 6. Part 1, Repetition 3 (the third presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.419±2.779	1.000	0.000	0.000	8.000
Maximum	8.839±3.153	8.000	8.000	3.000	22.000
Mean	5.632±2.155	5.600	7.250	1.539	10.000
Number with 0	0.645±0.889	0.000	0.000	0.000	4.000
Number with 1	0.500±0.805	0.000	0.000	0.000	4.000
Number with 2	0.548±1.066	0.000	0.000	0.000	5.000
Number with 3	0.419±0.860	0.000	0.000	0.000	4.000
Number with 4	0.306±0.589	0.000	0.000	0.000	2.000
Number with 5	0.387±0.610	0.000	0.000	0.000	2.000
Number with 6	0.871±0.735	1.000	1.000	0.000	3.000
Number with 7	0.645±0.546	1.000	1.000	0.000	2.000
Number with 8	0.984±0.896	1.000	0.000	0.000	3.000
Number with 9	0.081±0.275	0.000	0.000	0.000	1.000
Number with ≥10	0.226±0.422	0.000	0.000	0.000	1.000

Notes: see Table 5.34

Table 5.37: **The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 1 (the fourth presentation of the melody).**

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.968±4.852	1.000	0.000	0.000	32.000
Maximum	10.048±4.415	8.000	8.000	5.000	32.000
Mean	6.562±4.136	5.600	7.250	1.750	32.000
Number with 0	0.581±0.950	0.000	0.000	0.000	4.000
Number with 1	0.597±0.877	0.000	0.000	0.000	4.000
Number with 2	0.355±0.851	0.000	0.000	0.000	4.000
Number with 3	0.290±0.611	0.000	0.000	0.000	2.000
Number with 4	0.274±0.548	0.000	0.000	0.000	2.000
Number with 5	0.387±0.636	0.000	0.000	0.000	3.000
Number with 6	0.645±0.630	1.000	1.000	0.000	2.000
Number with 7	0.677±0.672	1.000	1.000	0.000	2.000
Number with 8	0.968±0.923	1.000	0.000	0.000	3.000
Number with 9	0.097±0.298	0.000	0.000	0.000	1.000
Number with ≥10	0.355±0.575	0.000	0.000	0.000	2.000

Notes: see Table 5.34

Table 5.38: The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 2 (the fifth presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.000±2.476	1.000	0.000	0.000	6.000
Maximum	8.161±2.299	8.000	8.000	4.000	17.000
Mean	5.283±1.891	5.600	5.600	1.750	10.000
Number with 0	0.710±0.982	0.000	0.000	0.000	3.000
Number with 1	0.645±0.832	0.000	0.000	0.000	3.000
Number with 2	0.435±1.018	0.000	0.000	0.000	4.000
Number with 3	0.435±0.898	0.000	0.000	0.000	4.000
Number with 4	0.419±0.714	0.000	0.000	0.000	3.000
Number with 5	0.387±0.636	0.000	0.000	0.000	2.000
Number with 6	0.903±0.670	1.000	1.000	0.000	3.000
Number with 7	0.790±0.681	1.000	1.000	0.000	2.000
Number with 8	0.871±0.799	1.000	0.000	0.000	2.000
Number with 9	0.113±0.319	0.000	0.000	0.000	1.000
Number with ≥10	0.145±0.399	0.000	0.000	0.000	2.000

Notes: see Table 5.34

Table 5.37: The important statistics for the location of boundaries within Stage 8, Part 1, Repetition 3 (the sixth presentation of the melody).

	Mean & SD	Median	Mode	Minimum	Maximum
Minimum	2.581±3.351	1.000	0.000	0.000	15.000
Maximum	8.177±2.287	8.000	8.000	4.000	17.000
Mean	5.710±2.440	5.600	7.250	2.000	15.500
Number with 0	0.516±0.864	0.000	0.000	0.000	4.000
Number with 1	0.516±0.695	0.000	0.000	0.000	3.000
Number with 2	0.371±0.794	0.000	0.000	0.000	3.000
Number with 3	0.516±1.127	0.000	0.000	0.000	5.000
Number with 4	0.323±0.672	0.000	0.000	0.000	3.000
Number with 5	0.274±0.548	0.000	0.000	0.000	2.000
Number with 6	1.016±0.665	1.000	1.000	0.000	2.000
Number with 7	0.823±0.800	1.000	1.000	0.000	3.000
Number with 8	0.935±0.744	1.000	1.000	0.000	2.000
Number with 9	0.065±0.248	0.000	0.000	0.000	1.000
Number with ≥10	0.161±0.451	0.000	0.000	0.000	2.000

Notes: see Table 5.32

Table 5.40: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 1 (the first presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.843**	1.000	
Mean	.977**	.927**	1.000
Number with 0	-.384**	-.256*	-.335**
Number with 1	-.346**	-.338**	-.370**
Number with 2	-.237	-.198	-.235
Number with 3	-.235	-.271*	-.276*
Number with 4	-.234	-.205	-.262*
Number with 5	-.275*	-.281*	-.283*
Number with 6	-.353**	-.378**	-.347**
Number with 7	-.243	-.325*	-.272*
Number with 8	-.317**	-.475**	-.373**
Number with 9	-.023	-.080	-.048
Number with ≥10	.505**	.666**	.563**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 5.41: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 2 (the second presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.182	1.000	
Mean	.728**	.641**	1.000
Number with 0	-.656**	-.301*	-.671**
Number with 1	-.498**	-.340**	-.678**
Number with 2	-.294*	-.471**	-.649**
Number with 3	-.230	-.289*	-.546**
Number with 4	-.252*	-.295*	-.501**
Number with 5	.049	-.115	-.267*
Number with 6	.306**	.061	-.134
Number with 7	.362**	.012	-.348**
Number with 8	.038	-.112	-.347**
Number with 9	.090	.790**	-.452**
Number with ≥10	.505**	.666**	.563**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 5.42: The correlations between the various statistics for the locations of boundaries within Stage 6, Part 1, Repetition 3 (the third presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.169	1.000	
Mean	.754**	.675**	1.000
Number with 0	-.642**	-.220	-.613**
Number with 1	-.455**	-.297*	-.668**
Number with 2	-.356**	-.407**	-.599**
Number with 3	-.377**	-.416**	-.639**
Number with 4	-.350**	-.079	-.416**
Number with 5	-.358*	-.197	-.437**
Number with 6	-.094	-.235	-.059
Number with 7	.348**	-.043	.294*
Number with 8	.615**	-.042	.555**
Number with 9	-.024	.167	.041
Number with ≥10	-.012	.793**	.389**

Notes: * Significant (one tailed) at $p<0.05$

** Significant (one tailed) at $p<0.01$

Table 5.43: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 1 (the fourth presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.674**	1.000	
Mean	.924**	.852**	1.000
Number with 0	-.380**	-.304*	-.428**
Number with 1	-.365**	-.342**	-.451**
Number with 2	-.204	-.275*	-.345**
Number with 3	-.235	-.218	-.362**
Number with 4	-.237	-.283*	-.366**
Number with 5	-.267*	-.258*	-.313*
Number with 6	-.208*	-.289*	-.190
Number with 7	-.204	-.260*	-.163
Number with 8	.154	-.241	.041
Number with 9	-.054	.096	.033
Number with ≥10	.315*	.748**	.517**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 5.44: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 2 (the fifth presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.274*	1.000	
Mean	.781**	.648**	1.000
Number with 0	-.594**	-.175	-.679**
Number with 1	-.502**	-.398**	-.721**
Number with 2	-.273*	-.541**	-.615**
Number with 3	-.265*	-.503*	-.591**
Number with 4	-.325*	-.451*	-.540**
Number with 5	-.323*	-.167	-.438**
Number with 6	.098	.149	.398**
Number with 7	.233	.221	.402**
Number with 8	.381**	.003	.413**
Number with 9	.104	.131	.115
Number with ≥10	.133	.725**	.361**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Table 5.45: The correlations between the various statistics for the locations of boundaries within Stage 8, Part 1, Repetition 3 (the sixth presentation of the melody).

Correlations	Minimum	Maximum	Mean
Maximum	.515**	1.000	
Mean	.870**	.798**	1.000
Number with 0	-.468**	-.263*	-.492**
Number with 1	-.433**	-.265**	-.488**
Number with 2	-.316*	-.425**	-.538**
Number with 3	-.302*	-.513*	-.510**
Number with 4	-.288*	-.411**	-.439**
Number with 5	-.240	-.144	-.290*
Number with 6	-.254*	.030	-.081
Number with 7	.070	-.126	.118
Number with 8	.338**	-.041	.246
Number with 9	.-.125	.385**	.093
Number with ≥10	.501**	.831**	.672**

Notes: * Significant (one tailed) at $p<0.05$
** Significant (one tailed) at $p<0.01$

Within-subjects ANOVAs using planned comparisons was performed on the minimum, maximum and mean distances between boundaries to see if there were any changes as a function of repetition (see Table 5.46). All variables changed as a function of repetition (in contrast to the previous melody), and contrast analysis indicated that these changes were distributed pretty much over the set of repetitions. Note that the η^2 and ω^2 values indicate that the changes are not insignificant, accounting for about 10% of the variance. The contrast analysis indicated that Repetitions 4 and 6 were the most similar, but previous analysis implied that Repetitions 5 and 6 were most similar (cf., Tables 5.34 through 5.39). The reason for the difference is likely due to the choice of contrasts and the focus on global measures as dependent variables in the ANOVA.

Table 5.46: The analysis of the changes in the boundary statistics across repetitions of the same melody, for Stages 6 and 8, Part 1.

WS-ANALYSIS		SS_{IV}	df	SS_{ERR}	df	*F*	*p*	η^2	ω^2
Minimum Distance Between Boundaries	Overall	313.174	5	4829.660	305	3.955	0.002	0.061	0.045
	Ψ_1	176.645	1	2032.355	61	5.302	0.025	0.080	
	Ψ_2	63.753	1	835.914	61	4.652	0.035	0.071	
	Ψ_3	4.645	1	853.355	61	.332	0.567	0.005	
	Ψ_4	24.774	1	377.227	61	4.006	0.050	0.062	
	Ψ_5	43.358	1	730.809	61	3.619	0.062	0.056	
Maximum Distance Between Boundaries	Overall	505.326	5	3353.498	305	9.192	0.000	0.131	0.116
	Ψ_1	193.750	1	1429.750	61	8.266	0.006	0.119	
	Ψ_2	115.185	1	503.981	61	13.942	0.000	0.186	
	Ψ_3	108.516	1	553.484	61	11.960	0.000	0.164	
	Ψ_4	37.430	1	321.237	61	7.108	0.010	0.104	
	Ψ_5	50.545	1	545.046	61	5.647	0.021	0.085	
Mean Distance Between Boundaries	Overall	361.323	5	3098.397	305	7.114	0.000	0.104	0.089
	Ψ_1	196.832	1	1427.138	61	8.413	0.005	0.121	
	Ψ_2	81.494	1	503.604	61	9.871	0.003	0.139	
	Ψ_3	22.536	1	498.880	61	2.756	0.102	0.043	
	Ψ_4	30.041	1	221.227	61	8.283	0.006	0.120	
	Ψ_5	30.420	1	447.548	61	4.146	0.046	0.064	

Notes Ψ_1 compared Repetition 1 to Repetition 3 (within Stage 6)
Ψ_2 compared Repetitions 1 and 3 to Repetition 2 (within Stage 6)
Ψ_3 compared Repetition 4 to Repetition 6 (within Stage 8)
Ψ_4 compared Repetitions 4 and 6 to Repetition 5 (within Stage 8)
Ψ_5 compared Repetitions 1, 2 and 3 to Repetitions 4, 5 and 6 (Stage 6 to Stage 8)
Maximum distance both violated the Mauchly Sphericity Test, but the multivariate results were the same. Minimum distance violated the Mauchly Sphericity Test, and the multivariate results were non-significant ($p<.183$). Mean distance violated the Mauchly Sphericity Test and the results were marginal ($p<.050$).

The associated correlation matrix (Table 5.47) implies that individuals were more consistent over the last two repetitions within a stage, and that the Repetition 1 (the first of Stage 6) was more similar to Repetition 4 (the first of Stage 8) than the other repetitions within Stage 6. The same is true in reverse: Repetition 4 was more similar to

Repetition 1 than it was to Repetitions 5 and 6.

Table 5.47: The correlation of the boundary statistics across repetitions of the same melody, for Stages 6 and 8, Part 1.

Correlation Within Measures			Repetition Number				
			2	3	4	5	6
Minimum Distance Between Boundaries	Repetition Number	1	0.158	0.140	0.361**	0.045	0.266*
		2	1.000	0.371**	0.182	0.540**	0.504**
		3		1.000	0.395**	0.534**	0.292*
		4			1.000	0.190	0.209
		5				1.000	0.599**
Maximum Distance Between Boundaries	Repetition Number	1	0.256*	0.164	0.518**	0.091	0.130
		2	1.000	0.516**	0.312*	0.696**	0.490**
		3		1.000	0.200	0.488**	0.441**
		4			1.000	0.242	0.326**
		5				1.000	0.525**
Mean Distance Between Boundaries	Repetition Number	1	0.197	0.181	0.482**	0.012	0.220
		2	1.000	0.740**	0.437**	0.712**	0.681**
		3		1.000	0.460**	0.660**	0.551**
		4			1.000	0.301**	0.332*
		5				1.000	0.698**

Notes: Repetitions 1, 2 and 3 are Stage 6 and Repetitions 4, 5 and 6 are Stage 8
* $p<0.05$
** $p<0.01$

Generally, the results indicate that subjects demonstrated high within-subject consistencies, as well as, high between-subject similarities, with some differential responding over time. The analysis of change implied that the last two repetitions of each stage were the most similar, or conversely, that most of the change occurred between Repetitions 1 and 2 (of Stage 6) and Repetitions 4 and 5 (of Stage 8), as well as, between stages.

Given that there exists an average boundary profile that can be said to represent the parsing of individual subjects, the question shifts to one of the prediction of those boundaries based upon the structure of music. In this analysis, the method of non-linear regression with asymmetric error terms (see Chapter 4 of this work) was applied to the average profile of Repetitions 2 and 3 and to the average profile of Repetitions 5 and 6. That is, for each subject, an individual profile for Stage 6 was created by averaging Repetitions 2 and 3 and an individual profile for Stage 8 was created by averaging Repetitions 5 and 6. These individual profiles were averaged within each stage over all subjects (the actual order of operations does not matter). This averaging seems valid given the high within-subject consistencies and the high between-subjects similarity.

For the extract from the melody "Softly Now the Light of Day", the analysis was very similar to that of the previous analysis (i.e., similar to "Three Blind Mice"), although there are three rules to consider and two different presentations (Stage 6 & Stage 8). In Figure 5.14 (see also Figure 5.13), the empirically determined boundaries are shown in the bottom panel. These can be compared with the best fitting theoretically determined predictions of Lerdahl and Jackendoff (1983) for each rule individually within Stages 6 and 8 (Figure 5.14, the top two panels). The analysis, using asymmetrical error terms is presented in Table 5.48.

As with "Three Blind Mice" previously, the results indicate that Rule 2b was the most important rule, even though the total number of times that it could be applied dropped to 5 from 11. (However, this melody contains 34 notes in comparison to "Three Blind Mice" which contains 48 notes). The total error of fit for Rule 2b was actually less than in "Three Blind Mice", implying that despite the lower number of applications, the rule still captured the important moments (however, the second melody is shorter than the first: the average errors are equivalent). Rule 3a could actually be applied at more locations (9), but still produced a much higher overall error than Rule 2b. Rule 3d produced the higher level of overall error on five potential locations.

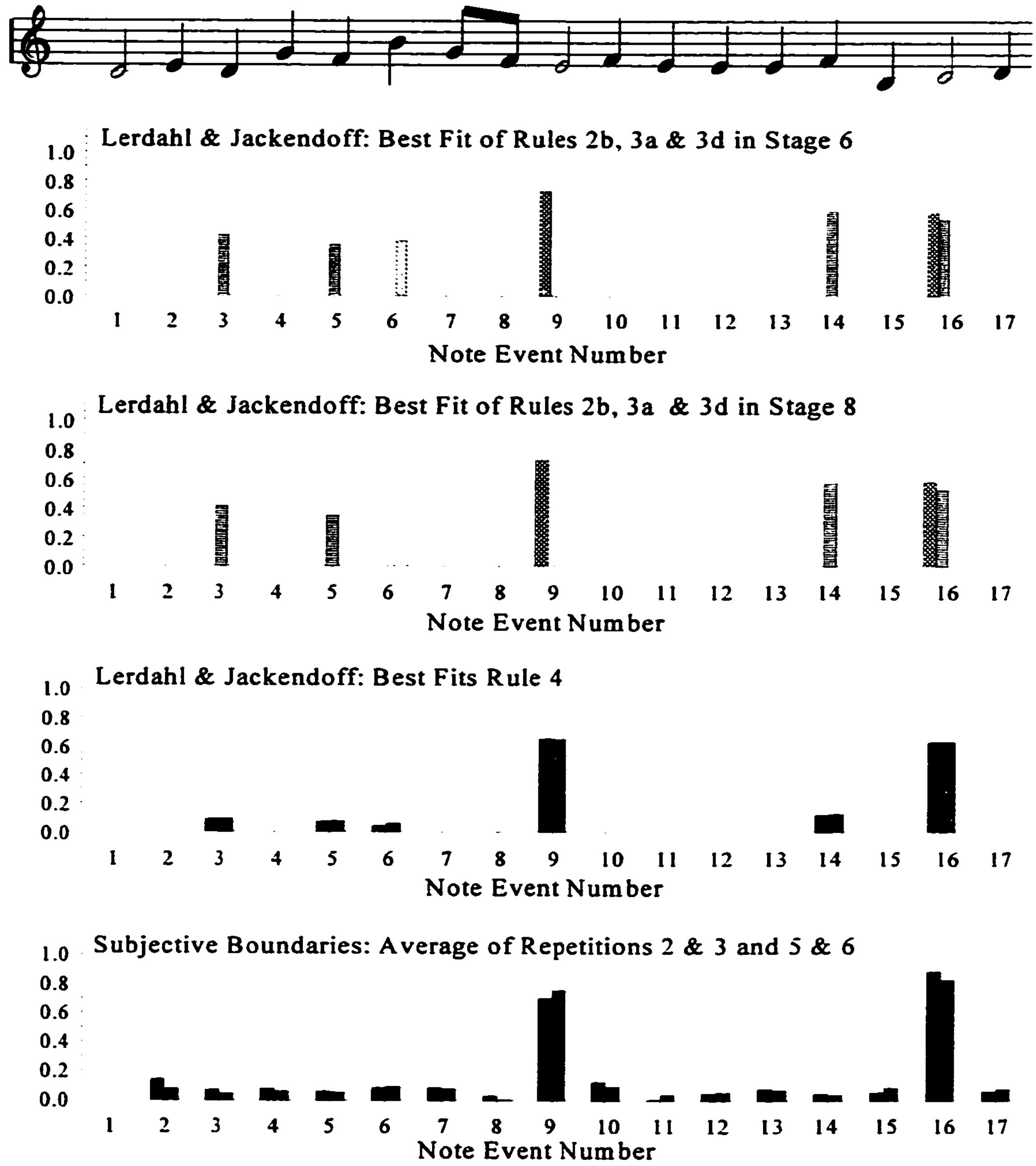


Figure 5.14a: The first part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a, 3d alone and in combination as Rule 4). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitionsfor all subjects.

▨ Rule 2b ■ Rule 4: Stage 6 ■ Average Profile (Repetitions 2 & 3)
▤ Rule 3a ■ Rule 4: Stage 8 ■ Average Profile (Repetitions 5 & 6)
▨ Rule 3d

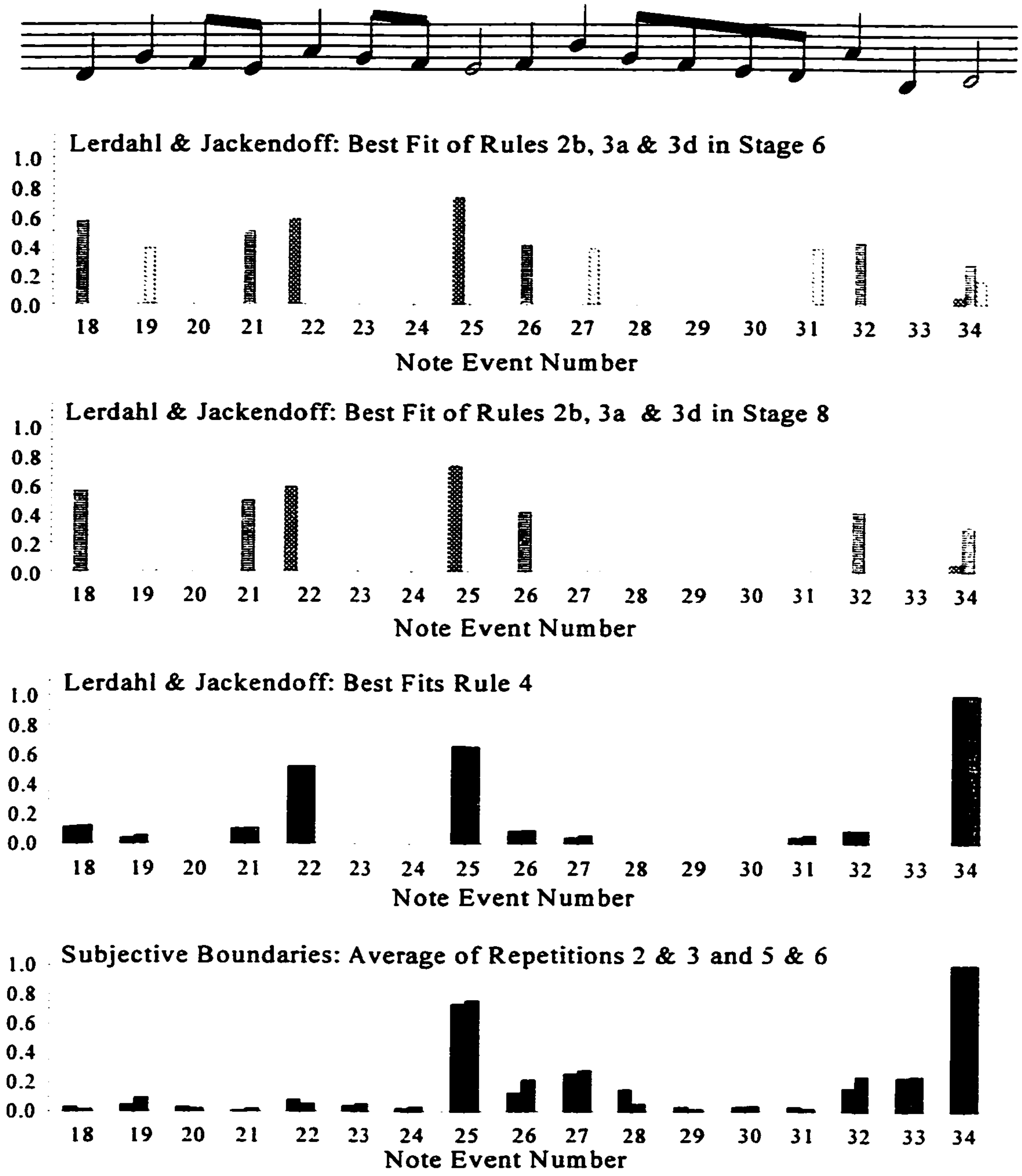


Figure 5.14b: The second part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (2b, 3a, 3d alone and in combination as Rule 4). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitionsfor all subjects.

▩ Rule 2b
▤ Rule 3a
▨ Rule 3d

■ Rule 4: Stage 6
■ Rule 4: Stage 8

■ Average Profile (Repetitions 2 & 3)
■ Average Profile (Repetitions 5 & 6)

Table 5.48: Rules 2 and 3 for the extract from "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6				
Rule	n	*r*	Equation	SS_{error}
2b	5	0.742**	b = 1.163*2b	7.418
3a	9	0.237	b = 0.708*3a	13.842
3d	5	0.284	b = 0.767*3d	16.335

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b	2	0.303	0.313	3	0.398	0.689
3a	2	0.704	0.648	7	1.703	2.753
3d	1	0.164	0.233	4	0.567	1.106

Notes: *r* = simple correlation (see Chapter 3 of this work)
n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Results for misses and false alarms separately provided much the same story. However, from this breakdown, it would seem that Rule 3d was preferable to Rule 3a: It produced a smaller error for misses and false alarms (bearing in mind that it occurs at fewer positions). It appears that the reason Rule 3d did not perform well is that it simply did not apply at the more important parts of the melody: For Rule 3d, the high value for the overall error, when compared to the miss and false alarm errors, implies that Rule 3d did not apply when the empirically determined boundaries were high (i.e., it did not apply in a lot of real situations); However, when it did apply, it could be aligned quite well with the empirical data. Rule 3a, on the other hand, produced a lot of false alarms, but even given the basis to ignore false alarms, it still produced a large error for misses. That is, on the occasions when Rule 3a did apply, it was not possible to match the value of the rule to the empirically determined boundary value with a precision that was even remotely close to that of Rule 2b or 3d.

For Stage 8 (Table 5.49, see also Figure 5.13 or 5.14), the results were essentially unchanged, producing the same slopes and errors as in Stage 6 (the gap between Rules 3a and 3d seemed to be narrowing if anything), as would be expected given the high similarities of the averaged boundaries in these two stages.

Table 5.49: Rules 2 and 3 for the extract from "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8				
Rule	n	r	Equation	SS_{error}
2b	5	0.738**	b = 1.159*2b	7.496
3a	9	0.241	b = 0.693*3a	14.314
3d	5	0.296	b = 0.773*3d	16.419

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b	3	0.200	0.319	2	0.416	0.682
3a	2	0.623	0.618	7	1.552	2.553
3d	1	0.15462	0.22732	4	0.525	1.047

Notes: r = simple correlation (see Chapter 3 of this work)
n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|\text{Deviation}| = \Sigma(X_i - X'_i)$

The extension of the analysis to Rule 4 (Intensification) was conducted in the same manner as for the previous melody (see also, Chapter 4 of this work). That is, the best fit combination for Rules 2b, 3a and 3d was determined using non-linear regression and asymmetrical error terms. Then each two-way interaction was added, followed by all two-way interactions. This analysis was conducted within both Stages 6 and 8.

In Stage 6 (Table 5.50; see also the third panel of Figure 5.14), overall error value, misses value and the false alarm value are quite low. In addition, the inclusion of interactions did not aid in prediction. In no case, did the interactions produce a significant contribution. Note that the combined action was a very good predictor of boundaries.

In Stage 8, the result was very similar (Table 5.49; see also Figure 5.14), with only minor fluctuations to the actual slopes. Generally, as in the previous melody, Rule 2b was the most important. For Stage 6, Rules 3a seemed more important than 3d. For Stage 8, more emphasis had been given to Rule 3d (it seemed to be equivalent to 3a).

Table 5.50: The Combination of Rules 2 and 3 (Rule 4) for the extract from "Softly Now the Light of Day", in Stage 6

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b = 1.045*2b+0.146*3a+0.084*3d	7.183
2b+3a+3d+2b*3a	16	b = 1.045*2b+0.146*3a+0.084*3d+0.000*2b*3a	7.183
2b+3a+3d+2b*3d	16	b = 1.045*2b+0.146*3a+0.084*3d+0.000*2b*	7.183
2b+3a+3d+3a*3d	16	b = 1.045*2b+0.146*3a+0.084*3d+0.000*3a*3d	7.183
all	16	b = 1.046*2b+0.146*3a+0.084*3d +0.000*2b*3a+0.000*2b*3d+0.000*3a*3d	7.183

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.386	0.769	8	0.468	1.004
2b+3a+3d+2b*3a	8	0.386	0.769	8	0.468	1.004
2a+3a+3d+2b*3d	8	0.368	0.769	8	0.468	1.004
2a+3a+3d+3a*3d	8	0.368	0.769	8	0.468	1.004
all	8	0.368	0.769	10	0.468	1.004

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 5.51: The Combination of Rules 2 and 3 (Rule 4) for the extract from "Softly Now the Light of Day", in Stage 8

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b = 1.044*2b+0.152*3a+0.115*3d	7.208
2b+3a+3d+2b*3a	16	b = 1.044*2b+0.152*3a+0.115*3d+0.000*2b*3a	7.208
2b+3a+3d+2b*3d	16	b = 1.044*2b+0.152*3a+0.115*3d+0.000*2b*	7.208
2b+3a+3d+3a*3d	16	b = 1.044*2b+0.152*3a+0.115*3d+0.000*3a*3d	7.208
all	16	b = 1.044*2b+0.152*3a+0.115*3d +0.000*2b*3a+0.000*2b*3d+0.000*3a*3d	7.208

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.437	0.988	8	0.535	1.148
2b+3a+3d+2b*3a	8	0.437	0.988	8	0.535	1.148
2a+3a+3d+2b*3d	8	0.437	0.988	8	0.535	1.148
2a+3a+3d+3a*3d	8	0.437	0.988	8	0.535	1.148
all	8	0.437	0.988	10	0.535	1.148

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X^{*}_i$
$|Deviation| = \Sigma(X_i - X^{*}_i)$

As with the previous melody ("Three Blind Mice" in Stage 4), the final analysis actually consisted of four separate ways to analyse parallelism: Parallelism due to pitch contour (Rule 6a, Pitch Pattern Parallelism A), analysed in a manner that include note durations, parallelism due to pitch contour (Rule 6b, Pitch Pattern Parallelism B), analysed in a manner that ignored note durations (i.e., note onset to note onset), parallelism due to time pattern (Rule 6c, Time Pattern Parallelism A), analysed in a manner that notated the timings of breaks between note events, and parallelism due to time pattern (Rule 6d, Time Pattern Parallelism B), analysed in a manner that compared the sequential note event durations.

In the analysis of Pitch Pattern Parallelism A (Rule 6a), Rule 6a was entered in conjunction with Rules 2 and 3, because Rule 6a is built on putative boundaries identified by Rules 2 and 3. The interactions of Rule 6a with Rules 2b, 3a and 3d were also computed and each interaction was considered separately as an addition to the model containing Rules 2b, 3a, 3d and 6a. Finally all rules and their two-way interactions were examined (not including 2-way interactions between Rules 2b, 3a and 3d). Table 5.52 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 5.53 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged), with Figure 5.15 providing the visual analogue. Note that Rule 4 (the best combination of Rules 2b, 3a and 3d) is provided in Figure 5.15 as a reference for the amount of improvement afforded by parallelism (Figure 5.15 also includes Rule 6b which will be discussed momentarily).

Note that the combination of all rules and interactions did provide some improvement overall (from 7.18 to 7.11 in Stage 6 & from 7.21 to 7.13 in Stage 8; 0.9% and 1.1%) and in the miss rate despite more data points being classified as misses (Reminder: A miss is any point where the prediction is less than the empirical data.). The improvement was mainly attributable to the parallelism at Note 16.

Table 5.52: The Combination of Rules 2, 3 (Rule 4) and 6a for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.045*2b+0.146*3a+0.084*3d	7.183
2b+3a+3d+6a	16	b=1.042*2b+0.110*3a+0.034*3d+0.090*6a	7.151
2b+3a+3d+6a+2b*6a	16	b=1.042*2b+0.110*3a+0.034*3d+0.090*6a + 0.000*2b*6a	7.151
2b+3a+3d+6a+3a*6a	16	b=1.014*2b+0.078*3a+0.082*3d+0.000*6a + 2.240*3a*6a	7.123
2b+3a+3d+6a+3d*6a	16	b=1.041*2b+0.113*3a+0.008*3d+0.083*6a + 0.693*3d*6a	7.150
all	16	b=1.012*2b+0.078*3a+0.000*3d+0.000*6a +0.000*2b*6a+2.235*3b*6a+1.905*3d*6a	7.114

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.386	0.769	8	0.468	1.004
2b+3a+3d+6a	8	0.317	0.711	8	0.504	1.025
2a+3a+3d+6a+2b*6a	8	0.317	0.711	8	0.504	1.025
2a+3a+3d+6a+3a*6a	8	0.348	0.756	8	0.445	0.930
2a+3a+3d+6a+3d*6a	9	0.318	0.722	7	0.502	1.024
all	8	0.331	0.774	8	0.453	0.937

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 5.53: The Combination of Rules 2, 3 (Rule 4) and 6a for "Softly Now the Light of Day" (Stage 8)

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.044+2b+0.152*3a+0.115*3d	7.208
2b+3a+3d+6a	16	b=1.041*2b+0.104*3a+0.046*3d+0.118*6a	7.153
2b+3a+3d+6a+2b*6a	16	b=1.041*2b+0.104*3a+0.046*3d+0.118*6a + 0.000*2b*6a	7.153
2b+3a+3d+6a+3a*6a	16	b=1.022*2b+0.083*3a+0.086*3d+0.043*6a + 1.664*3a*6a	7.141
2b+3a+3d+6a+3d*6a	16	b=1.040*2b+0.107*3a+0.008*3d+0.107*6a + 1.982*3d*6a	7.151
all	16	b=1.015*2b+0.080*3a+0.000*3d+0.000*6a +0.000*2b*6a+2.281*3b*6a+5.055*3d*6a	7.125

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.437	0.988	8	0.535	1.148
2b+3a+3d+6a	8	0.336	0.890	8	0.580	1.142
2a+3a+3d+6a+2b*6a	8	0.336	0.890	8	0.580	1.142
2a+3a+3d+6a+3a*6a	8	0.370	0.919	8	0.534	1.065
2a+3a+3d+6a+3d*6a	8	0.339	0.902	8	0.576	1.136
all	9	0.367	0.944	7	0.522	1.030

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

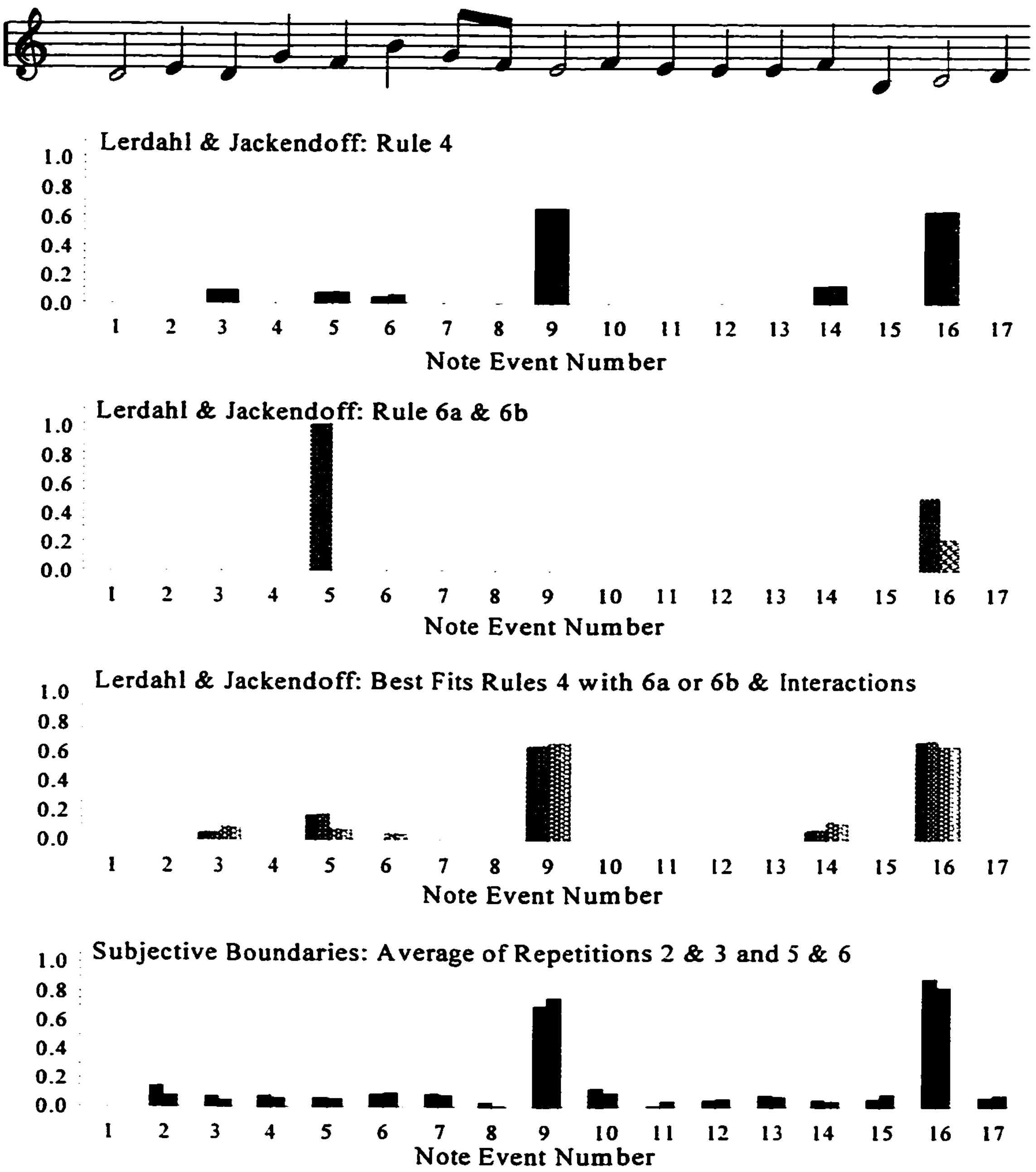


Figure 5.15a: The first part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (Rule 4 alone & in combination with 6a or 6b). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects.

■ Rule 4: Stage 6	■ Rules 4 & 6a	■ Best Fit for Rules 4, 6a & their interactions
■ Rule 4: Stage 8	▨ Rules 4 & 6b	▨ Best Fit for Rules 4, 6b & their interactions
■ Average Profile (Repetitions 2 & 3)		■ Average Profile (Repetitions 5 & 6)

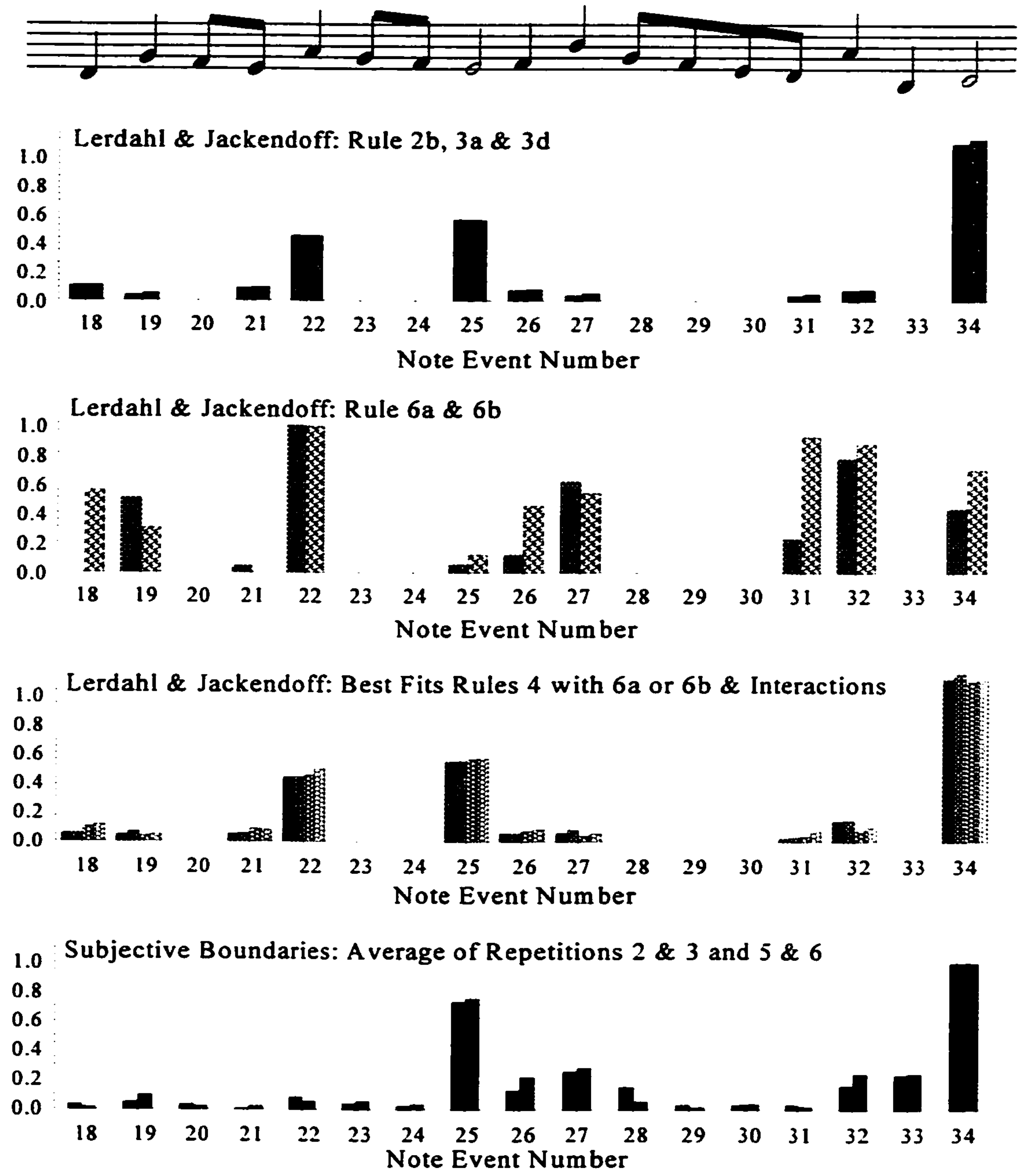


Figure 5.15b: The second part of the melody used in Stags 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (Rule 4 alone & in combination with 6a or 6b). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects.

■ Rule 4: Stage 6 ■ Rules 4 & 6a ■ Best Fit for Rules 4, 6a & their interactions
■ Rule 4: Stage 8 ▨ Rules 4 & 6b ▦ Best Fit for Rules 4, 6b & their interactions
■ Average Profile (Repetitions 2 & 3) ■ Average Profile (Repetitions 5 & 6)

As before, the simple correlations between the rules are shown in Table 5.54 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown; these are redundant with previous tables). Note that in contrast to the previous melody, "Three Blind Mice", the correlations between Rule 6a (Pitch Pattern Parallelism A) and Rules 2b (Attack-Point), 3a (Register) and 3d (Length) were not high because Rule 6a was constrained on the basis of any single application of any one of three rules (Rules 2b, 3a or 3d) instead of just two rules (Rules 2b & 3a in "Three Blind Mice").

Table 5.54: Correlation Matrix for "Softly Now the Light of Day", Pitch Pattern Parallelism A.

	Rule							
	2b	3a	3d	4	6a	2b*6a	3a*6a	3d*6a
Stage 6	.742**	.237	.284	.735**	.128	.427*	.369*	.328
Stage 8	.738**	.241	.296	.736**	.296	.400*	.377*	.343*
Rule 2b		.292	.386*	.988**	.296	.712**	.377*	.410*
Rule 3a			.178	.412*	.278	.312	.610**	.212
Rule 3d				.467**	.285	.367*	.243	.898**
Rule 4	.992**	.403*	.428**	1.000	.333	.762**	.445*	.486*
Rule 6a				.325		.538**	.671**	.369**
R 2b*6a				.761**			.382*	.386*
R 3a*6a				.439**				.266
R 3d*6a				.451**				

Notes: * $p < .05$

** $p < .01$

Since Rule 4 is a combination of Rules 2 and 3, it has different correlations within Stages 6 and 8: Stage 6 are shown above the diagonal, and Stage 8 are shown below the diagonal.

Generally, Pitch Pattern Parallelism A (Rule 6a) added little in the way of predictability to the model beyond that which could be explained by the low level rules. The same pattern was evident in both stages. What predictability was evident was due to both the rule and its interactions, particularly with Rules 3a and 3d (not Rule 2b), although in the final equation Rule 6a alone was not a factor (i.e., Rule 6a was dropped while the interactions were retained).

The analysis of Pitch Pattern Parallelism B (Rule 6b) proceeded in a similar manner. The rule was entered in conjunction with Rules 2 and 3 and then the interactions of Rule 6a with Rules 2b, 3a and 3d were considered separately. Finally all rules and their two-way interactions (not including 2-way interactions between Rules 2b, 3a and 3d) were examined. Table 5.53 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 5.54 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged). The visual representation is shown in Figure 5.16. The rule and all its interactions were not useful at all in either stage. The change in fit in Stage 6 did not reduce the error (from 7.18 to 7.18; 0.0%), and only slightly improved the fit in Stage 8 (from 7.21 to 7.20; 0.1%).

Table 5.55: The Combination of Rules 2, 3 (Rule 4) and 6b for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.045*2b+0.146*3a+0.084*3d	7.183
2b+3a+3d+6b	16	b=1.046*2b+0.146*3a+0.084*3d+0.000*6b	7.183
2b+3a+3d+6b+2b*6b	16	b=1.046*2b+0.146*3a+0.084*3d+0.000*6b + 0.000*2b*6b	7.183
2b+3a+3d+6b+3a*6b	16	b=1.046*2b+0.146*3a+0.084*3d+0.000*6b + 0.000*3a*6b	7.183
2b+3a+3d+6b+3d*6b	16	b=1.046*2b+0.146*3a+0.084*3d+0.000*6b + 0.000*3d*6b	7.183
all	16	b=1.046*2b+0.146*3a+0.084*3d+0.000*6b +0.000*2b*6b+0.000*3b*6b+0.000*3d*6b	7.183

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.386	0.769	8	0.468	1.004
2b+3a+3d+6b	8	0.386	0.769	8	0.468	1.004
2a+3a+3d+6b+2b*6b	8	0.386	0.769	8	0.468	1.004
2a+3a+3d+6b+3a*6b	8	0.386	0.769	8	0.468	1.004
2a+3a+3d+6b+3d*6b	8	0.386	0.769	8	0.468	1.004
all	8	0.386	0.769	8	0.468	1.004

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 5.56: The Combination of Rules 2, 3 (Rule 4) and 6b for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.044+2b+0.152*3a+0.115*3d	7.208
2b+3a+3d+6b	16	b=1.046*2b+0.131*3a+0.077*3d+0.052*6b	7.199
2b+3a+3d+6b+2b*6b	16	b=1.046*2b+0.131*3a+0.077*3d+0.052*6b + 0.000*2b*6b	7.199
2b+3a+3d+6b+3a*6b	16	b=1.046*2b+0.131*3a+0.077*3d+0.052*6b + 0.000*3a*6b	7.199
2b+3a+3d+6b+3d*6b	16	b=1.046*2b+0.131*3a+0.077*3d+0.051*6b + 0.000*3d*6b	7.199
all	16	b=1.046*2b+0.131*3a+0.077*3d+0.051*6b +0.000*2b*6b+0.000*3b*6b+0.000*3d*6b	7.199

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.437	0.988	8	0.535	1.148
2b+3a+3d+6b	8	0.397	0.955	8	0.565	1.160
2a+3a+3d+6b+2b*6b	8	0.397	0.955	8	0.565	1.160
2a+3a+3d+6b+3a*6b	8	0.397	0.955	8	0.566	1.160
2a+3a+3d+6b+3d*6b	8	0.397	0.955	8	0.565	1.160
all	8	0.397	0.955	8	0.565	1.160

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X^*_i$
$|\text{Deviation}| = \Sigma(X_i - X^*_i)$

The simple correlations between the rules are shown in Table 5.57 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown). Note the low correlations with the empirical data as well as the relatively higher correlations with other rules.

Table 5.57: Correlation Matrix for "Softly Now the Light of Day", Pitch Pattern Parallelism B.

	Rule							
	2b	3a	3d	4	6b	2b*6b	3a*6b	3d*6b
Stage 6	.742**	.237	.284	.735**	.120	.444**	.350*	.339
Stage 8	.738**	.241	.296	.736**	.145	.428*	.376*	.341*
Rule 2b		.292	.386*	.988**	.338	.783**	.428*	.459**
Rule 3a			.178	.412*	.314	.335	.695**	.254
Rule 3d				.467**	.465**	.526**	.404*	.893**
Rule 4	.992**	.403*	.428*	1.000	.392*	.807**	.518**	.535**
Rule 6b				.377*		.533**	.591**	.561**
R 2b*6b				.799**			.524**	.599**
R 3a*6b				.504**				.477**
R 3d*6b				.502**				

Notes: * $p < .05$
** $p < .01$
Since Rule 4 is a combination of Rules 2 and 3, it has different correlations within Stages 6 and 8: Stage 6 are shown above the diagonal, and Stage 8 are shown below the diagonal.

In the analysis of Time Pattern Parallelism A (Rule 6c), Rule 6c was entered in conjunction with Rules 2 and 3, followed by the interactions of Rule 6a with Rules 2 and 3 3. Each interaction was entered separately, and then together (not including 2-way interactions between Rules 2b, 3a and 3d). Table 5.58 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 5.59 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged), with Figure 5.16 providing a graphical representation (Figure 5.16 includes Rules 6d). Note that the combination of all rules and interactions did provide some improvement overall (from 7.18 to 7.04 in Stage 6 & from 7.21 to 6.99 in Stage 8; 1.9% & 3.0% respectively). The improvement was mainly attributable to Rule 6c and its interaction with 2b.

Table 5.58: The Combination of Rules 2, 3 (Rule 4) and 6c for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.045*2b+0.146*3a+0.084*3d	7.183
2b+3a+3d+6c	16	b=0.963*2b+0.062*3a+0.000*3d+0.131*6c	7.067
2b+3a+3d+6c+2b*6c	16	b=0.390*2b+0.037*3a+0.000*3d+0.137*6c +0.697*2b*6c	7.045
2b+3a+3d+6c+3a*6c	16	b=0.961*2b+0.053*3a+0.000*3d+0.130*6c +0.140*3a*6c	7.067
2b+3a+3d+6c+3d*6c	16	b=0.963*2b+0.062*3a+0.000*3d+0.131*6c +0.000*3d*6c	7.067
all	16	b=0.390*2b+0.037*3a+0.000*3d+0.137*6c +0.697*2b*6c+0.000*3b*6c+0.000*3d*6c	7.044

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.386	0.769	8	0.468	1.004
2b+3a+3d+6c	9	0.261	0.550	7	0.476	1.081
2a+3a+3d+6c+2b*6c	8	0.233	0.613	8	0.482	1.076
2a+3a+3d+6c+3a*6c	9	0.261	0.563	7	0.476	1.081
2a+3a+3d+6c+3d*6c	9	0.261	0.550	7	0.476	1.081
all	8	0.233	0.613	8	0.482	1.076

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X^*_i$
$|Deviation| = \Sigma(X_i - X'_i)$

Table 5.59: The Combination of Rules 2, 3 (Rule 4) and 6c for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.044+2b+0.152*3a+0.115*3d	7.208
2b+3a+3d+6c	16	b=0.931*2b+0.035*3a+0.000*3d+0.178*6c	6.994
2b+3a+3d+6c+2b*6c	16	b=0.878*2b+0.032*3a+0.000*3d+0.179*6c + 0.059*2b*6c	6.994
2b+3a+3d+6c+3a*6c	16	b=0.931*2b+0.035*3a+0.000*3d+0.178*6c + 0.000*3a*6c	6.994
2b+3a+3d+6c+3d*6c	16	b=0.931*2b+0.035*3a+0.000*3d+0.178*6c + 0.000*3d*6c	6.994
all	16	b=0.878*2b+0.032*3a+0.000*3d+0.179*6c +0.059*2b*6c+0.000*3b*6c+0.000*3d*6c	6.994

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.437	0.988	8	0.535	1.148
2b+3a+3d+6c	8	0.211	0.665	8	0.548	1.197
2a+3a+3d+6c+2b*6c	8	0.210	0.670	8	0.548	1.197
2a+3a+3d+6c+3a*6c	8	0.211	0.665	8	0.548	1.197
2a+3a+3d+6c+3d*6c	8	0.211	0.665	8	0.548	1.197
all	8	0.210	0.670	8	0.548	1.197

Notes: n = number of point that the rule applied (either as a miss or false alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

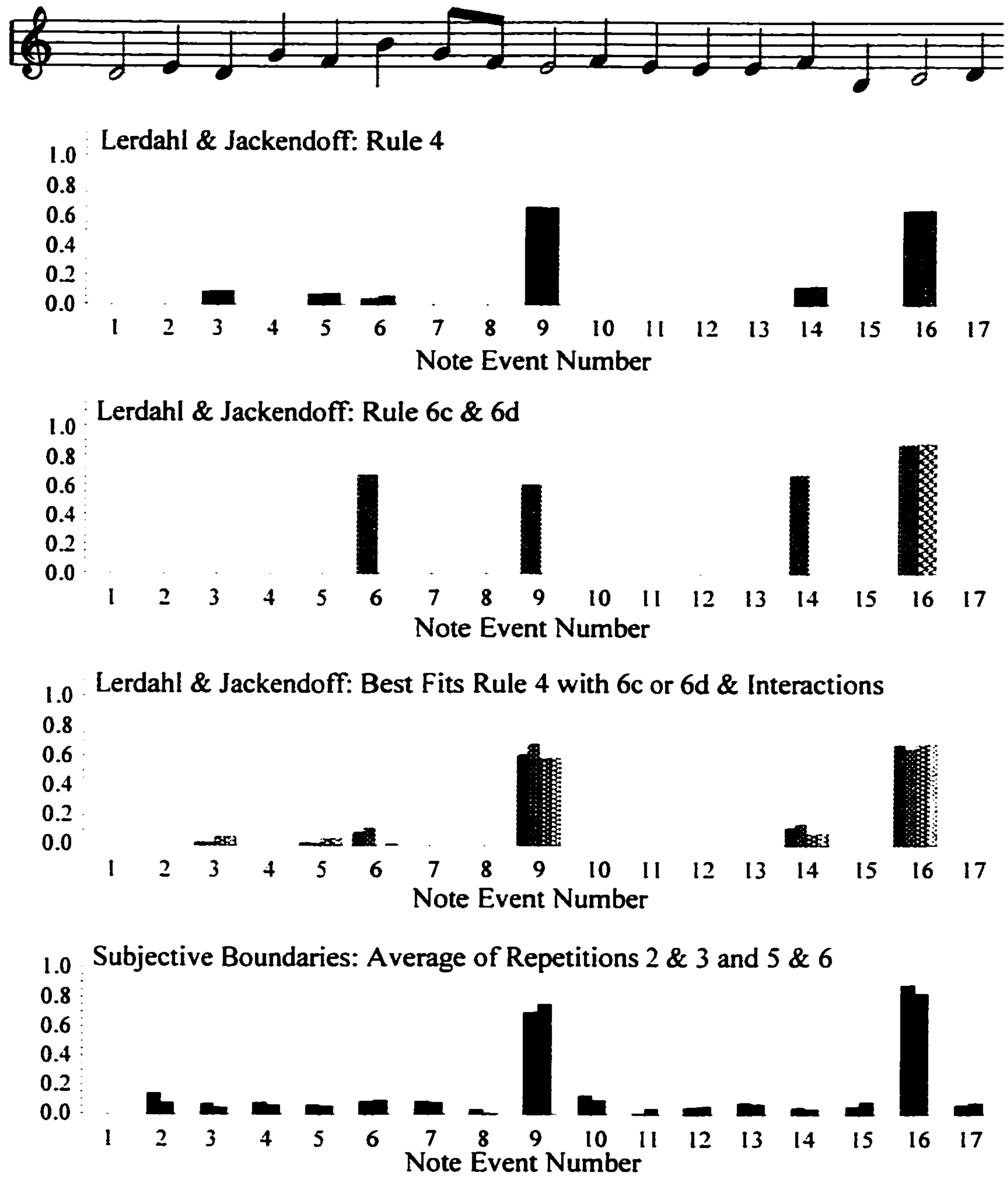


Figure 5.16a: The first part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (Rule 4 alone & in combination with 6c or 6d). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects.

■ Rule 4: Stage 6 ■ Rules 4 & 6c ■ Best Fit for Rules 4, 6c & their interactions
■ Rule 4: Stage 8 ▨ Rules 4 & 6d ▨ Best Fit for Rules 4, 6d & their interactions
■ Average Profile (Repetitions 2 & 3) ■ Average Profile (Repetitions 5 & 6)

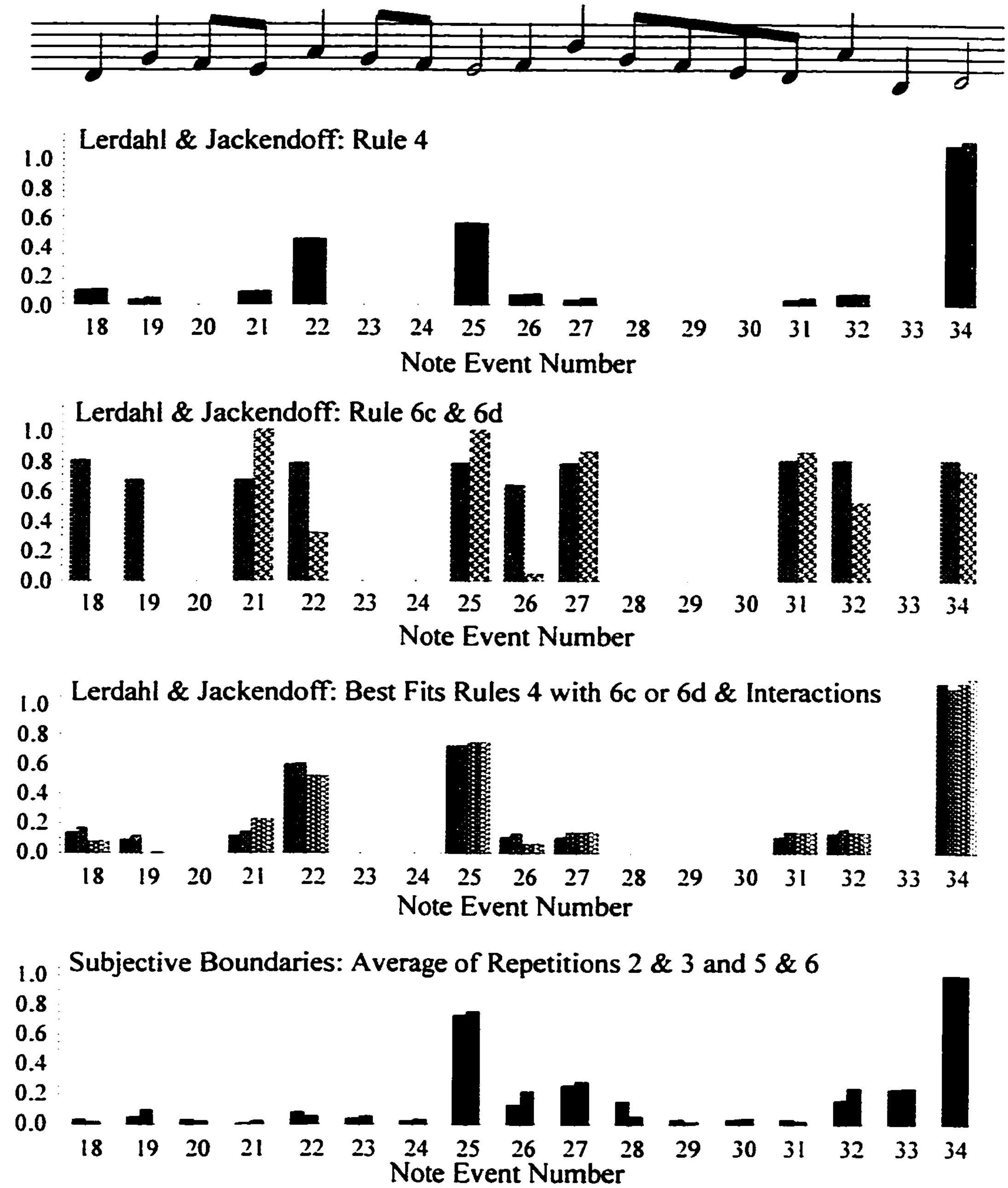


Figure 5.16b: The second part of the melody used in Stages 6 & 8, Part 1, and the predicted boundary locations given Lerdahl and Jackendoff's (1983) Group Preference Rules (Rule 4 alone & in combination with 6c or 6d). This can be compared with the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects.

■ **Rule 4: Stage 6** ■ **Rules 4 & 6c** ■ **Best Fit for Rules 4, 6c & their interactions**
■ **Rule 4: Stage 8** ▨ **Rules 4 & 6d** ▦ **Best Fit for Rules 4, 6d & their interactions**
■ **Average Profile (Repetitions 2 & 3)** ■ **Average Profile (Repetitions 5 & 6)**

As before, the simple correlations between the rules are shown in Table 5.60 (correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown).

Table 5.60: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism A.

	Rule							
	2b	3a	3d	4	6c	2b*6c	3a*6c	3d*6c
Stage 6	.742**	.237	.284	.735**	.339	.743**	.338	.306
Stage 8	.738**	.241	.296	.736**	.372*	.735*	.345*	.316
Rule 2b		.292	.386*	.988**	.501**	.994**	.386*	.411*
Rule 3a			.178	.412*	.510**	.326	.910**	.199
Rule 3d				.467**	.479**	.406*	.247	.997**
Rule 4	.992**	.403**	.428**	1.000	.568**	.989**	.493**	.493**
Rule 6c				.554**		.507**	.618**	.474**
R 2b*6c				.991**			.424*	.433*
R 3a*6c				484**				.269
R 3d*6c				455**				

Notes: * $p < .05$
** $p < .01$
Since Rule 4 is a combination of Rules 2 and 3, it has different correlations within Stages 6 and 8: Stage 6 are shown above the diagonal, and Stage 8 are shown below the diagonal.

Generally, Time Pattern Parallelism A (Rule 6c) did add some predictability to the model beyond that which could be explained by the low level rules. The same pattern was evident in both stages. This predictability was due to both the rule and its interaction with Rule 2b.

In the final analysis, Time Pattern Parallelism B (Rule 6d), Rule 6d was entered, as before, in conjunction with Rules 2 and 3 followed by the interactions of Rule 6d with Rules 2b, 3a and 3d. In the last analysis, all rules and their two-way interactions (not including 2-way interactions between Rules 2b, 3a and 3d) were examined. Table 5.61 provides the details for the first session (Stage 6: Repetition 2 and 3 averaged) and Table 5.62 provides the details for the second session (Stage 8: Repetitions 5 and 6 averaged) along with Figure 5.16. Note that the combination of all rules and interactions did provide some improvement overall (from 7.18 to 7.02 in Stage 6 & from 7.21 to 7.05 in Stage 8; 2.3% and 2.2% respectively). The improvement was mainly attributable to Rule 6d alone, though the interaction with Rule 3a did have a minor contribution in Stage 6.

Table 5.61: The Combination of Rules 2, 3 (Rule 4) and 6d for "Softly Now the Light of Day" (Stage 6).

Overall Results: Stage 6			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.045*2b+0.146*3a+0.084*3d	7.183
2b+3a+3d+6d	16	b=0.938*2b+0.097*3a+0.000*3d+0.159*6d	7.018
2b+3a+3d+6d+2b*6d	16	b=0.938*2b+0.097*3a+0.000*3d+0.159*6d +0.000*2b*6d	7.018
2b+3a+3d+6d+3a*6d	16	b=0.937*2b+0.096*3a+0.000*3d+0.158*6d +0.030*3a*6d	7.018
2b+3a+3d+6d+3d*6d	16	b=0.938*2b+0.097*3a+0.000*3d+0.159*6d +0.000*3d*6d	7.018
all	16	b=0.937*2b+0.096*3a+0.000*3d+0.158*6d +0.000*2b*6d+0.030*3b*6d+0.000*3d*6d	7.018

Breakdown for Misses and False Alarms: Stage 6						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.386	0.769	8	0.468	1.004
2b+3a+3d+6d	9	0.240	0.697	7	0.448	0.999
2a+3a+3d+6d+2b*6d	9	0.240	0.697	7	0.448	0.999
2a+3a+3d+6d+3a*6d	9	0.240	0.699	7	0.448	0.997
2a+3a+3d+6d+3d*6d	9	0.240	0.697	7	0.448	0.999
all	9	0.240	0.699	8	0.448	0.997

Notes: n = number of point that the rule applied (either as a miss or falses alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X^{*}_i$
$|Deviation| = \Sigma(X_i - X^{*}_i)$

Table 5.62: The Combination of Rules 2, 3 (Rule 4) and 6d for "Softly Now the Light of Day" (Stage 8).

Overall Results: Stage 8			
Rule	n	Equation	SS_{error}
2b+3a+3d	16	b=1.044+2b+0.152*3a+0.115*3d	7.208
2b+3a+3d+6d	16	b=0.943*2b+0.101*3a+0.022*3d+0.155*6d	7.051
2b+3a+3d+6d+2b*6d	16	b=0.943*2b+0.101*3a+0.022*3d+0.155*6d +0.000*2b*6d	7.051
2b+3a+3d+6d+3a*6d	16	b=0.943*2b+0.101*3a+0.022*3d+0.155*6d +0.000*3a*6d	7.051
2b+3a+3d+6d+3d*6d	16	b=0.943*2b+0.101*3a+0.022*3d+0.155*6d +0.000*3d*6d	7.051
all	16	b=0.943*2b+0.101*3a+0.022*3d+0.155*6d +0.000*2b*6d+0.000*3b*6d+0.000*3d*6d	7.051

Breakdown for Misses and False Alarms: Stage 8						
Rule	Misses			False Alarms		
	n	SS_{error}	|Deviation|	n	SS_{error}	|Deviation|
2b+3a+3d	8	0.437	0.988	8	0.535	1.148
2b+3a+3d+6d	9	0.322	0.896	7	0.493	1.087
2a+3a+3d+6d+2b*6d	9	0.322	0.896	7	0.493	1.087
2a+3a+3d+6d+3a*6d	9	0.322	0.896	7	0.493	1.087
2a+3a+3d+6d+3d*6d	9	0.322	0.896	7	0.493	1.087
all	9	0.322	0.896	7	0.493	1.087

Notes: n = number of point that the rule applied (either as a miss or falses alarm)
$SS_{error} = \Sigma[(\Delta X_i + 0.630)^4 - (\Delta X_i + 0.630)^1 + 0.472]$ where: $\Delta X_i = X_i - X'_i$
$|Deviation| = \Sigma(X_i - X'_i)$

As before, the simple correlations between the rules are shown in Table 5.63

(correlations between these rules and the interactions between Rules 2b, 3a and 3d are not shown).

Table5.63: Correlation Matrix for "Softly Now the Light of Day", Time Pattern Parallelism B.

	Rule							
	2b	3a	3d	4	6d	2b*6d	3a*6d	3d*6d
Stage 6	.742**	.237	.284	.735**	.420*	.711**	.458**	.359*
Stage 8	.738**	.241	.296	.736**	.426*	.696**	.453*	.359*
Rule 2b		.292	.386*	.988**	.296	.866**	.503**	.455*
Rule 3a			.178	.412*	.278	.367*	.665**	.256
Rule 3d				.467**	.285	.419*	.335	.851**
Rule 4	.992**	.403*	.428*	1.000	.333	.878**	.577**	.527**
Rule 6d				.514**		.610**	.652**	.537**
R 2b*6d				.876**			.576**	.483**
R 3a*6d				.568**				.396*
R 3d*6d				.497**				

Notes: * $p < .05$
** $p < .01$
Since Rule 4 is a combination of Rules 2 and 3, it has different correlations within Stages 6 and 8: Stage 6 are shown above the diagonal, and Stage 8 are shown below the diagonal.

Generally, Time Pattern Parallelism B (Rule 6d) did add some predictability to the model beyond that which could be explained by the low level rules. Essentially, the same pattern was evident in both stages. This predictability was due to the rule itself.

Boundary Efficacy: Stages 6 and 8, Part 2

As in Stage 4, a secondary click detection task was used to assess whether or not boundaries affect subsequent processing. Subjects listened to the melody, and simultaneously listened for clicks and then indicated the detection of a click as rapidly (and as accurately) as possible with a keypress. The time of detection was measured from the onset of the click, in milliseconds.

As in Stage 4, before the main analysis, subjects were equated by converting reaction times (RTs) to z-scores within each subject: All subjects were converted so that all had the same mean and standard deviation for RTs (0.00±1.00), so that the analysis could demonstrate which points in the melody produced longer RTs without regard to individual differences in RTs. Invalid trials were not included (14 in Stage 6 and 11 in Stage 8).

The predicted relationships would include faster RTs just after a boundary and slower RTs just before a boundary. From Figure 5.17, one can see this is not the general pattern. To paraphrase Calvin (Watterson, 1995, pp. 87-91,191), the data do not offer stupendous support for expectations. Visually, the data seem to be more indicative of a general trend to longer RTs at points where the melody is more complex.

Nonetheless, the same analysis as in Stage 4 was conducted. It must be remembered that the analysis actually compared boundary positions of the third repetition of each stage (not the average of the second and third repetitions) to the positions of clicks because this was how the actually experiment was conducted (i.e., clicks were placed within the melody on the basis of each subject's individual responding in the third repetition of each stage; see the previous Methods section).

In the initial assessment of reaction times (RT) to clicks, much of the information contained within the data was ignored so that RTs for clicks could be examined in the region of boundaries. To remind, in the actual analysis, RTs were examined for the following conditions:

clicks that fell on a note that was two notes in front of a boundary and at least three notes after the previous boundary

clicks that fell on a note that was one note in front of a boundary and at least three notes after the previous boundary

clicks that fell on a note that was a boundary

clicks that fell on a note that was one note after a boundary and at least three notes before the next boundary

clicks that fell on a note that was two notes after a boundary and at least three notes before the next boundary.

When this was done, in Stage 6, the number of clicks was reduced from a total of 482 (62 times 8 equals 496, minus 14 invalid trials) to 355, with 203 of those being on boundaries (hence, 127 clicks were placed between boundaries, but more than 2 notes from any boundary). In Stage 8, the number of clicks was similarly reduced from a total of 485 (62 times 8 minus 11 invalid trials) to 372, with 204 of those being on boundaries (hence, 120 clicks were placed between boundaries, but more than two notes from any boundary).

For Stage 6, the analysis (see Table 5.64) indicated that there was a significant tendency (see the last contrast, in particular) to be faster both before and after a boundary. There are also other differences. RTs to clicks on notes that were two notes before the boundary were much slower than RTs to clicks on notes that were one note before the boundary (which is the fastest condition). In Stage 8, the results (see Table 5.65) were similar in some ways but there were no significant differences. As before the fastest condition was that of click placed one note in front of the boundary. However, in this case, the slowest condition was that of click placed on a note that was two notes after a boundary.

Table 5.64: RT (z-scores) for clicks placed just before, on, or just after boundaries and the analysis of those scores, in Stage 6.

Boundary Position	n	v	Mean ± SD	Median	Mode	Min	Max
Two Before	67	1	-0.100 ± 0.907	-0.426	-1.984	-1.984	2.458
Before	11	1	-0.351 ± 1.108	-0.459	-1.416	-1.416	2.012
On	197	6	0.143 ± 0.948	-0.136	0.000	-2.035	2.459
Just After	23	0	0.075 ± 0.898	-0.230	0.000	-0.972	2.202
Two After	57	3	-0.219 ± 0.766	-0.415	-0.447	-1.504	2.257
Overall	355	11	0.021 ± 0.924	-0.230	0.000	-2.035	2.459

WS-ANOVA	SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Phi Overall	8.90	4	293.30	350	2.65	0.033	0.029	0.018
Ψ_1	0.23	1			0.27	0.602	0.000	
Ψ_2	3.04	1			3.62	0.058	0.010	
Ψ_3	0.06	1			0.07	0.797	0.000	
Ψ_4	5.58	1			6.66	0.010	0.018	

where: Ψ_1 compares the two Before conditions
Ψ_2 compares the two After conditions
Ψ_3 compares Before to After.
Ψ_4 compares Before and After to On.

Table 5.65: RT (z-scores) for clicks placed just before, on, or just after boundaries and the analysis of those scores, in Stage 8.

Boundary Position	n	v	Mean ± SD	Median	Mode	Min	Max
Two Before	69	1	-0.063 ± 1.028	-0.062	-0.915	-2.150	2.471
Before	5	0	-0.167 ± 1.425	-0.254	-2.017	-2.017	1.984
On	204	1	-0.012 ± 0.868	-0.290	0.000	-1.774	2.415
Just After	23	0	-0.027 ± 1.024	-0.392	-1.641	-1.641	2.228
Two After	71	2	0.176 ± 1.046	-0.117	-0.489	-1.931	2.472
Overall	372	4	0.035 ± 0.950	-0.223	0.000	-2.150	2.472

WS-ANOVA	SS_{IV}	df	SS_{ERR}	df	F	p	η^2	ω^2
Phi Overall	2.21	4	332.86	367	0.61	0.656	0.007	0.000
Ψ_1	0.14	1			0.15	0.695	0.000	
Ψ_2	1.61	1			1.78	0.183	0.005	
Ψ_3	0.43	1			0.48	0.491	0.001	
Ψ_4	0.03	1			0.03	0.867	0.000	

where: Ψ_1 compares the two Before conditions
Ψ_2 compares the two After conditions
Ψ_3 compares Before to After.
Ψ_4 compares Before and After to On.

It is difficult to relate these to any general rationale: In particular, clicks before boundaries were not slower than clicks after boundaries. The hypothesis held that RTs should increase as processing demands of the melody increased. Boundary points were associated with increased processing demands because a boundary point implies the closure of a unit. Generally, click on boundaries are slower than clicks off boundaries. This offers some support for the hypothesis. However, more generally, the results are difficult to interpret. That is, in Stage 6, unit closure seems to be able to explain the pattern: Clicks on the boundary or one note after the boundary were the slowest conditions. One could suppose that the act of boundary closure takes some time (particularly in a complex melody). However, in Stage 8, the condition where clicks were on notes that were two notes after the boundary was the slowest condition. It is this finding that is difficult to interpret.

A correlational analysis like that of Stage 4 was used in each of Stages 6 and 8 (with all the data). As before, in this analysis, variables coding for position of click within the melody (to the third order), boundary type (on or off a boundary), distance from the previous boundary (0 if on a boundary) and distance to the next boundary (0 if on a boundary) were used. Variables were entered hierarchically, with melody position first, boundary type second and the distances third. As can be gleaned from Figure 5.17, results did not indicate a particularly strong relationship (see Table 5.66) in either stage: The important relationship is that of click distance from boundaries after position in melody and type of boundary have been controlled for. Note that the type of click (on a boundary or not) did add predictability to the equation even after position within the melody had been controlled for in Stage 6[57], but the magnitude was not high at only 2% (generally, the same was observed in Stage 4). This was not observed in Stage 8. In neither stage did the combination of previous and post distances add any predictabililty.

---

[57] Properly, the analysis should account for the fact that z-scores have been used as the dependent variable by reducing the df for the error term by 61. However, this did not change the results.

Table 5.66: The correlational analysis of the RT to clicks within the melody "Softly Now the Light of Day".

Stage 6	simple r	part r	partial r	ΔR^2	$F(\Delta R^2)$	$p(\Delta R^2)$
Position	-0.022	-0.026	-0.026	0.003	0.535	0.659
Position^2	-0.003	-0.000	-0.000			
Position^3	0.008	0.003	0.003			
Boundary	0.130	0.070	0.071	0.015	7.285	0.007
Previous Boundary	-0.066	0.035	0.035	0.004	0.926	0.397
Next Boundary	-0.124	-0.053	-0.054			
Total Effect				0.022	1.793	0.099

Stage 8	simple r	part r	partial r	ΔR^2	$F(\Delta R^2)$	$p(\Delta R^2)$
Position	-0.019	-0.086	-0.087	0.031	5.194	0.015
Position^2	-0.036	-0.095	-0.097			
Position^3	0.062	0.144	0.146			
Boundary	-0.011	-0.016	-0.017	0.002	1.049	0.306
Previous Boundary	0.013	-0.029	-0.030	0.003	0.814	0.444
Next Boundary	0.013	0.049	0.050			
Total Effect				0.037	3.041	0.006

Generally the results of Stage 6 and 8 offer mild support of the notion that boundaries represent the closure of units. Note that the two stages provide similar values in most notes of the melody (see Figure 5.17), particularly in the openning stages. It must be remembered that different locations for the clicks would have been used within each stage for each subject (click positions were randomly chosen for each subject). However, clicks on boundaries would have been located in the same places because subjects were consistent in their locations of boundaries across repetitions. Generally, Figure 5.17 seems to imply that subjects adopted a strategy of listening for the clicks as the melody progressed (variation decreases with increasing note event, even though boundary locations continued to show the same range).

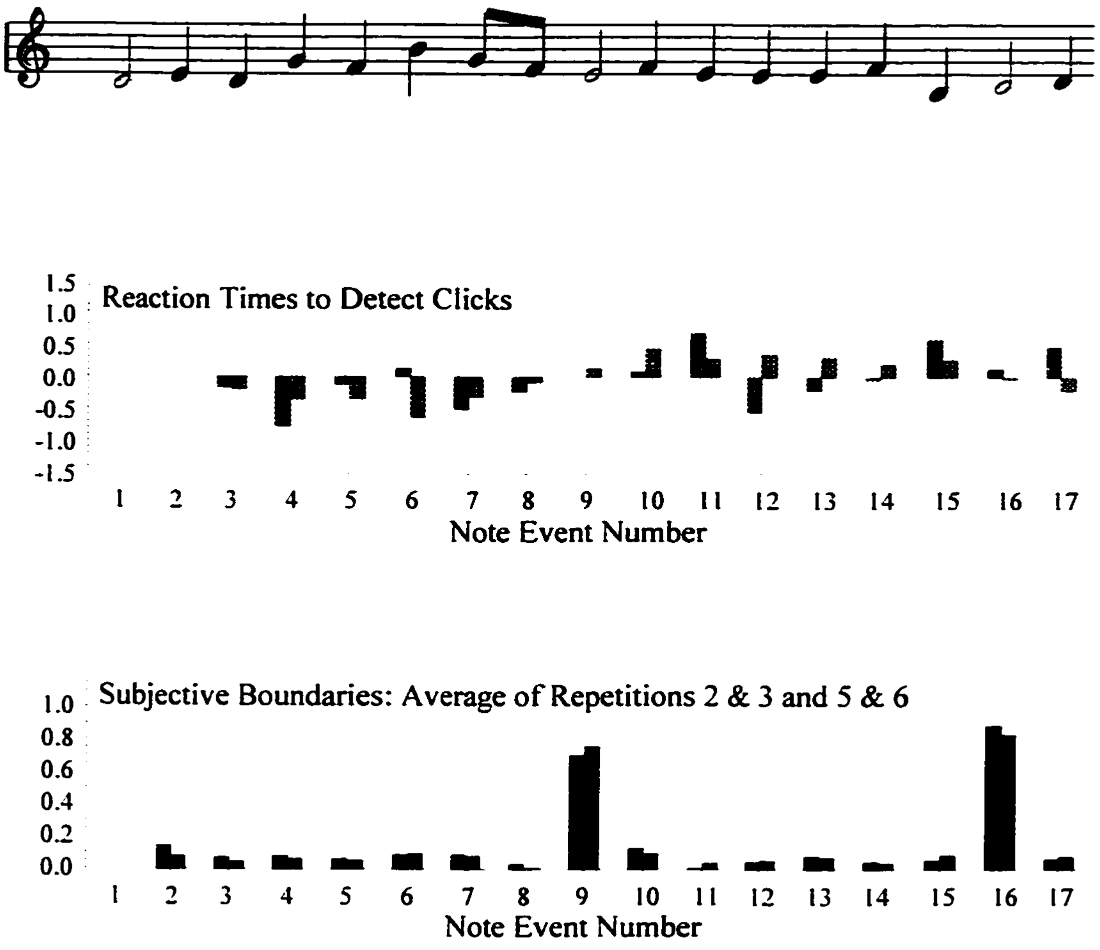


Figure 5.17a: The first part of the song used in Stages 6 & 8, Part 1, the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects, and the z-score of the time to detect clicks placed at various positions.

■ Average Profile (Repetitions 2 & 3)
■ Average Profile (Repetitions 5 & 6)
■ Click Profile (Stage 6)
▨ Click Profile (Stage 8)

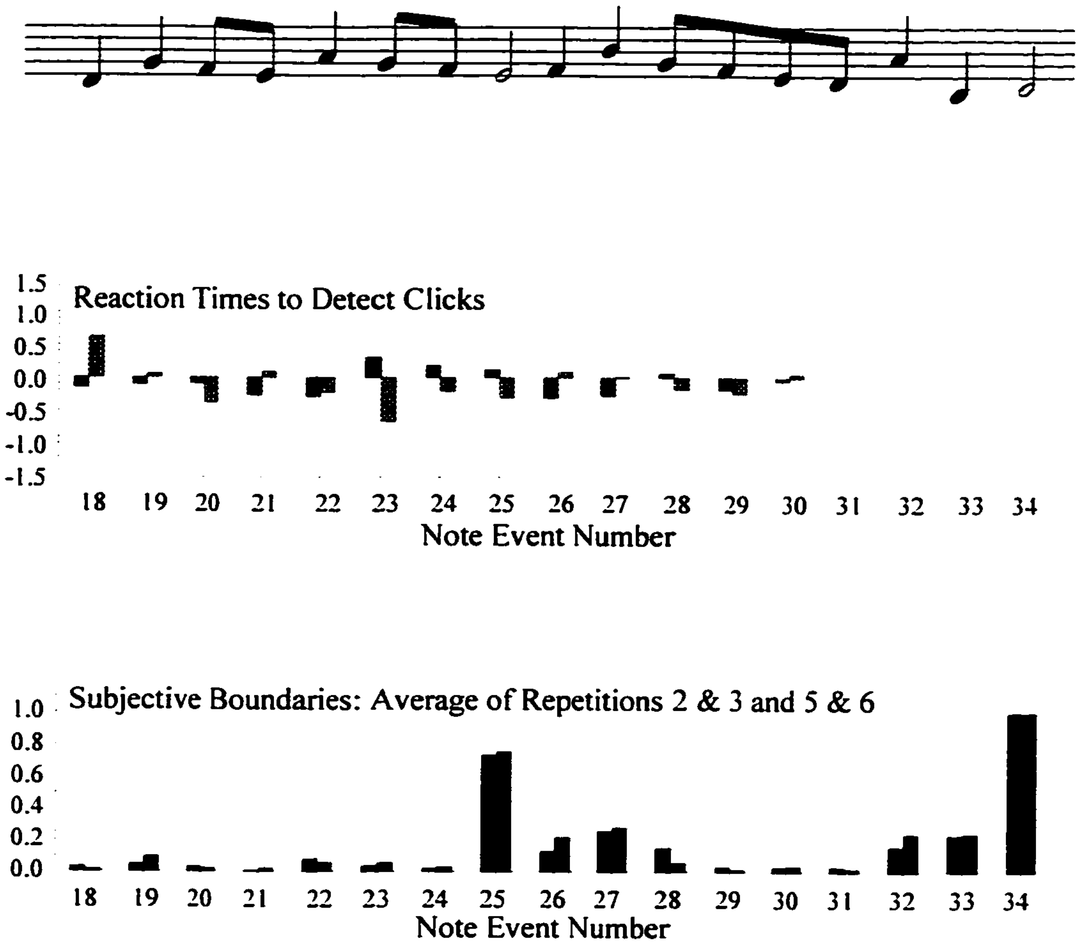


Figure 5.17b: The second part of the song used in Stags 6 & 8, Part 1, the location of subjective boundaries based on the average of the second and third repetitions or fifth and sixth repetitions for all subjects, and the z-score of the time to detect clicks placed at various positions.

■ **Average Profile (Repetitions 2 & 3)**	■ **Click Profile (Stage 6)**
■ **Average Profile (Repetitions 5 & 6)**	■ **Click Profile (Stage 8)**

Discussion

The main purpose of this experiment was the verification of the boundary assessments results of the previous experiment using a design that did not impose a memory task. In this regard, the experiment can be considered an unqualified success.

In this experiment, as in the last, individuals having a wide variety of musical backgrounds were able to parse both a familiar nursery rhyme and an unfamiliar, but tonal, song using the direct on-line procedure. Subjects' responses were highly consistent across repetitions and showed the expected dependency on familiarity. That is, subjects' responses were less consistent with an unfamiliar tune and subjects's responses became more consistent over repetitions. Furthermore, different subjects demonstrated a great deal of similarity in their responding.

More importantly, the different subjects in the two experiments parsed the song in very similar, though not identical, manners. That is, with respect to the placement of boundaries by subjects in the two experiments, inspection of the boundary profiles indicates that subjects performed the same in both experiments. The correlations between boundary profiles produced in Experiments 1 and 2, for the first melody are presented in Table 5.67 (correlations within each experiment have been presented previously -- they were equally high or higher). Note that all were very high.

Table 5.67: Correlations between boundary profiles of the two experiments for the melody "Three Blind Mice"

		Second Experiment			
	Rep	1	2	3	2 & 3
First Experiment	1	.983	.980	.995	.983
	2	.986	.987	.999	.988
	3	.982	.985	.999	.985
	2 & 3	.983	.979	.985	.983

Notes: all values are significant with $p<0.001$

2 & 3 is a mnemonic for the average of Repetitions 2 and 3

The same was true for the second melody, the extract from "Softly Now the Light" (Table 5.68). Again, note that all correlations were high. It must be remembered that these correlations represent two different groups of subjects, roughly equivalent in mean and breadth of training.

Table 5.68: Correlations between boundary profiles of the two experiments for the extract from the melody "Softly Now the Light of Day"

		Second Experiment							
	Rep	1	2	3	4	5	6	2 & 3	5 & 6
First Experiment	1	.766	.776	.762	.786	.773	.757	.771	.768
	2	.745	.815	.796	.794	.813	.792	.808	.805
	3	.741	.815	.803	.790	.819	.800	.811	.812
	4	.758	.805	.791	.798	.805	.787	.800	.799
	5	.730	.816	.802	.789	.814	.798	.811	.809
	6	.729	.806	.794	.777	.813	.797	.802	.808
	2 & 3	.745	.817	.801	.794	.818	.798	.811	.811
	5 & 6	.733	.815	.803	.787	.818	.802	.811	.812

Notes: all values are significant with $p<0.001$
2 & 3 is a mnemonic for the average of Repetitions 2 and 3
5 & 6 is a mnemonic for the average of Repetitions 5 and 6

Hence, it can be said with some certainty that the memory requirement of the Experiment 1 did not force an artifactual parsing of the melody.

With respect to the details of parsing, subjects tended to use groupings of six to nine notes, which is within the range of the magic number for short term memory (7±2), but there was a wide range. Units consisting of a single note event and units consisting of 10 or more note events were not common, but they were not entirely absent either.

By inspection of Figures 3.16 and 3.21, or 5.8 and 5.13, it appears that subjects preferred to place boundaries at relatively longer notes that ended a smooth contour. This can be seen at Note Events 3, 5, 10 14 and 23 of the first song ("Three Blind Mice"; Figures 3.20 or 5.8) and at Note Events 9, 16 and 25 of the second song (the extract from "Softly Now the Light of Day"; Figures 3.28 or 5.13). It is not likely a coincidence that these longer notes correspond to tonal centers within the tonality of the song.

As in the previous experiment, the second purpose of this experiment concerned the empirical verification of Lerdahl and Jackendoff's (1983) Group Preference Rules 2 through 6. In this, the experiment can be described to have a success that is consistent with the previous. That is, if one considers verification of the results of the previous experiment as the goal, then this aspect of the experiment was a complete success. If one considers verification of the rules of Lerdahl and Jackendoff as the goal, then this experiment is as successful as the previous (Chapters 3 and 4 combined). Of the rules tested, only one was found to have strong empirical verification (Rule 2b: boundaries due to changes in the lengths of the notes) which coincides with the previous observations. Of the low level rules (i.e., Rules 2 and 3), only Rule 2b (Attack Point Change) was found to consistently produce strong results. This is consistent with previous observations that boundaries tended to correspond to relatively longer notes. Both melodies in both parts produced similar correlations and similar relative errors in the non-linear regression analysis. Note that this comparison considers only those analyses that were based on the averaged profile of the last two repetitions of a stage (i.e., Repetition 2 & 3 averaged in Stages 4 and 6, and Repetitions 5 & 6 averaged in Stage 8). Table 5.69 provides the details of this similarity for the first melody.

Note that the degree of fit for the second experiment is slightly lower than for the first experiment. This implies that the memory task did induce a quantitative change in performance. One can conclude that this is only a quantitative change (i.e., a change in degree or amount) and not a qualitative change (i.e., a change in response strategies or methods) because the raw data from the two experiments, the boundary profiles, show such a high degree of similarity. That is, performance defined as adherence to the rules, was slightly lower in the second experiment, but the same rules were used with the same relative importance. Note that the best fit analysis of the combination of the low level rules (Rule 4: Intensification) resulted in the same degree of improvement in both experiments, even though the percentage improvements were not the same (i.e., the magnitudes were similar, but the percentages are not because the initial values were different). In both cases the difference between the overall error rate and the individual miss and false alarm error rates represented the fact that the rules do not account for all empirical boundaries.

Table 5.69: Comparisons across experiments of the applications of Rules 2b and 3a and their combination (Rule 4) in the melody “Three Blind Mice”.

Rule	Measure	Experiment 1	Experiment 2
Rule 2b	r	0.770**	0.704**
	SS_{error}	3.592	8.910
	Misses SS_{error}	0.907	0.687
	False Alarms SS_{error}	1.933	1.651
Rule 3a	r	0.340**	0.286**
	SS_{error}	12.777	18.668
	Misses SS_{error}	1.987	1.778
	False Alarms SS_{error}	4.166	3.793
Rule 4 (2b + 3a)	SS_{error}	3.540	8.874
	Misses SS_{error}	0.828	0.639
	False Alarms SS_{error}	1.991	1.698
	Change Overall	0.052	0.036
	% Improvement	1.45	0.40

The situation for the second melody is similar, though there are more comparisons to consider (see Table 5.70). Again, the best fit comparison resulted in the same degree of improvement in all analyses.

Table 5.70: Comparisons across experiments of the applications of Rules 2b, 3a and 3d and their combination (Rule 4) in the melody "Softly Now the Light".

Rule	Measure	Experiment 1		Experiment 2	
		Stage 6	Stage 8	Stage 6	Stage 8
Rule 2b	*r*	0.877**	0.883**	0.742**	0.738**
	Overall SS_{error}	1.650	1.520	7.418	7.496
	Misses SS_{error}	0.237	0.271	0.303	0.200
	False Alarms SS_{error}	0.454	0.567	0.398	0.416
Rule 3a	*r*	0.313	0.310	0.237	0.241
	Overall SS_{error}	9.104	10.286	13.842	14.314
	Misses SS_{error}	0.689	0.736	0.704	0.623
	False Alarms SS_{error}	1.708	1.844	1.703	1.552
Rule 3d	*r*	0.367*	0.334	0.284	0.296
	Overall SS_{error}	11.110	13.023	16.335	16.419
	Misses SS_{error}	0.147	0.162	0.164	0.154
	False Alarms SS_{error}	0.610	0.633	0.564	0.525
Rule 4 (2b + 3a + 3d)	Overall SS_{error}	1.467	1.380	7.183	7.208
	Misses SS_{error}	0.500	0.486	0.386	0.437
	False Alarms SS_{error}	0.551	0.650	0.468	0.535
	Change Overall	0.183	0.140	0.235	0.288
	% Improvement	11.09	9.21	3.16	3.84

For the more complex rule, parallelism (Rule 6; see Table 5.71 and 5.72) in its four guises, the results were also fairly consistent (Rule 4 is included as a baseline for comparison). To simplify this comparison, only the basic version of the rule is considered (i.e., only the main effect; interaction terms were not included in the error comparisons).

but as was seen before, the interaction terms generally did not contribute much. However, again, the second experiment does demonstrate higher levels of overall error.

Table 5.71: Comparisons across experiments of the applications of the four versions of Rule 6 in the melody "Three Blind Mice".

Rule	Measure	Experiment 1	Experiment 2
Rule 4	SS_{error}	3.540	8.874
Rule 6a (plus 2b + 3a)	SS_{error}	3.376	8.694
	Misses SS_{error}	0.820	0.639
	False Alarms SS_{error}	1.835	1.698
	Change Overall	0.164	0.180
	% Improvement	4.63	2.03
Rule 6b (plus 2b + 3a)	SS_{error}	3.204	8.543
	Misses SS_{error}	0.929	0.679
	False Alarms SS_{error}	1.554	1.326
	Change Overall	0.336	0.331
	% Improvement	9.49	3.73
Rule 6c (plus 2b + 3a)	SS_{error}	3.540	8.871
	Misses SS_{error}	0.828	0.639
	False Alarms SS_{error}	1.991	1.694
	Change Overall	0.000	0.003
	% Improvement	0.00	0.03
Rule 6d (plus 2b + 3a)	SS_{error}	2.932	8.274
	Misses SS_{error}	1.095	0.778
	False Alarms SS_{error}	1.116	0.958
	Change Overall	0.608	0.600
	% Improvement	17.18	6.76

Table 5.72: Comparisons across experiments of the applications of the four versions of Rule 6 in the melody "Softly Now the Light of Day".

Rule	Measure	Experiment 1		Experiment 2	
		Stage 6	Stage 8	Stage 6	Stage 8
Rule 4	SS_{error}	1.467	1.380	7.183	7.208
Rule 6a (plus 2b + 3a + 3d)	SS_{error}	1.416	1.353	7.151	7.153
	Misses SS_{error}	0.403	0.432	0.317	0.336
	False Alarms SS_{error}	0.598	0.676	0.504	0.580
	Change Overall	0.051	0.027	0.032	0.055
	% Improvement	3.48	1.96	0.45	0.76
Rule 6b (plus 2b + 3a + 3d)	SS_{error}	1.466	1.380	7.183	7.199
	Misses SS_{error}	0.491	0.432	0.386	0.397
	False Alarms SS_{error}	0.560	0.650	0.468	0.565
	Change Overall	0.001	0.000	0.000	0.009
	% Improvement	0.07	0.00	0.00	0.01
Rule 6c (plus 2b + 3a + 3d)	SS_{error}	1.285	1.211	7.067	6.994
	Misses SS_{error}	0.307	0.311	0.261	0.211
	False Alarms SS_{error}	0.570	0.656	0.476	0.548
	Change Overall	0.182	0.169	0.116	0.214
	% Improvement	12.41	12.25	1.61	3.00
Rule 6d (plus 2b + 3a + 3d)	SS_{error}	1.210	1.176	7.018	7.051
	Misses SS_{error}	0.268	0.319	0.240	0.322
	False Alarms SS_{error}	0.567	0.613	0.448	0.493
	Change Overall	0.257	0.204	0.169	0.157
	% Improvement	17.52	14.78	2.35	2.18

All these results are completely consistent with the notion that the raw data (i.e., the boundary profiles) from the two experiments show a high degree of similarity. Indeed, to the eye, there is little to distinguish them. Hence, on this basis one could conclude that the basic task and its findings are reasonably sound.

On the other hand, the results with the secondary task of click detection, intended as an assessment of boundary efficacy are much more equivocal. Generally, the results in Stage 4 ("Three Blind Mice") do offer some support for the hypothesis tested, but the results in Stages 6 and 8 do not. This is not completely consistent with previous literature (Gregory, 1978; Sloboda & Gregory, 1980), but the differences can easily be explained by the move to naturalistic stimuli, the use of individual subject-delineated boundaries and the use of multiple boundaries and clicks within a melody. All of these aspects of experimental design have been defended at length so that argument will not be reiterated here. Suffice it to say that the intent to gain external (or ecological) validity has once again butted heads with the problem of internal consistency and/or sensitivity.

The results did seem to indicate that reaction times to clicks on boundaries were longer (not a part of the original hypothesis), but this effect was not consistent in Stage 8. Note that the result was actually counter to that of Stoffer (1985; reaction times shorter on boundaries), though the stimuli used by Stoffer (binary and trinary forms) should not be compared directly with the simple melodies used herein. These results also seem, intuitively, to indicate that reaction times were longer in the more complex sections, but this cannot be rigorously verified until some quantifiable notion of complexity within music is applied (cf., Berent & Perfetti, 1993).

Also note that reaction times tended to show less differentiation over the course of the melody. When comparing the results of musicians and non-musicians in their click detection task, Berent and Perfetti noticed that reaction times decreased over the duration of the trial. They also commented that, in their task, non-musicians seemed to be using a strategy that allocated resources to click detection (and not to listening to the music; 1993, p. 215) effectively ignoring the music: Musicians did not seem to use this strategy. This is a possible explanation for some of the observed results -- the general lack of differentiation in reaction times in the latter half of the melodies. If some subjects, possibly those with less involvement in music, adopted such a strategy, the observed pattern would be predicted.

At this point, all that should be said is that this methodology needs to be refined before strong conclusions can be made. Since this task was secondary to the goal (and could have been eliminated completely given the goals of the analysis), it has not been pursued at this time.

General Discussion

The present thesis had three main goals. Firstly, the parsing of melody by listeners was assessed (also called, the empirical assessment of boundary location). Secondly, the location of these boundaries was related to Lerdahl and Jackendoff's (1983) theory of musical parsing. Thirdly, the efficacy of these empirical boundaries for the subsequent processing of music was tested. This work can be considered a success with respect to the first two goals, and a qualified success with respect to the third.

In the present work, it was assumed that listeners hear a melody as a series of smaller units, that is as motives, phrases, sections et cetera. To determine boundary locations between the units of melody, an on-line methodology for the determination of empirical boundary locations was developed. This method was employed in two experiments. However, to discuss and interpret the observed boundary locations, a suitable theoretical framework was required. The chosen framework was that of Lerdahl and Jackendoff (1983) because it is well placed within music theory (cf., Handel, 1993; Jackendoff, 1993; Jones & Holleran, 1992; Lerdahl, 1993; see particularly Lerdahl & Jackendoff, 1983) and because it has been empirically tested (e.g., Clarke & Krumhansl, 1990; Deliege, 1987, Krumhansl, 1996; Peretz, 1989). Lerdahl and Jackendoff's theory was also amenable to quantification. Finally, it was assumed that the unitization of a melody (boundary formation) would have an effect on the processing of that melody. In particular, it was assumed that the unitization of the melody would affect recall of the melody and that the processing of the melody (the creation of units or boundary formation) would have an impact on other concurrent processing.

The predictions of the theory were quantified and then two different experiments were conducted with two groups of subjects. Both experiments tested subjects having a wide range of training and involvement with music. Both experiments required subjects to parse a familiar nursery rhyme ("Three Blind Mice") and an unfamiliar, tonal piece ("Softly Now the Light of Day"). The secondary task in the first experiment required subjects to commit the melody to memory. The secondary task in the second experiment did not have this memory requirement. The empirical parsings produced in both experiments were related to the theoretical notions of Lerdahl and Jackendoff (1983).

The Methodology for Boundary Location

The basic methodology developed to determine the positions of boundaries required subjects to press a key each time they felt that one unit had ended and a new one started. This methodology was adapted from studies in social psychology (Newtson 1973, 1976; Hanson, & Hirst, 1989; Massad, Hubbard & Newtson, 1979; Newtson & Engquist, 1976; Newtson, Engquist & Bois 1977; Newtson, Hairfield, Bloomingdale & Cutino, 1987) and music perception (Deliege, 1987; Peretz, 1989). In both experiments, subjects created reliable parsings of both melodies. Reliability was assessed as consistency (within subject) and similarity (between subjects).

Consistency Analysis (Within-Subjects Analysis)

There was high consistency within subjects across different repetitions of the same melody. That is, each subject produced the same parsing on three repetitions of the same melody. Consistency did not seem to be related to musical training (i.e., all subjects performed about the same regardless of training). Consistencies were slightly lower in the unfamiliar piece, but still well above chance. The implication is that this task can be used to reliable parsings of a melody. It is important to remember that high consistencies were obtained regardless of the context of the task. That is, performance in the experiment that required memorizing the melody (see Chapter 3 of this work) was essentially the same as performance in the experiment that did not require memorizing the melody (see Chapter 5 of this work). Hence, boundary formation was not a consequence of a desire to remember the melody.

With respect to within-subject consistency, it seems clear that subjects can perform this task reliably on the first repetition. That is, on average, the parsing of the first repetition was very similar to the parsing of the second and third repetitions -- even for an unfamiliar melody. However, consistencies were slightly higher for sequentially adjacent repetition of each melody (i.e., Repetitions 1 & 2 were more similar than Repetitions 1 & 3). Peretz (1989) also asked subjects to parse melodies (she did not use an on-line task however). Subjects were permitted to replay a melody once. Only 3% of the stimuli needed to be presented a second time, supporting the notion that subjects can reliably parse on the first attempt. In a very similar task, Deliege (1987) allowed subjects as many repetitions of the musical extract as desired. She noted about 50 repetitions for every 100 stimuli, implying that subjects could parse on the first repetition (actually, the first parsing was the second time the subjects heard the musical extract, but this must be qualified by the fact that the extracts were from instrumental or orchestral sequences).

In the present work, detailed analyses in the both experiments (only the detailed analyses of the first experiment were presented) indicated that the different repetitions did not produce any major systematic changes. That is, there were some changes across repetitions, but those changes were minor relative to the backdrop of consistency. Hence, future studies may be able to trade multiple repetitions of the same melody (to assess/insure consistency and/or learning) for single repetitions of many melodies. This would allow the empirical examination of a large corpus in a relatively short time frame. This in turn, would allow for more complex model testing and development.

Conversely, if multiple repetitions are used, then it seems reasonable to average all repetitions with only a limited analysis of consistency. The use of an average of multiple repetitions would allow for a more accurate assessment of boundary locations of individual subjects (standard errors for each boundary position for each subject would be smaller when based on two or more repetitions[58]). This would be most useful if those boundary locations were to be a basis for subsequent tests or analyses. For example, in the experiments used herein, if one had known, a priori, that the within-subject consistency would be high, then the secondary tasks of boundary efficacy could have been based on the average individual boundary profile (i.e., the average of Repetitions 2

---

[58] At each note position in the melody, multiple presentations to individual subjects essentially produces a bimodal distribution with a mean of np and variance of npq (where, n is the number of repetitions, p is the probability of a boundary and $q=1-p$). In fact, this concept was the basis for the between-subjects comparison. Stable boundaries would be demonstrated by many/all subjects and less stable boundaries would be demonstrated by only a few subjects, with the same binomial distribution.

and 3 of each Stage), rather than on only the last boundary profile (i.e., just Repetition 3 of each Stage). Test sequence extraction and click positioning could have been based on two or three repetitions. This might have provided a more accurate assessment of the "true" individual boundary profile and hence a better assessment of the relationship between boundary location and measures of boundary efficacy (see below). In this view, an individual boundary profile based on multiple repetitions could be considered as a measure of the locations where the subject always placed a boundary, versus the locations where the subject sometimes placed a boundary, versus the locations where the subject never placed a boundary.

Such averaging might also enhance the comparisons between subjects (i.e., the Similarity Analysis: see below). That is, the profile based on the average of two repetitions for each subject would indicate which boundaries are the most stable, and which boundaries are more flexible for that subject. Groups of subject might be delineated on the basis of such finer distinctions.

Note that, even though this technique is advocated for future studies, this technique was not used in this thesis because the secondary boundary efficacy tasks (the memory recognition task and the click detection task) were tied to the third repetition of each melody. Since one goal of this study was an assessment of within-subject consistency, high within-subject consistency could not be assumed a priori and therefore, the averaging of multiple repetitions could not be a design factor for the tasks used in these experiments. After more studies have confirmed these results with other stimuli, the use of average profiles is advocated for the comparison of subjects and for the construction of stimuli for the secondary "boundary-efficacy" tasks.

On the other hand, the fact that there was high consistency, even when considering the first presentations of melody (even an unfamiliar melody) argues that it may be difficult to see the acquisition (learning) of a melody using this task. That is, one may not be able to plot the developmental course as subjects become more familiar with a melody. One might expect that the initial parsing of a melody changes as experience with the melody increases. Individuals might be expected to regroup note events as the perception of the whole exerts a greater influence on the relationships between individual note events. Indeed, the data provided herein do tend to support the notion that subjects work from the global to the particular: The number of boundaries tended to increase slightly with repetition number. However, the effect was not very pronounced. Longer, more complex, melodies might show more of an effect. A greater delay between repetitions might also show greater within-subject variability, particularly when combined with longer, more complex, pieces. In addition, though it was not attempted at this point, it might be fruitful to examine separately those subjects who demonstrated the lowest and the highest consistencies. This in turn, might lead to the development of an empirical criteria for more precisely defining consistency within the parsing of melody (the definitions used herein were based on theoretical notions of consistency: see Table 3.8).

Similarity Analysis (Between-Subject Analysis)

There was high similarity between different subjects within the third repetition. That is, different subjects having a wide range of training parsed both the familiar and unfamiliar melodies in much the same manner. A cluster analysis did not reveal any distinct groups of subjects for either melody. Again, the experimental context (boundary formation assessed within experiments that did or did not require memorizing the melody) did not affect results. Since subjects tended to parse all melodies in a consistent manner across repetitions and since subjects produced similar parsings on the third repetition, one can conclude that most subjects would likely parse a melody in the same fashion when first hearing the melody. This has some important implications for theories of music cognition. If unitization is the same over a wide range of training[59], then training must contribute something other than knowledge of how to create units -- perhaps a richer knowledge of how units relate to each other within the global structure of the piece. To

---

[59] This is not an argument that the parsing of melodies is innate. The lack of a training effect may imply that the mechanisms for parsing are innate, or that those mechanisms are learned very early in life. It can only be said that, by the time listeners reach the age of the subjects who participated in the experiments of the current thesis, they have acquired, essentially, the same mechanisms for parsing music.

resort to the analogy with speech once more, individuals may be able to parse a speech sentence (that is in their own language) into its component words without fully understanding the meanings of each individual word or the relations between those words (cf., Clark & Clark, 1977, particularly Chapters 2 and 5). For example, an individual lacking training in computer science may be able to parse the speech of a computer professional without comprehension. It is possible that in music, as in speech, training and/or experience provides one with a richer understanding of the "meaning" of units and the relationships between units.

As mentioned previously, the analysis of similarity could benefit from the high degree of consistency demonstrated within each subject, across different repetitions. That is, the detailed comparison between subjects (i.e., the similarity analyses) could be based on individual subject profiles that represent the average of many repetitions for each subject. This would improve the precision of the measure of each subject's profile (a continuum instead of a dichotomy). This, in turn, would also result in a more accurate comparison between subjects.

Although similarities were high enough in both experiments that it was reasonable to consider all subjects (in each experiment) to consist of one single group, it is possible that a more fine-grained analysis might reveal somewhat distinct subgroups. Deliege (1987) and Peretz (1989), for example, implied differences in the boundary profiles of subjects with and without musical training. The differences were implied because the analyses were directed at the subjects' use of the Lerdahl and Jackendoff's (1983) rules and not boundaries per se (although different rule use does imply different boundaries, it does not demand different boundary profiles). In addition, in both studies music groups were defined, a priori, to represent ends of the training continuum (effectively none versus a great deal). By sampling the continuum of training, the present work may have obscured differences that might exist between the ends (but see the later tonality discussion). It should be pointed out that adopting a more stringent criteria for the delineation of boundary profiles was considered and rejected in the present work. By careful examination of the clustering profiles for Stages 6 and 8 (these two stages used the same melodies), one can see that in general, a tighter criteria would not have resulted in consistent homogeneous subgroups across stages. That is, the subjects who would have been grouped together in Stage 6 would not have remained together in Stage 8. This implied that creating groups based on a more stringent criteria would have amounted to an artificial separation based on noise. Once again, it must be mentioned that the use of the average of Repetitions 2 and 3 of each stage (as implied by the consistency analysis), might have improved such an analysis.

Given that this work has demonstrated that the average profile based on all subjects is generally representative of individual subjects, it might be possible to look for subgroups that represent minor but consistent deviations from that average profile. These minor deviations may be related to stylistic preferences, to training or to the development of the unitization of each melody. Although each experiment did have about 60 subjects, it is possible that the cluster analysis did not reveal systematic groups because the pattern of differences obeys different statistical rules (a pattern that is not easily captured within a cluster analysis). To expand on this notion, consider a span of nine note events. Within that span of nine notes events, there could be:

1 one-unit grouping consisting of nine notes
8 different two-unit groupings (1 and 8 notes, 2 and 7 notes, 3 and 6 notes, etc.)
28 different three-unit groups.

Hence, within a single nine-note span, there are many possible unitizations, even if one assumes that each unit must have at least one note event and all notes in a unit must be adjacent. Furthermore, there are several different nine note spans within a usual melody (i.e., "Three Blind Mice" had 48 notes and "Softly Now the Light of Day" had 34 notes). As a hypothetical example, consider two groups ("A" and "B") each consisting of 20 subjects. Everyone in Group "A" used the same structure to subdivide the first nine-note span into two units of four and five notes each. Everyone in Group "B" divided that same nine-note span into three three-note units. Assume that there is a second nine-note span in melody that is repetition of the first nine-note span (avoiding the term "phrase"). In the second nine-note span of the melody, ten subjects from Group "A" divided the span into three three-note sequences ("A1") while the remaining ten divided the second nine-note span into a group of four and a group of five ("A2"). Hence, half the subjects in group A used a consistent parsing for the two nine-note spans; the other half did not. Similarly, in Group "B", ten subjects ("B1") divided the second nine-note span into three three-note groups, and ten subjects ("B2") divided the second nine-note span into a four and a five note group. There would be, in reality, only two patterns of parsing (i.e., a four-note and a five-note parsing or 3 three-note parsings), but the net result would be four different boundary profiles (4 different groups). Now consider this process for a melody of 36 or 45 notes (4 or 5 nine-note spans) with "only" 60 subjects. Difficulties like this imply that the cluster analysis would not produce distinct groups even though there might be distinctly different subpatterns at different points of the melody. Note that in the boundary profiles of both melodies (in both experiments), there were many points in the melody that had some boundaries indicated. Future studies might need more subjects to detect such differences. Future analysis will be aided by the fact that the consistency analysis has indicated that different repetitions can be averaged to create a more accurate individual boundary profile.

On the other hand, in place of ever increasing numbers of subjects, analytic methods could be refined or developed to capture these relationships (i.e., there is similarity here with the notion of Markovian chains). As noted, the cluster analysis used herein is actually a fairly blunt tool for such exploration (this was the reason that, as was noted previously that cluster analyses were augmented by examination of profiles at each stage of the clustering; see Chapter 3, Tonality). One possible analysis would use the average profile as a baseline. Individual subjects could then be compared to the baseline and branches of similar subjects could be developed (this was actually used to determine the optimal alignment between subjects). However, this level of analysis will require more information about the basis of parsing. That is, because there are so many ways to divide the pie, theories like that of Lerdahl and Jackendoff (1983) are needed to help distinguish noise (happenstance subgroups of subjects) from consistent subgroups with meaningful boundary profiles (even groups with somewhat idiosyncratic parsings). As stated in the introduction, the parsing of melody is simply too complex to be approached atheoretically. Unfortunately, theories, including that of Lerdahl and Jackendoff, are not yet well enough developed to provide such a fine-grained analysis. Hopefully, this work has provided empirical feedback to refine theories.

The Parsing of Melody

In both experiments, subjects were requested to parse into the smallest units that made sense. In music theory, the smallest unit is the usually considered the motive, consisting of about three to four notes (see Chapter 1 of this work; one might also consider the interval formed by two notes to be a unit). In some instances, such as the opening of "Three Blind Mice", the units delineated seemed to correspond to the classical notion of a motive (see Figure 3.16). In other instances, such as the middle of the extract from "Softly Now the Light of Day" (particularly Note Events 19-21 and 22-24, but also Note Events 2-4 and 9-11; see Figure 3.21), most subjects seemed to have ignored an obvious motive interpretation. In the melody "Three Blind Mice", there were strong boundaries at Note Events 3, 6, 10, and 14, with weaker boundaries near 23 and 34. These corresponded to units of 3, 3, 4, 4, 9, 11 and 14 (mean 6.8) notes. For the extract from the melody "Softly Now the Light of Day", there were strong boundaries at Note Events 9, 16, and 25 with a weaker boundary near 27, corresponding to units of 9, 7, 9, 2 and 7 (mean of 6.8) or 9, 7, 9, 9 (mean of 7.5). Hence, units seemed to be a little larger than the usual motive size. Herein, boundaries delineated units with sizes between six- and nine-note events, as defined by means, medians and modes. It is interesting that boundaries tended to delineate units of six to nine notes which is close to the short-term memory limit of 7±2. All the central tendencies reflected this value, but there were some larger units (larger than the short-term memory limit). Measures were taken of the individual unit sizes used by each subject. These measures also indicated that, although units the size of the motive were common, there were units that were much larger than the motive.

On closer consideration, it is difficult to imagine how one could produce a mean value for the unit size that is less than about six note events. One needs two -- more likely, three -- note events to define a unit. To produce a *mean* unit size of three or four notes, given that the minimum unit size would necessarily be two or three, would require that all units be three or four notes. Music theories never state that all motives must be three or four notes and often examples provided are longer than this. Although the more detailed analysis demonstrated that a range of sizes was used, that analysis also indicated that most units were less than 10 notes long[60]. Occasionally units of size one were indicated, but more often units of three to five notes were indicated. Results were consistent with the notions that parsing reflects short-term memory constraints (i.e., 7±2 items) and notes represent the basic items of short-term memory.

One might wish to argue that the mean sizes of units is suggestive of parsing at a level that is closer to the phrase, though tending in the direction of the motive (i.e., not tending to structures beyond the phrase). That is, if the motive is considered a short three or four note segment of music while the phrase is generally conceived as two or more motives, then units of seven or so notes should be considered as phrases. It is possible that some structure having attributes of both the phrase and motive is the basis of musical perception (i.e., something between the motive and the phrase, though closer to the phrase). It is also possible that it might requires some practice or training to parse a melody at the level of the motive. By analogy with speech, it may be difficult for subjects to parse at the level of the phoneme There is a similar analogy in visual processing: Visual primitives are not noticed by individuals who attend and process at the level of the object . The point is that if this is true, then the tasks developed in this thesis could be used to explore larger units, and possibly the development of the organization of larger units. However, parsing into smaller units would not seem reasonable or might require

---

[60] One might also argue that the rare large units represented an aspect of performance on the boundary location. If a subject "forgot" or missed a particular key press, there was no way for a subject to retrospectively re-insert that missing boundary. Hence, the large units would not accurately represent the unitization of the melody. Multiple repetitions was the method chosen to assess this effect. Had consistencies been low, this might have been explored as an explanation. The use of the average profile of two or more repetitions might reduce this concern.

extensive training. This is simply a caution for future work using this task.

Comparisons with Lerdahl and Jackendoff's Group Preference Rules

The Low-Level Rules

Lerdahl and Jackendoff's theory of musical perception has been criticized as a theory that applies only to a "listener" who has a visual representation of the musical score to guide his parsing of the melody. However, this work successfully related the empirical boundaries generated by subjects on-line (without a score, and even without high familiarity with the melodies; See familiarity scores for the second melody used) to the theoretical notions advocated within the Group Preference Rules of Lerdahl and Jackendoff (1983). Essentially, the data confirmed the importance of Rule 2b (Attack Point Changes). In both melodies, this was the most important predictor, regardless of the analysis used. The other rules tested, Rule 3a (Register Changes) and 3d (Length Changes) were much less important. Of the remaining low level Group Preference Rules, Rule 2a (Slur/Rest) did not apply within the melodies used, and Rules 3b (Dynamics Change) and 3c (Articulation Change) were controlled at the level of stimulus presentation.

These results are consistent with both Deliege (1987) and Peretz (1989), although Deliege used more rules (but Rule 2b was still the best) and Peretz only used Rules 3a and 3d (possibly Rule 2b, given the examples). Deliege noted that parsing was always consistent with Rule 2b (particularly in Experiment 1, which most closely resembles the present work), while parsing was often consistent with Rules 3a and 3d. It should also be noted that Deliege found essentially no difference in the use of Rule 2b, between subjects differing in their level of musical training. She did find a difference between these groups in their use of Rules 3a and 3d. The fact that Rule 2b is not used differently as a function of training may explain the lack of differences in the previous similarity analysis. That is, if all subjects use Rule 2b equally (in naturalist stimuli) and if Rule 2b is the single most important rule, then boundary profiles would not demonstrate many differences as a function of training. Subjects in Peretz' study tended to respond in accordance with Rule 3d more often than Rule 3a, with some differences between those who had received musical training and those who had not. Again, it must be pointed out that examples provided implied that Rule 3d could have been interpreted as Rule 2b. If so, this result would then be consistent with Deliege and the present work.

It is important to realize that it is Rule 2b (Attack-Point) that captures the longer note that may end a phrase (i.e., the notion of cadence; see Chapter 1 this work). Rule 3d (Length Change) does not capture this effect. Rule 2b proposes a boundary after a single relatively longer note (not a shorter note!). A cadence is often defined by ending on a relatively longer, more tonally important note (see Chapter 1 of this work).

Even though it deals with relative length, Rule 3d proposes a boundary when there are two notes of the same duration followed by two notes of a different (but equal to each other) duration. The boundary is placed between the two sets of notes, before the first long note if the pair of longer notes are Notes 3 and 4. This does not correspond to the traditional notion of a cadence.

Rule 3a addresses changes in pitch height (with no consideration of diatonicism or tonality). It is interesting that Rule 3a was not important for parsing in the present work. The implication is that large pitch jumps do not matter with respect to parsing. This can be related to the work of Deliege (1987) who found that large (greater than an octave) and small (less than an octave) pitch jumps had the same impact on parsing. It could be that a difference in size did not matter because register change as a whole did not matter.

It is important to remember that the interaction of the different rules did not contribute to the prediction. One might have predicted that the interaction of Rule 2b and 3a for example might have indicated a boundary. This would be similar to the notion that the phrase ending corresponds to a arch of motion in pitch space. On more careful consideration, however, it would seem that the interaction of these two rules would not represent such a structure. The reason is that both rules (all low-level rules) are local in time. Each rule only considers four notes of the melody (sometimes, as in Rule 2b, it would be more appropriate to consider this to be three note events). As such, these rules, and their interactions cannot address large scale structure. Such large scale structures are within the domain of Lerdahl and Jackendoff's (1983) high level rules.

Generally, the present work along with Deliege (1987) and Peretz (1989) implies a hierarchy of rule importance. The existence of a hierarchy may make the design of stimuli for use in experimental studies that rely on the perception of a boundary (implicit or explicit) somewhat easier. That is, one can be certain that the application of Rule 2b has the best chance of producing a boundary. However, it must be noted that even though Rule 2b has the best chance, it will not always produce a boundary. Other, as yet, ill defined factors must have a role.

Asymmetrical Error Terms and Non-linear Regression

In addition, in this work, the analysis was refined by the use of asymmetrical error terms. Asymmetrical error terms allowed false alarms to be weighted differently than misses, so that each rule was given the maximal chance to explain empirical boundaries while effectively ignoring any over prediction. When nonlinear regression and asymmetrical error terms were used, Rule 2b was still the most important rule, while Rules 3a and 3d were still much less important. That is, even when false alarms were effectively ignored in the assessment of error, Rule 2b outperformed Rules 3a and 3d.

The use of asymmetrical error terms allowed one to create an approximate scaling factor for each rule (i.e. a "slope") when that rule was used in isolation. That is, when each rule was initially quantified (i.e., rendered in an algorithmic form), the scale factor for converting each rule, as defined by theory, to an empirical boundary was not known. For example, the algorithms defining Rules 2b and 3a could have required that Rule 2b be multiplied by 100 to produce boundaries while Rule 3a only need be multiplied by 1 to create boundaries. The analysis of each rule in isolation defined the best case scaling factor for each rule. Note that when applied in isolation, all rules produced scaling factors ("slopes") that were more or less equivalent in magnitude (i.e., all were somewhere near a value of 1.0)[61], even though all rules were not equally effective at predicting boundaries.

---

[61] The initial application of the rules to the grammatical coding of the lyrics of a number of melodies (see Chapter 2 of the present work) allowed the actual algorithms defining each rule to be adjusted so that all rules were roughly comparable in their scaling factors.

The real importance of the asymmetrical error terms and the associated scale factors was critical for the attempt to quantify some of the more ambiguous higher level Group Preference Rules. Lerdahl and Jackendoff's (1983) structure clearly implies that boundaries at any level of the hierarchy are created by the application of rules from increasingly higher levels of the hierarchy starting at the bottom. Ultimately all boundaries at any level must have been identified as a potential boundary by the application of the low-level rules. The empirical boundary profiles of subjects may correspond to any level of the hierarchy, and likely correspond to some intermediate level. Hence, when comparing the low level rules to the empirical data (the boundary profiles of subjects), the low-level rules may generate a potential for a boundary at more locations than actually exist in the empirical data. Asymmetrical error terms were necessary to extend the analysis from the low-level rules to the higher levels of the hierarchy because only the use of asymmetrical error terms could weight false alarms differently than misses. Asymmetrical error terms were necessary so that the different low-level and intermediate-level rules could be properly combined. The analysis of the intermediate-level rules examined Rules 4 and 6; Rule 5 was considered untestable at this time.

The Intermediate-Level Rules

Rule 4 (Intensification) was conceived as the optimal combination of Rules 2 and 3, where "optimal combination" was defined within the notion of non-linear regression using asymmetrical error terms (the standard notion of an optimal combination using multiple linear regression was also provided as a comparison). The best combination of rules placed the greatest emphasis on Rule 2b (Attack-Point), assigning reduced roles for Rules 3a (Register Change) and 3d (Length Change). When all rules were used in combination, the slope of Rule 2b did not change dramatically from its value when used alone. Conversely, the slopes of Rules 3a and 3d were dramatically reduced from their use in isolation. In particular, the use of asymmetrical error terms with non-linear analysis demonstrated that Rules 3a and 3d could not explain anything that Rule 2b had not explained. That is, the three rules often predicted boundaries at the same points in the melody. Each rule also predicted boundaries at unique points in the melody. However, at the points where they overlapped, only Rule 2b was needed to explain boundaries. At the points where they were unique, Rules 3a and 3d did not predict boundaries while Rule 2b did. It is important to realize that if all the rules had been equally effective at locations where their application was unique, then all rules would have been granted equal importance when using the asymmetrical error terms (their respective slopes would not have changed). As a hypothetical example, consider a melody with seven empirical boundaries. Assume the Rule 2b, 3a and 3d all explain the first four boundaries equally well, but only Rule 2b can explain the fifth boundary, only Rule 3a can explain the sixth boundary and only Rule 3d can explain the seventh boundary. The use of asymmetrical error terms would make all rules equally important because the asymmetrical error terms would effectively ignore the over prediction generated where the rules overlap[62]. This point is critical. The usual linear multiple regression would not be able to capture this. In this example, the usual multiple regression technique would marginalize two of the three rules because they overlapped at boundaries one through four. The only rule that would have been retained in a multiple regression approach would have been the rule that maximally predicted the first four boundaries. This would have made predicting two of the last three boundaries impossible.

In summary, the asymmetric error term (in conjunction with non-linear regression) indicated that only Rule 2b had predictive validity. This is a significant finding for research in this area, both for experimental tests of boundary formation and for the design of stimuli that make some use of a potential boundary. However, given that it is based on only two melodies, the conclusions should not be taken as overly binding.

An attempt was made to quantify parallelism (Rule 6), using four different types of parallelism: Pitch Pattern Parallelism A (pitch height, including duration), Pitch

---

[62] Note that the effectiveness would depend on how the asymmetry had been defined. The most effective would be to give a zero error to over prediction, but this was not used for reasons that were detailed during the exposition of these concepts.

Pattern Parallelism B (pitch height based onset to onset, ignoring duration), Time Pattern Parallelism A (temporal alignment of breaks between notes) and Time Pattern Parallelism B (duration of notes). Generally the analysis indicated that both Pitch Pattern Parallelism A and Time Pattern Parallelism B were effective in both melodies. Pitch Pattern Parallelism B was only effective in "Three Blind Mice" while Time Pattern Parallelism A was only effective in "Softly Now the Light of Day"(see Tables 5.70 and 5.71).

The implication is that individuals retain, at least over short spans of time/notes information about pitch height that contains durations. This would correspond the basic notion of a frequency contour over time. It would also seem that individuals retain categorical information about the relative durations of notes, for example, in the six note sequence:

quarter-quarter-half-quarter-quarter-eighth-eighth-quarter.

and not as a veridical representation of the times when each note ended and the next started:

|----|----|--------|---|----|--|--|----   where | is the break between notes

The categorical durations used in music and particularly the limited range of categorical duration used in simple melodies would make this an efficient way to store such information. For example, in the previous example, without any knowledge of actual durations or even the relative durations of notes, a subject simply store:

short-short-long-short-short-veryshort-veryshort-short

It must be remembered that in this work, all versions of parallelism were restricted to a somewhat simplistic level because the quantification and the experimental tests were novel. There were too many unknowns for a less restrictive exploration. How many notes can there be each subunit? How can two subunits be compared if they are not perfectly parallel? This work had to define parallelism, then quantify parallelism and then compare those predictions to empirical data. This work also had to consider parallelism within a framework of other possible causes of boundary formation.

Parallelism was constrained to occur only at locations defined by Rules 2 and 3. This first restriction seems to be in keeping with the model of Lerdahl and Jackendoff

(1983; see Chapter 4 of this work) but it might be useful to consider lifting this restriction for the second parallel unit. That is, it is possible that a listener, having heard a well delineated unit of music (properly marked by applications of Rules 2 and/or 3) might impose boundaries in a second unit that has the same initial structure but fails to fulfil the criteria for a boundary at the end (i.e., Rules 2 and 3 do not apply at the place that one would expect given the previous unit). This alternative analysis was one of the other analyses of parallelism (see Chapter 4 of the present work), but the results were too confusing to interpret. The problem was that there were simply too many unknowns in the general analysis of parallelism (e.g., the most appropriate basis for parallelism?, what constitutes parallel?, what span of notes should be used to compute parallelism?). As such, this alternative analysis produced some degree of parallelism at every point within the melody: as such, some constraints had to be imposed and the cited restriction was one of those constraints. Note that this constraint is consistent with the theory of Lerdahl and Jackendoff (1983), although the theory is not very explicit about parallelism is general. However, this point is worth further explorations.

A more important concern is that parallelism was restricted to the sequentially consecutive structures: the AA parallelism within the larger AABC structure. That is, the AA parallelism of larger structures like ABAC were not considered. This restriction was imposed because of the exploratory nature of the work. For example, how many notes can there be in B subunit within the ABAC structure before the AA parallelism is missed? This is a restriction that should be lifted as soon as possible now that there is some evidence that these methods of quantifying parallelism may have some validity.

In addition, parallelism was limited to a minimum span of three note events (defined as number of notes or as the duration of an equal number of beats). For Pitch pattern parallelism, the problem is that 2 two-note pairs always produce a quantified parallelism of $r=1.0$, or $r=-1.0$, or $r=0.0$. That is, any sequence like c-d-f-b treated as 2 two-note pairs (i.e., c-d and f-b) produces a correlation of $r=1.0$; any sequence like c-e-b-a, treated as 2 two-note pairs, produces $r=-1.0$; any sequence like c-c-f-g treated as 2 two-note pairs produces $r=0.0$. That is, if there are two pairs of notes and both pairs contain intervals other than unison, then those two pairs would be considered perfectly parallel correlations of r=1.0 or -1.0), *regardless of the magnitude of either interval.* If there are two pairs of notes, and either pair contains a unison, then the two pairs would be considered totally unrelated (correlations of r=1.0).

Similarly, in Time Pattern Parallelism B any two pairs of notes that shared a common direction of temporal change would also be perfectly parallel (correlation of r=1.0). Any two pair of notes that different in their directions of temporal change would be perfectly antiparallel correlation of r=-1.0). If either pair did not have any temporal change, then there would be no parallelism (correlation of r=0.0).

This is, in part, a limitation of the correlation measure used to assess this type of parallelism. More importantly, parallelism was limited at the other end to a maximum span of ten note events (defined as number of notes or as the duration of an equal number of beats), so longer structures could not be detected (the melodies used were 48 and 34 notes in length respectively, so parallelism longer than 10 notes might be problematic).

Finally as implied previously, a curious consequence of the correlational measure used to assess parallelism meant that flat repeating pitch patterns were not considered parallel, particularly Pitch Pattern Parallelism B (e.g., c-c-c-c-d-d-d-d would produce a correlation of r=0.0 for a two spans of 4 notes). Similarly, repeating strings of temporal patterns would not be considered parallel, particularly for Time Pattern Parallelism B (e.g., q-q-q-q-h-h-h-h would produce a correlation of r=0.0 for two spans of 4 note). As was stated earlier (see Chapter 4 of this work), this is a consequence of the measure used. The correlation is a pattern similarity measure that compares the relative change in two operand while being completely insensitive to absolute values of the operands. If there is no change in one operand, there is nothing to compare. A number of alternatives were considered that considered both absolute magnitudes (e.g., Euclidian, power) or pattern (e.g., sine) but all were rejected as less appropriate. The problem is that the correlation seems ideal (is close to ideal). That is, the correlation will find that the sequence c-d-e-f (within an octave) is more similar to the sequence g-a-b-c (within an octave) than to the sequence e-d-c-b (within an octave). However other measures like the Euclidian will find the reverse because they depend on absolute magnitudes (the correlation depends only on the pattern). Attempts to alter these other measures to reflect that which the correlation does invariably resulted, mathematically, in the correlation. A method will be needed to capture all types of parallelism. Note that this method will require some elements of pattern similarity and some elements of absolute similarity because human processing is not completely pattern based. For example, two sequences of c-d-e-f are considered more similar if placed in the same octave than if placed many octaves apart. It may be that, in the interim, the correlational analysis can be minimally adapted to treat, as a special case, the situation in which either comparison sequence has a note-to-note variance of 0.0. Of course, this still means that differences between a sequence like c-c-c-c, on the one hand, and sequences like c-d-e-f, c-e-g-C, c-d-e-e or c-e-g-e, on the other hand, would have to be carefully quantified. That is, settling for a modified correlation would still leave a lot of work to be done. In addition, such modifications would have to reflect knowledge and intuitions of music.

Efficacy of Boundaries

With respect to efficacy, the experiment can be considered a qualified success. In Experiment 1, subjects were presented with four-note extracts from the melody. Subjects had to decide whether or not each extract was in the melody. About one-half the extracts were altered (one note was changed by ±1 or ±2 semitones) such that they could not have come from the melody. Extracts were analysed as a function of the relationship between the position of extract in the melody and the position of boundaries. Most importantly, extracts that crossed a boundary were expected to produce lower performance while extracts that did not cross a boundary were expected to produce higher performance (for full details of the complex analysis, one must look to Chapter 3 of this work). This analysis was conducted for both Literal (those that had not been altered) and Foil test sequences (those that had been altered). In the melody "Three Blind Mice", performance was aligned with predictions. That is, individuals were more accurate recognizing test sequences that did not cross a boundary than they were with test sequences that did cross a boundary. This was generally true of both Literals and Foils. In the second melody, the extract from "Softly Now the Light of Day", the results were much more difficult to interpret within that simple construct, particularly for Foils. However, the second melody was more ambiguous tonally, and the second melody was more unfamiliar to subjects. It is possible that this memory task is more capable of tapping performance for well known melodies.

Generally, test sequence recognition performance could be explained, in part, by the relation between the test sequence location and boundaries. That is, the different conditions representing the relationship between boundary location and test sequence explained (significantly) some of the variance in performance, but the nature of that relationship is not clear. At two different levels, the results were similar to those obtained by Peretz (1989). As in the present work, the overall analysis by Peretz indicated that performance was higher for test sequences that did not cross boundaries (note: Peretz did not analyse Foils). However, as in the present work, more detailed analysis by Peretz yielded inexplicable complexities. Peretz actually analysed as a function of boundary type (Rule 3a or 3d or parallelism) and found that boundaries based on Rule 3a and 3d actually produced higher performance for test sequences that crossed boundaries while only boundaries based on parallelism produced lower performance for test sequences that crossed boundaries (and there was a speed-accuracy tradeoff for Rule 3a and 3d). Peretz altered the task so that the test sequence was presented before the melody, requiring subjects to monitor the melody until the test sequence was presented. This version produced results that were more consistent with expectations, but for boundaries based on Rule 3d only. Although the present work did not analyse for type of boundary, the more detailed analyses within the two studies do imply that performance in such a task is only partially explained by boundaries. Dowling also demonstrated (implicitly; see Fig. 1) that the simple "cross and do not cross" dichotomy does not capture all the differences in such test sequences.

In future work, one obvious problem to explore is that the test sequences did not form "natural" units. That is, all test sequences were four notes long, and as such, did not represent those units that the subjects had identified when parsing the melody. It would seem reasonable that one should test with the identified units. Unfortunately, the use of empirically determined units of varying lengths creates a number of theoretical difficulties. Should performance be better for longer units (more information for comparison) or should performance be better for shorter units (lower processing loads)? When crossing boundaries, does the position of the boundary within the test sequence affect performance? It seems that based on this work, that the location of the boundary within the test sequence would be a major concern (cf., Chapter 3 of this work). However, one must then consider whether or not this effect would vary with the length of the test sequence. This, in turn, raises the question of whether one measures boundary location in terms of absolute number of notes or as a percentage of sequence length. The use of variable length test sequences also creates numerous methodological difficulties.

The second area of research could focus on the role of other factors in the accurate detection of extracted sequences. That is, the present rules implied some of the variance could be explained by the boundary - sequence relationship. In addition to the boundary sequence relationship, in the present work, the roles of the location of change, the size of change and the tonality of change were examined. Each of these added some predictability, but a major portion of variance remained unexplained.

Many issues related to this boundary efficacy test were not addressed within this work for two reasons. Firstly, the boundary efficacy task was not of primary importance (as stated, the main goals were the development of an online task to determine boundary locations and the comparison of those boundary locations to a quantified version of Lerdahl & Jackendoff's [1983] theory). Secondly, there were no answers to the general questions within the literature. For example, previous studies (Dowling, 1973; Peretz, 1989; Tan et al., 1981) have used fixed length test sequences: In Tan et al., test sequences were only two notes long. Dowling's sequences were actually designed to test recognition of temporal pattern that crossed a boundary with stimuli of unknown grouping pattern (as defined by Lerdahl & Jackendoff, 1983). The probe technique advocated by Peretz would limit the data to one point per person per melody (given the time requirements of the main goals of the current work). Finally, as stated earlier, it would be much more difficult to create the programs to present variable length stimuli and to analyse the results. It is important to realize that this boundary efficacy task does not contradict the notion that boundaries are important for subsequent processing, even though this task may not provide strong support. That is, though the support may be weak, there is no interpretation of the data that would contradict the idea that parsing affects subsequent processing.

In Experiment 2, a click detection task was utilized. Subject were asked to detect, as quickly as possible, a click within the melody while listening to the melody. Generally, subjects tended to be slower for clicks that were aligned with a note that was on a boundary and faster for clicks that were aligned with a note that was not on a boundary. However the effect was not very pronounced and the effect did not follow predictions identically. More specifically, the results for the melody "Three Blind Mice" tended to follow the prediction that detection would be slowed prior to a boundary and speeded after a boundary (cf. Gregory, 1978; Sloboda & Gregory, 1980). However the effect was weak, and restricted to the opening section of the melody (implying that reaction times to clicks may be a local phenomena -- dependent on the immediate context surrounding click and not on the broader context of the whole stimulus). In the second melody, the extract from "Softly Now the Light of Day", the results were much more difficult to link to this construct. Generally, the results for the click detection task were much more difficult to relate to notion of boundaries in melody than those of the memory task. It is possible that the click detection task was too easy (hence, no systematic variance in response). However, this interpretation is blunted by the fact that performance is closer to prediction with the more familiar melody (hence easier). It seems more reasonable that the task design encouraged individuals to attend only to the clicks, effectively ignoring the music (cf., Berent & Perfetti, 1993). Subjects were more capable of this in the second unfamiliar melody, possibly because, by that point in the experiment they simply had more practice with the technique of ignoring the music. Before the last click task in Part 2 of Stage 8, they would have heard the melody "Softly Now the Light of Day" seven times. This is supported by the observation that reaction times become more consistent as the piece progressed. Again it is important to recognize that even if the support is weak, there is no interpretation of the data that could actually contradict the notion that parsing is important for subsequent processing.

Relations to Tonality

To emulate Lerdahl and Jackendoff (1983, p. 52) "The importance of [tonality] in [western-tonal] musical structure cannot be overestimated". Throughout this work, the relationship between tonality defined as the general knowledge about pitch relationships within a defined key (defined by the notions of Krumhansl and coworkers as presented in Krumhansl, 1990) and boundary formation was explored[63]. Subjects were tested in a variation of Krumhansl's probe tone task (cf., Frankland & Cohen, 1990). In the first experiment, based on a cluster analysis of their responses, most subjects were placed within one of three major groups: Triadic 1, Triadic 2 and Proximity. In the second experiment, based on a cluster analysis of their responses, most subjects were placed within one of two major groups: Triadic and Proximity. The boundary profiles of these groups of subjects were then analysed to determine whether or not there were any systematic differences in boundary profiles that could be explained by tonality group. Stated succinctly, none was found, despite the depths of the analysis[64]. More precisely,

[63] The analysis of the relationship between tonality and boundary formation required a thorough analysis of tonality so to insure that the relationship between tonality and boundary formation was not an artifact of the design. In both experiments, the analysis of tonality identified predicted subgroups that exhibited the expected relationship between training and tonality. Hence, the lack of a relationship between tonality and boundary formation was not a consequence of the lack of an effect of tonality.

[64] For this reason, only the tonality results of the first experiment were presented, though greatly abbreviated. Note that the second experiment retained tonality as an aspect of design for several reasons: perfect compatibility with first experiment, secondary questions and analyses not related to the work presented within this thesis (e.g., more detailed analysis of the link between training and tonality profiles, an exploration of the changes in the tonality profile across an experiment), the lack of a complete analysis of the first experiment before the

there were few minor differences. In the first experiment, there were also minor differences in the average responses of the three tonality groups in the boundary efficacy task (recognition memory for test sequences extracted from the melody): Subjects in the proximity group demonstrated a pattern that seemed to indicate sensitivity to size of change (and not tonality) while subjects in the tonality group seemed to demonstrate a pattern that reflected the use of information about tonality. These results were more evident in the second melody (see Chapter 3 of this work).

Although the lack of a strong tonality effect was somewhat surprising and seemingly counter-intuitive (indeed, the corpus of music which this study addresses is called Western *Tonal* Music), it is in fact, consistent with the notions of Lerdahl and Jackendoff (1983). That is, implications of tonality do not enter the model until the higher levels of the hierarchy (however, see Jackendoff, 1993; Lerdahl, 1993). Recall that subjects were asked to parse into the smallest units that made sense. If this parsing does represent lower level units, then only some of those units will have implications of tonality while others will not, hence negating any general effects. As with any study, several caveats are worth mentioning.

Firstly and most importantly, as noted by Krumhansl and coworkers (1990), within western tonal music, note durations are associated with tonality: relatively more important notes within the tonal hierarchy tend to have longer durations in a piece. Within Lerdahl and Jackendoff's (1983) work, longer notes may be associated with Rules 2b (Attack-Point) or 3d (Length Change). That is both of these rules deal with note duration. More importantly, it is Rule 2b that is associated with a single longer note (Rule 3d requires 2 longer and 2 shorter notes). Hence, the most important rule for defining boundaries is also the rule that would be associated with the more important notes of the tonal hierarchy. It is, therefore, difficult to completely separate Rule 2b from tonality. The work of Sloboda and Gregory (1980) implied that both physical changes (i.e., Rule 2b) and structural changes (i.e., tonality) were important for defining a boundary, but that

---

commencement of the second and the overriding belief that tonality *should* matter.

physical changes are more important than structural changes. It is probably best to conceive that physical (i.e, Rule 2b) and structural (i.e, tonality) relationships as existing in a boot-strapping process. That is, physical changes (akin to those defined by Lerdahl and Jackendoff), provide some idea of the importance of structural changes but in turn, structural changes reinforce which types of physical changes matter over the long run.

It is also important to remember that the rules presented herein do not represent the complete set advocated by Lerdahl and Jackendoff (1983). In particular, tonality is more important within higher level rules. As such, tonality (i.e., higher level rules) may be important for explaining why a particular instance of Rule 2b does not result in a boundary (i.e., note events 16, 21, 31 & 36 of "Three Blind Mice"). That is, tonality (and other attributes contained within the high level rules) may be critical for the elimination of false alarms.

A second caveat is that the stimuli used herein were relatively simple (minimal pitch range, minimal use of accidentals, minimal durational change, monophonic single line) compared to that which Lerdahl and Jackendoff (1983) had intended for their analysis. Given this, one might then see more effects of tonality in more complex pieces.

A third caveat is that, from an analytical point of view, the relationship between tonality and boundary was explored on the basis of boundaries created for the third repetition of each melody. Given that this study has shown a high degree of similarity between different repetitions, it might be fruitful to reanalyse the relationship on the basis of the average of last two repetitions. This might show more effects at points where boundaries were ambiguous. Also from the point of analysis, it must be remembered that most of the subjects in this experiment demonstrated a tonality profile that could be classified as triadic. Only a few subjects in each experiment were not so defined. Hence, the lack of a strong association between tonality and boundary profile may be more reflective of the homogeneity of the sample than a lack of any true relationship (the few non-triadic subjects could simply be randomly interspersed). However, it must be remembered that in both experiments there was a large range in musical training.

In the future, it might be more informative to create groups based on boundary profiles (see previous discussion of Similarity Analysis and issues of cluster analysis) and to then examine the differences between those groups to see if any can be explained by tonality. This technique would have the advantage of detecting any differences between boundary profiles (not just tonality), such as training, or new rules of parsing. The analysis in the present work was not conducted in this fashion because the cluster analysis for boundary profiles indicated no distinct subgroups.

General Considerations for Future Directions

The methodology of boundary location produced results that were consistent with theoretical notions of boundary location. As such, the main recommendation for future work is that more experiments in this line should be conducted. Particularly, empirical assessments of boundary location within a number of different pieces needs to be done. The development of a corpus of empirical data may allow those with expertise in music theory to refine their notions of grouping principles, which in turn will allow for more precision in the various empirical tests. In fact, one further experiment in this series using different stimuli and the memory recognition efficacy test has been completed, but it was not analysed in time for inclusion within this work.

With respect to modelling, the present work has provided some support for the constructs of Lerdahl and Jackendoff (1983), although there were some qualifications (e.g., the focus on Rule 2b). In addition, the present work extended the model to include the quantification of intensification and parallelism. Hence, it might be useful to point out that although the algorithms used to quantify the model were created in the programming language C, there is no reason why these same algorithms could not be written in a spreadsheet like Quattro or Excel or within a statistical package like SPSS[65]. A

---

[65] Initially, it was planned that the algorithms would be included as an appendix. However, including a minimally useful set of algorithms was found to double the length of this thesis. Interested individuals may request copies from me. Note that although the language of choice was C, I have extensive experience with SPSS and know that the algorithms could be transferred, but it would not be straightforward. I have some experience with Excel and I think that the algorithms could be transferred (the low-level analysis would be trivial, but the non-linear regression with asymmetric error terms would be difficult, if possible).

spreadsheet would easier for constructing algorithms to analysis the low-level rules, but SPSS would be easier for analysis based on non-linear regression and associated asymmetric error terms. A combination might be necessary. Thus a large body of data pertaining to the quantification of melody and the testing of Lerdahl and Jackendoff's rules could be created efficiently. More importantly, the availability of a body of quantified results on parallelism might enable theorists like Lerdahl and Jackendoff to refine their notions of parallelism, to enable the design of better methods of quantifying and testing the concept.

The current work only tested three rules. More work needs to examine the remaining rules. In particular, Rule 2a (Rest/Slur), particularly the rest aspect of Rule 2a would be expected to be very important for boundary location. A rest, within a single melody line, denotes the absence of sound. As such, it is difficult to imagine how it could fail to delineate a boundary. The slur aspect of the rule might be expected to be less important. In fact, the two aspects are so different that, as noted before, I have difficulty understanding the reason for their inclusion in a single rule (again, the works that Lerdahl and Jackendoff, 1983, considered were generally much more complex than those utilized in the present work; a rest might be more important within a simple single-line melody).

In a similar vein, the empirical data of the present work could be compared with other models that address the same issues (e.g., Boltz & Jones, 1986; Deutch & Feroe, 1981). Given the results (efficacy of Rule 2b), it would seem that these models would have equivalent predictive validity. Such an analysis was not attempted herein to allow for a detailed exploration of the one particular model that I thought provided the most comprehensive structure. For example, the rigid hierarchical model of Lerdahl and Jackendoff (1983) provides a specific context for comparing the empirical data of subjects who parsed at some unknown level of segmentation, with the theoretical predictions of the model (i.e., the analysis using asymmetric error terms and non-linear regression). The lack of such context makes it difficult to compare empirical data with theoretical predictions within other models, particularly when such models are combined with the use of naturalistic stimuli. Furthermore, it seems unreasonable to compare a detailed quantitative analysis of one model to a superficial qualitative analysis of another model. That is, if such a comparison provided differential levels of support for different models, that difference might be more reflective of the types of analysis used: All models seem to work when examined superficially (one might call this face validity: if a model lacks face validity, it is not considered further). It is only a detailed quantitative analysis that actually demonstrates the real efficacy of a model. Finally, it should be noted that detailed analyses of other models were not attempted so that there could be some focus on the issue of tonality. The irrelevancy of tonality to parsing (within the tasks developed for this thesis) was not known until after tonality had been analysed.

Comparisons between the model of Lerdahl and Jackendoff (1983) and other models in music cognition also demonstrates a particular quirk of Lerdahl and Jackendoff's model -- a quirk that is not obvious on first reading. The model of Lerdahl and Jackendoff is really two distinct models. One model deals with the low level parsing of a melody (as analysed within this thesis). This model is similar to many others in music cognition. The second model is really a music-theoretic analysis that can be applied to complex western-tonal pieces by the use of high-level rules (i.e., time-span reduction, prolongation reduction). This analysis is similar to many other in music theory. The two models are not linked. Lerdahl and Jackendoff claim that the two are linked through the intermediate-level rules (i.e., intensification, symmetry, parallelism) but they do not provide a statement of these rules. That is, they do not define intensification, symmetry or parallelism well enough to be used within the analysis of a piece, even if that analysis is only qualitative. The lack of such a definition severes the link between the low-level rules and the high-level analysis. As such, the high-level grouping structure that is the basis of their music-theoretic analysis of a piece (i.e., the high-level rules: time-span and prolongation reduction) cannot be based on the the low-level grouping structure (i.e., the low-level rules). The low-level rules identify many potential boundaries. Only a few of those will be retained as high-level boundaries. Yet, Lerdahl and Jackendoff never state how the multitude of low-level boundaries are filtered to create the high-level boundaries. *As such, the high-level grouping structure that they use as the basis of their* *music-theoretic analysis (i.e., time-span and prolongation reduction) is really only based on their intuitions of the high-level grouping structure.* West, Howell and Cross have raised the same concern, albeit in a different form: "An unfortunate facet of the model . . . is that, despite having a formal output, the structural descriptions must be derived intuitively and, therefore, depend on the analytic capabilities and propensities of the model builder." (1985, p. 39). The lack of a link between the low-level rules and the high-level analysis is further evidenced in the examples chosen to illustrate the model. Examples that pertain to the low-level rules generally used simple monophonic melodies. However, examples that pertain to the high-level rules generally used harmonic, polyphonic or contrapuntal works. It is difficult to see how the low-level rules have been extended to such complex pieces (e.g., should the low-level rules be applied exclusively to the upper voice?, exclusively to the bass line? to the upper voice and bass, but separately? to some weighted average of all voices?, to the implied roots of chords?). The anonymous fifteenth-century instrumental piece, *Dit le Bourrguignon*, (Example 3.37, 1983, p. 66) is the only example to indicate the application of low-level rules in a complex piece, and yet, this example only indicates a few of the potential applications of low-level rules without qualifying any of the exclusions. To make the theory as comprehensive as they claim it to be, Lerdahl and Jackendoff complete the link between the low level rules and the high-level analysis in a manner that makes clear how the units defined at the lowest levels are combined to create the grouping structure that is the basis of the high-level analysis. This link must be made for both monophonic and harmonic, polyphonic or contrapuntal (or even, polytonal) pieces. That is, Lerdahl and Jackendoff must define intensification, symmetry and parallelism and Lerdahl and Jackendoff must demostrate the use of the low level rules in more complex pieces.

When or while filling the gaps in the model, it might also be fruitful to consider modifications and extensions to the rules. Most importantly, Lerdahl and Jackendoff must define the ladder for pitch relations. That is, should the low-level rules be based on absolute frequency, log-scaled frequency, the chromatic scale, the diatonic scale or tonality. If one of the more perceptual scales is chosen (i.e., diatonicism or tonality), then all steps between all possible intervals must be rigorously defined. For example, if pitch relationships are defined on a diatonic scale, then how should non-diatonic tones be represented? Furthermore, the use of a diatonic or tonal scale presupposes that listeners are processing events relative to some internal notion of key. How do listeners abstract tonality? How do listeners detect key changes? It should be noted that there are answers to many of these questions in the literature (Cohen, 1991; Krumhansl, 1990), but those answers must be eventually quantified within a single comprehensive framework. Further refinements could use the results of this thesis (even though these results are based on only two melodies). For example, Lerdahl and Jackendoff could also use the analysis of Intensification (Rule 4) to refine their model, perhaps to argue that in the case of conflict between rules, Rule 2b (Attack-Point) would take precedence. Lerdahl and Jackendoff could build upon the analysis of Parallelism (Rule 6), adding their intuitions of parallelism in music, to create a quantifiable version. Finally, Lerdahl and Jackendoff must carefully consider what is meant by Symmetry (Rule 5) for listeners (i.e., how can a system that is processing events on-line utilize the concept of symmetry): The same concern can be raised for Rule 1 (avoid groups containing a single note).

In the initial stages of this work, Lerdahl and Jackendoff's model was not the only approach to the analysis of parsing that was considered (see Chapter 1 of this work). In fact, an alternative model to that of Lerdahl and Jackendoff (1983) was developed. Development of this model was discontinued to allow for a greater focus on Lerdahl and Jackendoff's model[66]. In the model, boundaries were assessed on the basis of change only. That is, the currently sounding note was compared to previously sounded notes on a number of dimensions. Many of these dimensions were the same as Lerdahl and

---

[66] This decision was motivated, in part, by the fact that when developing a model one must also defend the constructs of the model by demonstrating its links to the empirical literature and other models that address the same issues. For example, Lerdahl and Jackendoff devote substantial space to contrasts and comparisons between their model and those of Meyer, Schenker and Narmour, to mention only a few. Such development would have detracted from the empirical assessment of boundaries.

Jackendoff (pitch height, duration, intensity, timbre). In many ways, the alternative model was similar to Lerdahl and Jackendoff (another factor in the decision to focus on Lerdahl and Jackendoff). However, the critical difference was that the currently sounding note was compared to only previous notes. Compared with Lerdahl and Jackendoff's model, this should minimize the cognitive load, because there was no retrospective boundary placement. In Lerdahl and Jackendoff, the boundary is placed between Notes 2 and 3 of a four note span. Hence, in the scheme of Lerdahl and Jackendoff, the listener would have to process the four note span, deciding whether or not to place a boundary one note back in time, while tracking any new notes that occurred during the processing. The alternative model reduced that load by one note event; that is, the listener would be expected to remember one less note while trying to process the previous notes (recall that there is a short term limit of 7±2). A second critical difference was that the alternative model also examined tonality as a low-level rule. That is, using the results of Cohen (1991) and Krumhansl (1990), it was assumed that the listener had an internal representation of tonality and that the currently sounding note was placed/heard for its function within the hierarchy of tonality (i.e., as a tonic, dominant, leading tone, etc.) and that the role of the currently sounding note was compared to previously sounded notes. This allows one to track changes in the tonality of the individual notes of a sequence, to "know" when one has arrived at the dominant or tonic. Finally, the other major difference was that in the ongoing analysis of the currently sounding note, difference spans of previously sounded notes could be used (note that Lerdahl and Jackendoff used a fixed four-note window). This variable span could be short for listeners lacking experience with music and long for listeners with extensive experience with music. In the initial analysis of empirical data (using correlational analysis only), this alternative model performed as well as that of Lerdahl and Jackendoff.

This alternative model does not preclude any of the higher level analysis of Lerdahl and Jackendoff (1983). The alternative model also lends itself to the analysis of symmetry (see Chapter 4 of this work). The nice thing about this alternative model is the manner in which it dovetails with the work of Jones and colleagues (cf., Boltz & Jones, 1986; Jones, 1976a, 1976b 1981b, 1982, 1993; Jones & Boltz, 1989). It was in fact, based in their constructs. In the alternative model, the previous span of sounded notes (which may vary in length as a function of training) determined the interpretation of the currently sounding note and simultaneously, expectancies for future notes.

A second major area of work is the issue of parallelism. Notions of parallelism must be rendered algorithmic if empirical tests are to be conducted. Such notions must clearly identify and operationally define the underlying constructs (i.e., contour, pitch height, duration, harmonic motion, while avoiding the use of vague, or overused, terms like rhythm). Any precursors to the perception of parallelism must also be defined. For example, one cannot discuss parallelism due to harmonic motion in the absence of some notion of how (or if) key was defined. The harmonic parallelism that might be detected could depend on whether one presupposed "key" defined by a tonal hierarchy (as in Krumhansl, 1990) or by the intervallic rivalry hypothesis(as in Brown, 1988; Brown, Butler, & Jones, 1994; Butler, 1989, van Egmond & Butler, 1996). An attempt was made in the current work to both define and test parallelism. Although successful, these tests must be considered as preliminary (see previous discussions, including Chapter 4 of the current work). In fact, personally, I view the methods developed in this work more as an example of how one might assess parallelism and not so much as the definitive work defining parallelism. It is clear that parallelism must be defined on a multiplicity of dimensions: Two basic dimensions of pitch and duration were identified and explored in this work, but there are many other candidates, as well as possible interactions (i.e., one definition of pitch pattern parallelism actually involved durations). To this end, retrospective classification studies of the structures typically implicated in music (cf., Huron, 1996; Vos & Troost, 1989) might be most useful.

For the future, it is advocated that the boundary location methods used herein be replicated with a large base of material so that theory building can rest on a large base of empirical data. In addition, as knowledge accrues, refinements to the basic procedures described herein can be used to gather more knowledge. Subjects could be encouraged to parse at different levels of the hierarchy. This may lead to further confirmations of

Lerdahl and Jackendoff's model.

The current tests of efficacy should be refined and perhaps different tests of boundary efficacy can be developed. Performance may be explained by other factors not examined within the experiments of the current work. For example, the complexity of the melody as a whole, the complexity of the section of the melody from which the extract was taken (or the complexity of the extract) or the complexity of the section in which the clicked placed might be implicated. Though complexity is not an easy construct to define within music, complexity has been implicated in many music cognition tasks (cf., Berent & Perfetti, 1985; Boltz & Jones, 1986; Cohen, Trehub & Thorpe, 1989; Deliege, 1996; Burbridge & Jones, 1982; Jones, Maser & Kidd, 1978; Jones, Sommerel & Marshburn, 1987; Palmer & van de Sande, 1995). In the current work, both boundary efficacy tasks were utilized within real-world stimuli, using an on-line task. Thus, there is no doubt that many other factors may play causal roles in performance. Even though neither efficacy test could be argued to be a perfect demonstration of the effect of boundaries on music, their continued use and development is advocated. In both tasks some of the variance in responding (accuracy with test sequences; reaction times to clicks) could be explained by boundary location. That is, both tasks do demonstrate some systematic differential responding and as such both tests demonstrate some promise (the recognition memory perhaps more so).

As an alternative, the probe technique of Peretz (1989) seems promising though it does limit the amount of data that can be collected (i.e., only one datum per presentation of a melody). Another alternative is a procedure developed by Deliege and coworkers (Deliege, Melen, Stammers & Cross, 1994; Cross, Melen, Stammers, Deliege, 1996). In this procedure, the melody is segmented into a number of sequences. In the cited works the segmentation was determined by the Group Preference Rules of Lerdahl and Jackendoff (1983), but they could be determined from empirically delineated boundaries. These segments are presented to subjects in random pairings. The subject must identify the correct orderings of the segments. Note that this method has some similarity to the boundary studies of Newtson and coworkers (Newtson 1973, 1976; Hanson, & Hirst, 1989; Massad, Hubbard & Newtson, 1979; Newtson & Engquist, 1976; Newtson, Engquist & Bois 1977; Newtson, Hairfield, Bloomingdale & Cutino, 1987) in the area of social psychology. Regardless of the boundary efficacy test that is used, it seems clear that to fully interpret the results of a boundary efficacy tests, the notions of units, boundaries and musical complexity and their role in melody processing must be further developed. Indeed, it is arguable that complete understanding of the role of units and boundaries in music might have to wait until there is a better understanding (perhaps consensus is a better word) of meaning in music (cf., Collier, 1996; Gregory, 1996; HaCohen & Wagner, 1996; Watt & Ash, 1996; see also Chapter 1 of the current work).

Summary

The present work has demonstrated an on-line task for the empirical assessment of parsing. The present work has also modelled the empirical data by quantifying the predictions of Lerdahl and Jackendoff (1983). The present work extended the notions of Lerdahl and Jackendoff to the analysis of symmetry and parallelism. Finally, the present work has also added some information pertaining to the utility of the memory recognition and click detection boundary-efficacy tasks.

Generally, the results of the current work represent a successful integration of empirical data with a complex, holistic theory of music perception.

Appendix: Instructions

Experiment 1: Stage 1

In the following phase of the experiment, you will hear a number of sequences.

Each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the "fit" of the single note to the preceding eight notes. Feel free to use any critera that you like as the basis for your assessment: There is no right or wrong answer, but try to make the judgement as natural or "instinctive" as possible. At first this may seem awkward, but eventually it becomes routine.

When making your judgement, try to use the full range of a scale from one to three with a one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

The use of three fingers (of either hand) on the number keys (top row of the keyboard) or the keypad is the easiest.

Finally, do not worry if you should change your mind: This task is only practice and it will be repeated in Phases 3, 5 and 7.

Experiment 1: Stage 2

A song can be considered as a sequence (or series) of musical motives. phrases, themes or musical ideas. As such, music is analogous to speech which consists of a sequence of words, phrases, clauses, sentences, or semantic ideas. Any song can be parsed into its component musical ideas, and a break-point is the end of one musical idea, and the beginning of the next musical idea.

In the first part of this phase of the experiment, you will hear a short song or an extract of a song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. At first this may seem a little contrived, but simply relax and indicate your choice at the level that seems most reasonable, or natural (or comfortable) to you. When parsing, try to parse at the level of the motive (i.e., akin to the word in a sentence), which is usually between 2 and 10 notes.

IT IS IMPORTANT TO REALIZE THAT THIS IS MUSIC. AND MUSIC IS VERY PERSONAL: THERE IS NO RIGHT OR WRONG CHOICE.

The song will be repeated three times, but the task is the same for each repetition: Use the repetitions to refine your responses.

The second part of this phase of this experiment occurs after the third repetition. You will hear 16 four-note subsequences. You will decide whether or not each of those subsequences was a part of the song you had just parsed: Some were, some were not.

Experiment 1: Stage 3

As in Phase 1, you will hear a number of sequences and each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes. Feel free to use any critera that you like but try to make the judgement as natural or "instinctive" as possible: Again, there is no right or wrong answer.

When making your judgement, try to use the full range of a scale from one to three with a one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

The use of three fingers (of either hand) on the number keys (top row of the keyboard) or the keypad is the easiest.

In this phase, there will be 2 runs through the 26 sequences.

Again, do not worry if you should change your mind: The task will be repeated in Phases 5 and 7.

Experiment 1: Stage 4

This phase is the same as Phase 2. Just to remind you, a song, like speech, can be considered to consist of a sequence of motives, phrases, themes or musical ideas. Any song can be parsed (i.e., divided) into its component musical ideas, and a break-point is the end of one musical idea, and the beginning of the next musical idea.

In the first part of this phase of the experiment, you will hear a short song or an extract of a song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. This may still seem a little contrived, but simply relax and indicate your choice at the level that seems most reasonable, or natural (or comfortable) to you. Again, it is important to realize that there can be no right or wrong choice.

The song will be repeated three times, but the task is the same for each repetition: Use the repetitions to refine your responses.

As in Phase 2, the second part of this phase of this experiment occurs after the third repetition. You will hear 16 four-note subsequences. You will decide whether or not each of those subsequences was a part of the song you had just parsed: Some were, some were not.

Experiment 1: Stage 5

As in Phases 1 and 3, you will hear a set of sequences. Each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes. Feel free to use any critera that you like as the basis for your assessment: Try to make the judgement as natural or "instinctive" as possible.

Again the scale is the range from one to three, with one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

In this phase, as in Phase 3, there will be 2 runs through the set of 26 sequences.

Experiment 1: Stage 6

As was stated previously, a song can be considered as a sequence (or series) of musical motives, phrases, themes or musical ideas. As such, it is possible to parse any song into its component musical ideas. A break-point is the end of one musical idea, and the beginning of the next musical idea.

As was done previously in Phase 2 and 4, in the first part of this phase, you will hear a short song or an extract of a song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. This may still seem a little contrived, but once again, relax and indicate your choice at the level that seems most reasonable, or natural (or comfortable) to you. It is important to realize that there can be no right or wrong choice.

Again, the song will be repeated three times, but the task is the same for each repetition: As before, the repetitions are for you to refine your responses.

The second part of this phase of the experiment is the same as before. After the third repetition, you will hear 16 four-note subsequences. You will decide whether or not each of those subsequences was a part of the song you had just parsed: Again, some were, some were not.

Experiment 1: Stage 7

This is the final run through the set of sequences. Once again, as in Phases 1, 3 and 5, you will hear a set of sequences and each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes, using any criteria that you like, using the one to three scale, with one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

As in Phases 3 and 5, there will be 2 runs through the set of 26 sequences.

Experiment 1: Stage 8

This is the final phase of this experiment. Once again, it is the same as the previous Phases 2, 4 and 6. It is necessary to parse a song into its component phrases, themes or musical ideas.

As before, in the first part of this phase, you will hear a short song or an extract of a song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. Please indicate your choice at the level that seems most reasonable, or natural (or comfortable) to you.

Again, the song will be repeated three times, but the task is the same for each repetition. Undoubtedly, you will notice that the song is the same as that of Phase 6. My apologies for the boredom, but the intent is to allow you further time to refine your responses.

The second part of this phase of the experiment is the same as before. After the third repetition, you will hear 16 four-note subsequences. You will decide whether or not each of those subsequences was a part of the song you had just parsed: Again, some were, some were not.

Experiment 2: Stage 1

General Instructions: Phases 1, 3, 5 and 7

(The instructions are repeated so there is no need to memorize this.)

In the following phase of the experiment, you will hear a number of sequences. Each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the "fit" of the single note to the preceding eight notes; that is, judge the fit of the last note to the preceding 8 notes.

Feel free to use any critera that you like as the basis for your assessment: There is no right or wrong answer, but try make the judgement as natural or "instinctive" as possible. At first this may seem awkward, but eventually it becomes routine.

When making your judgement, try to use the full range of a scale from one to three with a one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

If you are left handed, then the use of three fingers on the number keys (top row of the keyboard -- left side) is the easiest. However, you must remember to hold your arm off the keyboard. If you are right handed, the use of three fingers on the right-hand keypad is the easiest. You may also use the three arrow keys, but it will not seem to record your responses as numbers even though it does.

Finally, do not worry if you should change your mind: This task is only practice and it will be repeated in Phases 3, 5 and 7.

Experiment 2: Stage 2

General Instructions: Phases 2, 4, 6 and 8

(These instructions are repeated so there is no need to memorize this.)

A song can be considered as a sequence (or series) of musical motives, phrases, themes or musical ideas. As such, music is analogous to speech which consists of a sequence of words, phrases, clauses, sentences, or semantic ideas. Any song can be parsed (i.e., divided) into its component musical ideas, and a break-point is the end of one musical idea, and the beginning of the next musical idea.

Part 1: In the first part of this phase of the experiment, you will hear a short song. All you have to do is hit a key each time you feel that the music has reached a break-point. At first this may seem a little contrived, but simply relax: It quickly becomes routine. When parsing, try to parse at the level of the motive -- the smallest unit that makes sense to you. The motive in music is analogous to the word in a sentence, and it is usually considered to be between 2 and 10 notes. When responsing, you will likely find it easiest to use the spacebar to indicate your choices.

IT IS IMPORTANT TO REALIZE THAT THIS IS MUSIC, AND MUSIC IS VERY PERSONAL: THERE IS NO RIGHT OR WRONG CHOICE.

The song will be repeated three times, but the task is the same for each repetition: Use the repetitions to refine your responses.

Part 2: The second part of this phase of this experiment occurs after the third repetition of the song. In the second part, you will get a fourth repetition of the same song, but your task is different. In this repetition, you will listen for a "click" and any time you hear a "click" you will press the space bar (or any other key) as quickly as possible.

Experiment 2: Stage 3

General Instructions: Phase 3

As in Phase 1, you will hear a number of sequences and each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes. Feel free to use any critera that you like but try to make the judgement as natural or "instinctive" as possible: Again, there is no right or wrong answer.

When making your judgement, try to use the full range of a scale from one to three with a one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

The use of three fingers (of either hand) on the number keys (top row of the keyboard) or on the keypad or on the arrow keys is possible.

In this phase, there will be 2 runs through the 26 sequences. Again, do not worry if you should change your mind: The task will be repeated in Phases 5 and 7.

Experiment 2: Stage 4

General Instructions: Phase 4

This phase is the same as Phase 2. Just to remind you, a song, like speech, can be considered to consist of a sequence of motives, phrases, themes or musical ideas. Any song can be parsed (i.e., divided) into its component musical ideas, and a break-point is the end of one musical idea, and the beginning of the next musical idea.

In the first part of this phase of the experiment, you will hear a short song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. This may still seem a little contrived, but simply relax: Please indicate your choice at the lowest level (i.e., smallest units) that make sense to you. Again, please remember that there can be no right or wrong choice.

The song will be repeated three times, but the task is the same for each repetition: Use the repetitions to refine your responses.

As in Phase 2, the second part of this phase of this experiment occurs after the third repetition. Once again, you will hear a fourth repetition of the song, but at this time you will listen for "clicks". Any time you hear a click, hit the space bar as rapidly as possible.

Experiment 2: Stage 5

General Instructions: Phase 5

As in Phases 1 and 3, you will hear a set of sequences. Each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes. Feel free to use any critera that you like as the basis for your assessment: Try to make the judgement as natural or "instinctive" as possible.

Again the scale is the range from one to three, with one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

In this phase, as in Phase 3, there will be 2 runs through the set of 26 sequences.

Experiment 2: Stage 6

General Instructions: Phase 6

As was stated previously, a song can be considered as a sequence (or series) of musical motives, phrases, themes or musical ideas. As such, it is possible to parse any song into its component musical ideas. A break-point is the end of one musical idea, and the beginning of the next musical idea.

As was done previously in Phases 2 and 4, in the first part of this phase, you will hear a short song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. This may still seem a little contrived, but once again, relax and indicate your choice at the lowest level (i.e., smallest units) that makes sense to you. It is important to realize that there can be no right or wrong choice.

Again, the song will be repeated three times, but the task is the same for each repetition: The repetitions are so that you may refine your responses.

The second part of this phase of the experiment is the same as before. After the third repetition, you will a fourth repetition of the same song. Any time you hear a "click" hit the space bar as rapidly as possible.

Experiment 2: Stage 7

General Instructions: Phase 7

This is the final run through the set of sequences. Once again, as in Phases 1, 3 and 5, you will hear a set of sequences and each sequence will consist of eight notes, followed by a short pause, and then a single note. Please judge the fit of the single note to the preceding eight notes, using any criteria that you like, using the one to three scale, with one representing a good fit, and a three representing a poor fit:

good fit 1 2 3 poor fit

As in Phases 3 and 5, there will be 2 runs through the set of 26 sequences.

Experiment 2: Stage 8

General Instructions: Phase 8

This is the final phase of this experiment. Once again, it is the same as the previous Phases 2, 4 and 6. It is necessary to parse a song into its component phrases, themes or musical ideas.

As before, in the first part of this phase, you will hear a short song. All you have to do is hit a key (any key) each time you feel that the music has reached a break-point. Please indicate your choice at the lowest level (i.e., smallest units) that seems reasonable to you.

Again, the song will be repeated three times, but the task is the same for each repetition. Undoubtedly, you will notice that the song is the same as that of Phase 6. My apologies for the boredom, but the intent is to allow you further time to refine your responses.

The second part of this phase of the experiment is the same as before. After the third repetition, you will a fourth repetition of the song. Any time you hear a "click", hit the space bar as rapidly as possible.

References

Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science, 8 (1)*, 12-15

Adorno, T. W. (1990). Music, language and composition 9(S. Gillespie, Trans.). *The Musical Quarterly*, *77*, 401-414.

Agmon, E. (1990). Music theory as a cognitive science: Some conceptual and methodological issues. *Music Perception*, *7*, 285-308.

Allen, L. G. (1979). The perception of time. *Perception and Psychophysics, 26 (5)*, 340-354.

Anderson, J. R. (1990). *Cognitive Psychology and Its Implications*. New York: W. H. Freeman.

Anders, G., Diamond, S., Euverard, E., Geng, V., Grossman, P., Rule, S., Tebbel, J. R. & Williams, P. (1975). *Listening to Music*. Toronto: Holt Rinehart and Winston.

Bartett, J. C. (1993). Tonal structure of melodies. In T. J. Tighe and W. J. Dowling (Eds.) *Psychology and Music* (pp. 39-62). Hillsdale, NJ: Lawrence Erlbaum Associates

Bartlett, J. C. & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6 (3)*, 501-515.

Berent, I. & Perfetti, C. A. (1993). An on-line method in studying music parsing. *Cognition, 46*, 203-222.

Besson, M. & Macar, F. (1987). An event-related potential analysis of incongruity in music and othe non-linguistic contexts. *Psychophysiology, 24 (1)*, 14-25.

Bevington, P. R. & Robinson, D. K. (1992). *Data Reduction and Error Analysis for the Physical Sciences*, Toronto: McGraw Hill.

Bharucha, J. J. (1987a). MUSACT: A connectionist model of musical harmony. *Proceedings of the Cognitive Science Society*. Hillsdale NJ: Erlbaum.

Bharucha, J. J. (1987b). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5 (1)*, 1-30.

Bharucah, J. J. (1992). Tonality and learnability. In M. R. Jones & S. Holleran (Eds,), *Cognitive Basis of Musical Communication* (pp. 213-224). Washington, DC: American Psychological Association.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13 (3)*, 383-400.

Boltz, M. (1989). Perceiving the end: Effects of tonal relationships on melodic completion, *Journal of Experimental Psychology: Human Perception and Performance, 15 (4)*, 749-761.

Boltz, M. (1991a). Some structural determinants of melody recall. *Memory& Cognition, 19*, 239-251.

Boltz, M. (1991b). Time estimation and the attentional perspective. *Perception & Psychophysics, 49 (5)*, 422-433.

Boltz, M. & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology, 18*, 389-431.

Boltz, M., Marshburn, E., Jones, M. R. & Johnson, W. W. (1985). Serial-pattern structure and temporal-order recognition. *Perception & Psychophysics. 37 (3)*, 209--217.

Boretz, B. A. (1989). The logic of what? *Journal of Music Theory, 33.1*, 107-116.

Brown, H. (1988) The interplay of set content and temporal context in a functional theory of tonality perception. *Music Perception, 5 (3)*, 219-250.

Brown, H. Butler, D. & Jones, M. R. (1994). Musical and temporal influences on key discovery. *Music Perception, 11 (4)*, 371-407.

Brown, M. & Dempster, D. J. (1989). The scientific image of music theory. *Journal of Music Theory, 33.1*, 65-106.

Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception, 6 (3)*, 219-242.

Burbridge, S. & Jones, M. R. (1982). The function of rule-governed serial contexts in recognition of embedded tones. *Psychomusicology, 2 (2)*, 47-63.

Carterette, E. C., Kohl, D. V. & Pitt, M. A. (1986) Similarities among transformed melodies: The abstraction of invariants. *Music Perception, 3*, 393-410.

Christ, W., DeLone, R., Kliewer, V., Rowell, L. & Thomson, W. (1966). *The Materials and Structure of Music, I.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Christ, W., DeLone, R. & Winold, A. (1975). *Involvement with Music.* New York: Harper's College Press.

Clarke, E. F. & Krumhansl, C. L. (1990). Perceiving musical time, *Music Perception, 7 (3)*, 213-251.

Clark, H. H & Clark, E. V. (1977). *Psychology and Language: An Introduction to Psycholinguistics.* Toronto, On: Harcourt, Brace, Jovanovich.

Cohen, A. J. (1991). Tonality and perception: Musical scales primed by excerpts from the *Well Tempered Clavier* of J. S. Bach. *Psychological Research, 28*, 255-270.

Cohen, A. J. & Frankland, B. (1990). Scale and serial order information in melodic perception: Independence or interdependence. *Canadian Acoustics, 18*, 3-10.

Cohen, A. J., Trehub, S. E. & Thorpe, L. A. (1989). Effects of uncertainty on melodic information processing. *Perception & Psychophysics, 46 (1)*, 18-28.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science, 1 (3)*, 98-105.

Cohen, J. & Cohen, P. (1982). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Collier, G. L. (1996). Towards a model of music and emotion. *4th International Conference on Music Perception and Cognition: Proceedings*, 161-166.

Cook, N. (1989). Music theory and "good comparison": A viennese perspective. *Journal of Music Theory, 33.1*, 117-141.

Cooper, W. H. & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology, 2*, 179-184.

Croonen, W. L. M. & Kop, P. F. M. (1989). Tonality, tonal scheme, and contour in delayed recognition of tone sequences. *Music Perception, 7*, 49-68.

Cross, I. (1985). Music and change: On the establishment of rules. In P. Howell, I. Cross & R. West (Eds) *Musical Structure and Cognition.* (pp. 1-20) London: Academic Press.

Cross, I., Melen, M. M., Stammers, D. & Deliege, I. (1996). The cognitive representation of tonal musical structure. *4th International Conference on Music Perception and Cognition: Proceedings*, 59-63.

Cuddy, L. L. (1993). Melody comprehesion and tonal structure. In T. J. Tighe and W. J. Dowling (Eds.) *Psychology and Music* (pp. 19-38). Hillsdale, NJ: Lawrence Erlbaum Associates

Cuddy, L. L. & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28*, 255-270.

Cuddy, L. L., Cohen, A. J. & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology,: Human Perception and Performance, 7*, 869-883.

Cuddy, L. L., Cohen, A. J. & Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology, 33*, 148-157.

Cuddy, L. L. & Lunney, C. A. (1995). Expectancies generated by intervals: Perceptual judgements of melodic continuity. *Perception and Psychophysics, 57*, 451-462.

Cuddy, L. L. & Lyons, H. I. (1981). Musical pattern recognition: A comparison of listening to and studying tonal structures and tonal ambiguities. *Psychomusicology, 1*, 15-33.

Daniels, A. J. and Wagner, L. (1975) *Music*. NY: Holt, Rhinehart & Winston.

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's Grouping Preference Rules. *Music Perception, 4*, 325-360.

Deliege, I. Melen, M. Stammers, D. & Cross (1996). Musical schemata in real-time listening to a piece of music. *Music Perception. 14 (2)*, 117-160

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde and retrograde inversion. *Perception and Psychophysics, 12 (5)*, 417-421.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception and Psychophysics, 14 (1)*, 37-40.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85 (4)*, 341-354.

Dowling, W. J. (1982). Contour in context: Comments on Edworthy. *Psychomusicology, 2*, 47-48.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50 (4)*, 305-313.

Dowling, W. J. (1993). Procedural and declarative knowledge in music cognition and education. In T. J. Tighe and W. J. Dowling (Eds.) *Psychology and Music* (pp. 5-18). Hillsdale, NJ: Lawrence Erlbaum Associates

Dowling, W. J. & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1*, 30-49.

Dowling, W. J. & Fujitani D. S. (1971). Contour, Interval and Pitch Recognition in Memory for Melodies. *Journal of the Acoustical Society of America, 49 (2)*, 524-531.

Edworthy, J. (1985). Interval and contour in melody processing. *Music Perception, 2*, 375-388.

Edworthy, J. (1982). Pitch and contour in music processing. *Psychomusicology, 2*, 44-46.

Estes, W. K. (1997). Significance testing in psychological research: Some persisting issues. *Psychological Science, 8 (1)*, 18-21.

Frankland, B. & Cohen, A. J. (1990). Expectancy profiles generated by major scales: Group differences in ratings and reaction times. *Psychomusicology, 9*, 173-192.

Frankland, B. W. & Cohen, A. J. (1996). Using the Krumhansl and Schmuckler key-finding algorithm to quantify the effects of tonality in the interpolated-tone pitch-comparison task. *Music Perception, 14 (1)*, 57-83.

Garrett, M. F., Bever, T. G. & Fodor, J. (1966). The active use of grammar in speech perception. *Perception and Psychophysics, 1*, 30-32.

Goldstein, E. B. (1984). *Sensation and Perception*, Belmont CA: Wadsworth Publishing Company.

Green, D. M. (1965). *Form in Tonal Music: An Introduction to Analysis*. Toronto: Holt, Rinehart and Winston Inc.

Gregory, A. H. (1978). Perception of clicks in music. *Perception & Psychophysics, 2*, 171-174.

Gregory, A. H. (1996). Cross-cultural differences in perceiving the emotional content of music. *4th International Conference on Music Perception and Cognition: Proceedings*, 161-166.

Haberlandt, K. (1994). *Cognitive Psychology*. Toronto: Allyn and Bacon.

HaCohen, R. & Wagner, N. (1996). Analysis of musical meaning using the "Semantic Integral Method": The case of Wagner's Leitmotifs. *4th International Conference on Music Perception and Cognition: Proceedings*, 155-160.

Handel, S. (1993) *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Hanson, C. & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General, 118 (2)*, 136-147.

Harris, R. J. (1997). Significance tests have their place. *Psychological Science, 8 (1)*, 8-11.

Hayes, W. L. (1995). *Statistics*. Toronto: Harcourt Brace Jovanovitch College Publishers.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30*, 47-61.

Howell, D. C. (1992). *Statistical Methods in Psychology*, Belmont: Duxbury Press.

Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science, 8 (1)*, 3-7.

Huron, D. (1996). On the kinematics of melodic contour: Deceleration, declination, and arch-trajectories in vocal phrases. *4th International Conference on Music Perception and Cognition: Proceedings*, 133-138.

Idson, W. L. & Massaro, D. W. (1978). A bidirectional model of pitch in the recognition of melodies. *Perception and Psychophysics, 24 (6)*, 551-565.

Jackendoff, R. (1992). Musical processing and musical affect, In M. R. Jones & S. Holleran (Eds.), *Cognitive Basis of Musical Communication* (pp. 51-68). Washington, DC: American Psychological Association.

Jonas, O. (1982). *Introduction to the Theory of Heinrich Schenker*. (John Rothgeb, Trans). London: Longman.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*, 323-355.

Jones, M. R. (1976b). Levels of structure in the resconstruction of temporal and spatial serial patterns. *Journal of Experimental Psychology: Human Learning and Memory, 2 (4)*, 475-488.

Jones, M. R. (1981a). A tutorial on some issues and methods in serial pattern research. *Perception and Psychophysics, 30 (5)*, 492-504.

Jones, M. R. (1981b). Music as a stimulus for psychological motion: Part 1, some determinants of expectancies. Psychomusicology, 1 (2), 34-51.

Jones, M. R. (1982). Music as a stimulus for psychological motion: Part 2, an expectancy model. Psychomusicology, 2 (1), 1-13.

Jones, M. R. (1992). Attending to musical events. In M. R. Jones & S. Holleran (Eds,). *Cognitive Basis of Musical Communication* (pp. 91-115). Washington, DC: American Psychological Association.

Jones. M. R. & Boltz, M. (1989). Dynamic attending and response to time. *Psychological Review, 96 (3)*, 459-451.

Jones, M. R. & Holleran, S. (1992). Cognitive bases of musical communication: An overview. In M. R. Jones & S. Holleran (Eds,). *Cognitive Basis of Musical Communication* (pp. 1-6). Washington, DC: American Psychological Association.

Jones, M. R., Maser, D. J. & Kidd, G. R. (1978). Rate and structure in memory for auditory patterns. *Memory and Cognition, 6 (3)*, 246-258.

Jones, M. R., Summerell, L. & Marchburn, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology, 39A*, 89-121.

Jones, M. R. & Raltson, J. T. (1991). Some influences of accent structure on melody recognition. *Memory and Cognition, 19*, 8-20.

Jusczyk, P. W. & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology Human Perception and Performance, 19 (3)*, 627-640.

Katz, B. (1995). Harmonic resolution, neural resonance, and positive affect. *Music Perception, 13 (1)*, 79-108.

Kallman, H. J. & Massaro, D. W. (1979). Tone chroma is functional in melody recognition. *Perception and Psychophysics, 26 (1)*, 32-36.

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception, 4*, 185-214.

Kerlinger, F. N. (1986). *Foundations of Behavioural Research.* Toronto: Harcourt Brace Jovanovitch College Publishers.

Kidd, G. Boltz, M. & Jones, M. R. (1984). Some effects of rhythmic context on melody recognition. *American Journal of Psychology, 2*, 153-173.

Klein, M., Coles, M. G. H. & Donchin, E. (1984). People with absolute pitch process tones without producing a P300. *Science, 223*, 1306-1309.

Kraut, R. (1992). On the possibility of a deteminate semantics for music. In M. R. Jones & S. Holleran (Eds.), *Cognitive Basis of Musical Communication* (pp. 11-22) Washington, DC: American Psychological Association.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*, 346-374.

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension and musical ideas. *Music Perception, 13 (3)*. 401-432.

Krumhansl, C. L. & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1 (1)*, 70-73.

Krumhansl, C. L. & Kiel, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory and Cognition, 10 (3)*, 243-251.

Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 579-594.

Konovalov, V. F.; Otmakhova, N. A. (1984) EEG manifestations of functional asymmetry of the human cerebral cortex during perception of words and music.

[Abstract]
Human Physiology, 9 (4), 250-255.

Kuhn, R. (1978). Music as representational art. *British Journal of Asthetics, 18,* 120-125.

Lantz, M. E. & Cuddy, L. L. (1996). The effects of surface cues in the perception of pitch structure: Frequency of occurrence and duration. *4th International Conference on Music Perception and Cognition: Proceedings,* 281-286.

Lantz, M. E. & Cuddy, L. L. (1998). Total and relative durations in the internal representation of surface structure in music. *Eighth Annual Meeting of the Canadian Society for Brain, Behaviour and Cognitive Science*: Ottawa.

Lee, C. S. (1985). The rhythmic interpretation of simple musical sequences: Towards a perceptual model. In P. Howell, I. Cross & R. West (Eds) *Musical Structure and Cognition.* (pp. 53-70) London: Academic Press.

Lerdahl, F. (1988a). Tonal pitch space. *Music Perception, 5 (3),* 315-350.

Lerdahl, F. (1988b). Cognitive constraints on compositional system. In J. A. Sloboda (Ed.), *Generative Processes in Music* (pp. 231-259). Oxford: Clarendon Press

Lerdahl, F. (1992). Pitch space journeys in two Chopin preludes. In M. R. Jones & S. Holleran (Eds.), *Cognitive Basis of Musical Communication* (pp. 171-196). Washington, DC: American Psychological Association.

Lerdahl, F. (1996). Calculating tonal tension. *Music Perception, 13(3).* 319-363.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Lloyd, N. (1968). *The Golden Encyclopedia of Music.* New York: Golden Press.

Loftus, G. R. (1996) Psychology will be a much better science when we change the way we analyse data. *Current Directions in Psychological Science. 5 (6),* 161-171.

MacCallum, B. (1986). *A Reanalysis of the Unit of Perception of Ongoing Behaviour.* Unpublished master's thesis, Dalhousie University, Halifax, Nova Scotia, Canada.

Marr, D. (1982). *Vision.* San Fransico: W. H. Freedman & Company.

Massad, C. M., Hubbard, M. & Newtson, D. (1979). Selective perception of events. *Journal of Experimental Social Psychology, 15*, 513-532.

Meyer, L. B. (1956). *Emotion and Meaning In Music.* Chicago: University of Chicago Press.

Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and Perception.* Cambridge, MA: Belknap Press.

Millward, C. & Flick, J. (1985). *Handbook for Writers: Canadian Edition.* Toronto: Holt, Rinehart and Winston of Canada.

Narmour, E. (1991). The top-down and bottom-up systems of musical implications: Building on Meyer's theory of emotional syntax. *Music Perception, 9 (1)*, 1-26.

Narmour, E. (1992). The influence of embodied registral motion. In M. R. Jones & S. Holleran (Eds.) *Cognitive Basis of Musical Communication* (pp. 69-90). Washington, DC: American Psychological Association.

Narmour, E. (1996). Analyzing form and measuring perceptual content in Mozart's Sonata K. 282: A new theory of parametric analogues. Music Perception, 13 (3), 265-318.

Newtson D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*, 28-38.

Newtson, D. (1976). Foundations of attribution: The unit of perception of ongoing behaviour. In J. Harvey, W. Ickes & R Kidd (Eds.), *New Directions in Attribution Research.* Hillsdale, N.J.:Erlbaum.

Newtson, D. & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Psychology, 12*, 436-450.

Newtson, D. Engquist, G. & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 12*, 847-862.

Newtson, D. Hairfield, J. Bloomingdale, J & Cutino, J. (1987). The structure of action and interaction. *Social Cognition, 5 (3)*, 191-237.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to 'demand characteristics and their implications'. *American Psychologist, 17 (11)*, 776-783.

Osborne, K. & Gale, A. (1976). Bilateral EEG differentiation of stimuli. [Abstract] *Biological Psychology; 4 (3)*, 185-196

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15 (2)*, 331-346

Palmer, C. & Krumhansl, C. L. (1987a). Independent Temporal and Pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception and Performance, 13 (1)*, 116-126.

Palmer, C. & Krumhansl, C. L. (1987b). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics, 41 (6)*, 505-518.

Palmer, C & van-de-Sande, C (1996). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21 (5)*, 947-962

Paulus, W. (1992). Event-related potentials evoked by music lack a dissonance correlate.
Psychomusicology, 11 (2), 152-156.

Peretz, I. (1989). Determinants of clustering music: An apprasial of task factors. *International Journal of Psychology, 24 (2)*, 157-178.

Peretz, I. & Babai, M. (1989). Chunking melodies in the cerebral hemispheres. *Annual Meeting of the Psychonomic Society*:Atlanta.

Polanka, M. (1995). Factors affecting eye movements during the reading of short melodies. *Psychology of Music, 23 (2)*, 177-183.

Rahn, J. (1989). Notes on methodology in music theory. *Journal of Music Theory, 33 (1)*, 143-154.

Raffman, D. (1992). Proposal for a musical semantics. In M. R. Jones & S. Holleran (Eds.), *Cognitive Basis of Musical Communication* (pp. 23-32). Washington, DC: American Psychological Association.

Reynolds, A. G. & Flagg, A. G. (1983). Cognitive Psychology. Toronto, On: Little, Brown and Company.

Russo, F. A. & Cuddy, L. L. (1996). Predictive value of Narmour's principles for cohesiveness, pleasingness, and memory of western melodies. *4th International*

Conference on Music Perception and Cognition: Proceedings. 439-443.

Sadie, S. (Ed.). (1980). *The New Grove Dictionary of Music and Musicians.* (Vols 1-20). London: Macmillan Publishers Limited.

Scarr, S. (1997). Rules of evidence: A larger context for the statistical debate. *Psychological Science, 8 (1)*, 16-17.

Serafine, M. L., Crowder, R. G. & Repp, B. H. (1986). Integration of melody and text in memory for songs.*Cognition, 16 (3)*, 285-303.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58*, 75-125.

Schellenberg, E. G. & McKinnon, M. C. (1996). Melodic expectancy: A comparison of adults and children. *4th International Conference on Music Perception and Cognition: Proceedings*, 295-298.

Schiffman, H. R. (1982). *Sensation and Perception*. Toronto: John Wiley and Sons, Inc.

Shepard, R. N. (1982a). Geometrical approximations to the structure of musical pitch. *Psychological Review, 59 (4)*. 305-333.

Shepard, R. N. (1982b). Structural Representations of Musical Pitch. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 344-390). Toronto, On: Academic Press.

Shrout, P. E. (1997). Should significance tests be banned? Introduction to a special section exploring the pros and cons. *Psychological Science, 8 (1)*, 1-2.

Sims, W. L. (1988). Movement responses of pre-school children, primary grade children, and pre-service classroom teachers to characteristics of musical phrases. *Psychology of Music, 16 (2)*, 110-127.

Sloboda, J. A. (1977). Phrase units as determinants of visual processing in music reading. *British Journal of Psychology, 68*, 555-570.

Sloboda, J. A. (1992). Empirical studies of emotional response to music. In M. R. Jones & S. Holleran (Eds,), *Cognitive Basis of Musical Communication* (pp. 33-50). Washington, DC: American Psychological Association.

Sloboda, J. A. & Gregory, A. H. (1980). The psychological reality of musical segments. *Canadian Journal of Psychology, 34 (3)*, 274-280.

Sloboda, J. A. & Parker, D. (1985). Immediate recall of melodies. In P. Howell, I. Cross & R. West (Eds) *Musical Structure and Cognition.* (pp. 143-167) London: Academic Press.

SPSS-X User's Guide. (1988). Chicago:SPSS Inc.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates

Stoffer, T. (1985). Representation of phrase structure in the perception of music. Music Perception; 3 (2). 191-220.

Tan, N., Aiello, R. & Bever, T. G. (1981). Harmonic structure as a determinant of melodic organization. *Memory and Cognition, 9 (5),* 533-539 .

Taruskin, R. (1989). Reply to Brown and Dempster. *Journal of Music Theory, 33 (1),* 155-164.

van Egmond, R. & Butler, D. (1996). The perception of key and mode in trichords. *4th International Conference on Music Perception and Cognition: Proceedings,* 43-48.

Vos, P. G. & Troost, J. M. (1989) Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception, 6 (4).* 383-396.

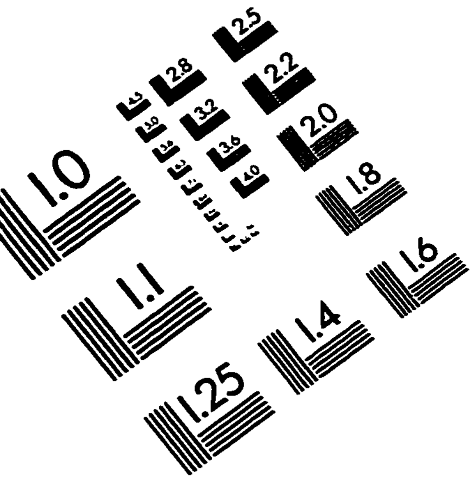
Warren, R. M. & Byrnes, D. L. (1975). Temporal discrimination of recycled tonal sequences: Pattern matching and naming of order by untrained listeners. *Perception and Psychophysics, 18,* 273-280.

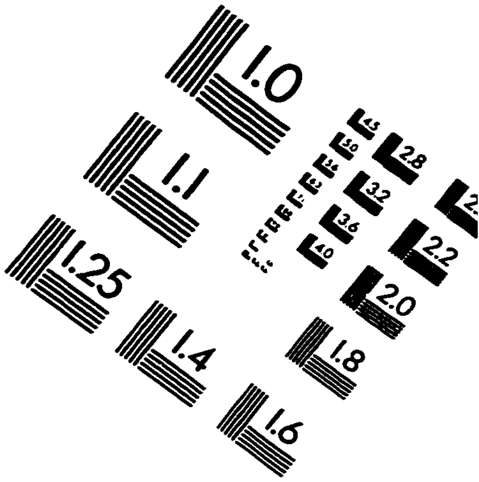
Watkins, A. J. & Dyson, M. J. (1985). On the perceptual organization of tone sequences and melodies. In P. Howell, I. Cross & R. West (Eds) *Musical Structure and Cognition.* (pp. 71-120) London: Academic Press.

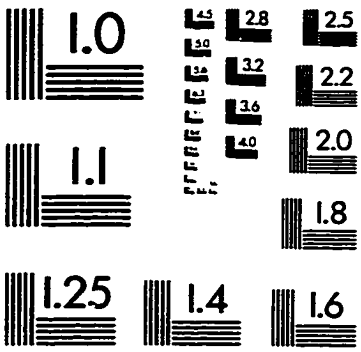
Watt, R. J. & Ash, R. L. (1996). An experimental investigation of the psychological content of music. *4th International Conference on Music Perception and Cognition: Proceedings,* 161-166.

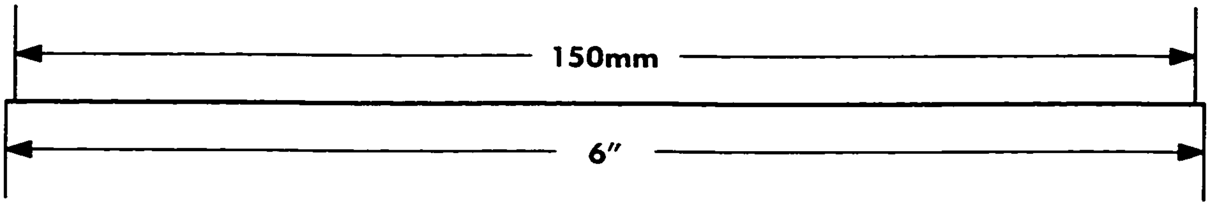
West, R., Howell, P. & Cross, I. (1985). Modelling perceived musical structure. In P. Howell, I. Cross & R. West (Eds) *Musical Structure and Cognition.* (pp. 143-167) London: Academic Press.

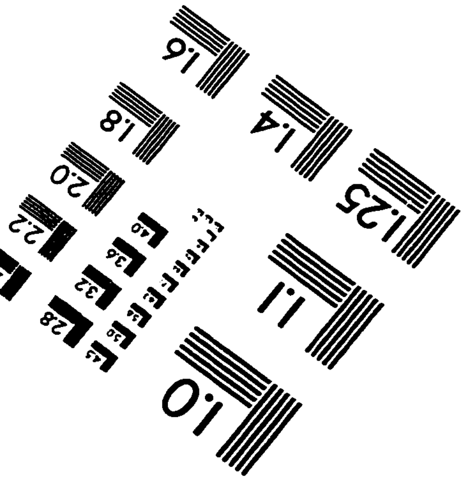
IMAGE EVALUATION
TEST TARGET (QA-3)

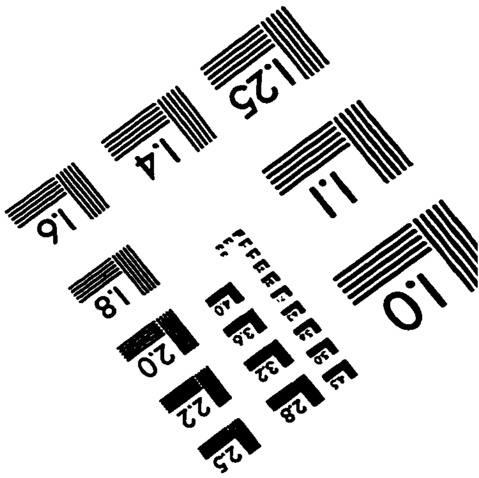












APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved