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.

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.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

PERCEPTUAL ASYMMETRIES IN A DIATONIC CONTEXT

By

MICHAEL RICHARD WOLOSZYN, B.SC. HONS.

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Doctor of Psychology

McMaster University

PERCEPTUAL ASYMMETRIES IN A DIATONIC CONTEXT

DOCTOR OF PHILOSOPHY (1999)
(Psychology)

McMaster University
Hamilton, Ontario

TITLE: Perceptual Asymmetries in a Diatonic Context

AUTHOR: Michael Richard Woloszyn, B.Sc. Hons. (McMaster University)

SUPERVISOR: Dr. J.R. Platt

NUMBER OF PAGES: vii, 156

Abstract

When investigating the perceived similarity of two musical stimuli, systematic asymmetries emerge which depend on the nature of the two elements being compared, relative to the musical context in which they appear. At the level of individual tones, Krumhansl (1979) found that if the order of presentation was diatonic/nondiatonic (relative to a tonal context), similarity ratings were lower than if that order was reversed. She concluded that when a change results in an element becoming less stable in terms of it's position on the tonal hierarchy, similarity perception will be lower than if a change increases it's stability. At the level of melodies, Bartlett and Dowling (1988) obtained a similar result, but claimed the asymmetry was due to violations of scalar structure, having little or nothing to do with the tonal hierarchy. To test between these different accounts, a series of experiments was conducted in which pairs of diatonic melodies were presented for similarity ratings. Each melody consisted of a context sequence and a target sequence. The context sequences were designed to promote either a C-major or a D-minor tonal hierarchy. These respective keys share similar notes in their scales, but the tonal hierarchies are inverted with respect to one another. According to Krumhansl's account of the asymmetry, noticeable changes in one key context should not be noticeable in the other, due to the reversal in stability of the component tones of the alteration. According to Bartlett and Dowling, no differential sensitivity should be observed, since all changes were within the scale structure of both contexts. In the first experiment, repetition of the tonic was employed as the key-instantiating stimulus, the result being that strong asymmetrical perception arose, both on the measures of similarity ratings, as well as alteration detection ability. Subsequent experiments employed triadic contexts (suggested by Krumhansl and others to be strongly key-instantiating), and note-frequency controlling contexts (to rule out the possibility of note repetition playing a role in the similarity ratings). The results supported the hypothesis that asymmetric perception is a result of the dynamic tone quality differences between scale degrees in a tonal melody. Two subsequent control studies ruled out the possibility of target sequences themselves being responsible for the asymmetries, and confirmed that listeners perceived the melodies in the keys specified in the experiment. A model based loosely on Tversky's explanation of asymmetric perception was put forth to explain these data, as well as those of Bartlett and Dowling.

Acknowledgements

The process of obtaining a degree of this sort involves so much more than learning the basics of methodology and the rapid acquisition of a knowledge-base. The experience also involves personal growth, and the realization that one's ideas can be worth something. I have many people to thank, each of whom contributed to a different aspect of my development on this long winding road to doctorhood. So.

Thanks to mom and dad, for understanding the trip I've been (and still am) taking. Your support and faith meant a lot, in spite of what I may have said.

Equal thanks to John Platt, for being the perfect supervisor both in and out of the office. They should make hot-tubs a necessary part of graduate equipment.

Thanks to Ron Racine for being so critical of my work. There's something to be said for hard-headed reductionism, and steering clear of those nasty constructs.

Thanks also to Lorraine Allen, first for letting me into graduate school in the first place, and second, for keeping me on my conceptual toes by forcing me to explain the musical concepts to a nonmusician.

Thank you Erie, for being my surrogate mom, and for helping me turn from a cringing, frightened new graduate student into the somewhat less frightened creature you see before you. Thanks also to Sally, Wendy, and Bev, for smoothing out the red-tape no matter what the situation. Truly, you are the ones running that department.

Thanks to all the graduate students and post-docs I've been in contact with over the years (Lloyd, Chapman, Bill, and Yoni to name a few), as well as that wild and wacky bunch at the Phoenix. You've all helped shape my psyche in new and weird ways.

In short, thanks to everyone in the department, faculty staff and students alike. It really has been fun.

Finally, thanks to my wife Csandra, without whom the final push would undoubtedly not have occurred. Huggles and snuggles, soulmate.

TABLE OF CONTENTS

List of Illustrations

Chapter I: Introduction

This chapter consists of several distinct parts. In the first section, the goal will be to familiarise the reader with some of the musical terminology and theory that will be used throughout the text. Once equipped with this knowledge, relevant issues concerning melody perception (what a melody is, what features seem to play a role in the storage and retrieval of melodies, and what underlying representations may exist in listeners) will be discussed. Finally, the discourse will turn toward the particular topic addressed by the experiments in subsequent chapters.

1.1 Music Terminology and theory

Music, in general terms, consists of the presentation of a set of frequencies. When those frequencies form part of a musical composition, they are called 'notes' in music terminology, or musical tones, in psychological terminology. While there are exceptions, most music has been written with reference to a single note, called the 'tonic'. This note serves as the focal point, from which the melody deviates and returns to, in the course of a song.

All music consists of notes drawn from a set of possible frequencies, called the 'scale'. Scales are determined by the interrelationships formed between the notes in the set, in terms of their frequency ratios. Several different methods for determining these interrelationships exist, and they are collectively known as 'tuning systems'. One feature most tuning systems have in common is that of the octave, which refers to two notes in a 2:1 frequency ratio with respect to one another. The perceptual effect of hearing two notes in an octave relationship is that they sound very similar to one another. Music theorists have acknowledged this for some time, and have provided notes in an octave relationship with the same name. 'Octave doubling', or the playing of the same notes in different octaves, frequently occurs in music in order to emphasise a particular line. Therefore, when discussing tuning systems, it is necessary only to describe the set of notes that fall between two notes standing an octave apart, since those relationships will repeat when one looks at a different octave level.

Although there are many different possible tuning systems, resulting in many different sets of musical notes, the music most Western listeners are familiar with is based on the 'Equal Tempered scale'. This system constructs the set of musical tones by dividing the octave into twelve logarithmically equal steps. The result is more or less the set of notes one would find on any piano today. The distance between two adjacent notes on a piano are equal, in log-frequency values, to the distance between any two other adjacent notes. This distance, when based on the number of steps between two notes in the system, is called a 'musical interval'.

A simpler method for dividing up the octave, which historically preceded the Equal Tempered scale, would be to determine that set of notes which stand in the least complex frequency ratios with the octave notes. For example, the octave consists of two notes which are in a 2:1 frequency ratio relative to one another (one note is twice the frequency of the other). If one were to find that note which stands in a 3:2 frequency ratio with an octave note, then a 4:3 ratio, and so on, one would be constructing a Just Intonation scale. The advantage of the Equal Tempered system is that one can transpose a piece of music from one tonic to another, and the same set of interrelationships between the notes would result. For example, you can sing the song 'Happy Birthday' beginning on any note you wish, and it will still sound exactly like the Happy Birthday you are familiar with, in part because the interval relationships between the notes are exactly alike. The same cannot be said for Just Intonation sets, since the frequency ratios are set up with respect to a single tonic. If one were to play a melody beginning on a note other than that tonic, the frequency ratios between notes would no longer be simple, and the song would sound like it was being played on a badly tuned instrument. As a result, the Equal Tempered system has become the standard for Western composition. The remainder of this thesis is concerned only with the perception of music written from within this system.

Once a tuning system has been used to construct the set of possible notes, one can begin to talk about further subsets of notes within that system. The most general set, called the chromatic scale, consists of all twelve notes that fall within an octave (all the black and white notes in an octave, on the piano). Within this, however, there are other sub-sets of notes, which are highly relevant from a compositional point of view. One of the most important of these is the 'diatonic scale', or the 'diatonic set'.

The most parsimonious way of describing the diatonic set in print would be to refer the reader to the film 'The Sound of Music'. One of the songs in that movie, "Doe, a Deer", is based on the 'do-re-mi' scale, which, in music theoretic terms, is a major diatonic scale. If one were to begin with any note as the 'do', or tonic, the steps between adjacent notes in the scale correspond to the set {2,2,1,2,2,2,1}. That is, if you begin on any note on the piano, the next note in the diatonic set would be two notes higher (going from left to right on the piano), the third would be two away from the second, and so on. One finds that a majority of the music written in the last several hundred years has employed this scale, although more recent efforts have deviated significantly from its use (I refer the reader to 20th Century composers such as John Cage, or Schoenberg's twelve-tone serialist pieces as examples).

The notes that were constructed via the tuning system have been given names by musicians, using letters of the alphabet. If one looks at the piano, the lowest possible note, at the far left of the keyboard, is given the name 'A'. The white note immediately adjacent to it is named B, then C, and so on, up to G. The white note adjacent to G, being an octave from A, is again called A. This cycling continues to cover the entire keyboard. So, with this naming system, one can see that musicians have incorporated the notion that octave-related notes share a high degree of similarity.

The 'black notes' on the piano may be conceptualised as 'deviations' from the white notes (this is not truly how music theorists conceptualise them, but this discussion is to aid the reader in understanding the rest of this thesis, not to turn her/him into a seasoned music theorist). The names for these black notes derive from their positioning relative to the white notes. For example, the black note to the immediate right of the note A is labelled A-sharp (written A#). The note immediately to the right of any white note is referred to as "that-note-name--sharp". For reasons that will not be taken up here, one could also label any note immediately to the left of a given white note as "that-note-name--flat. So, each of the black notes has two names. In our example, A# could also be referred to as B-flat.

Another issue pertinent to the work presented here is a distinction within the diatonic scale known as 'major' versus 'minor' scales, also known as the major or minor 'mode'. The diatonic scale described earlier was actually a major scale. To be colloquial for a moment, major scales can best be characterised as 'happy-sounding', while minor scales are 'sad/somber' sounding. The main contributing factor for this is due to the identity of the third note of each scale. For example, by using the set of intervals presented earlier, we could construct a major diatonic scale by beginning on C, then moving along with D-E-F-G-A-B-C. The note E is a prominent factor in determining that this is a major (or 'happy') scale. If one were to replace that E with E-flat, however, and play the first few notes of the scale, the mood suddenly shifts to a more sad and somber one. So, Happy Birthday is just that, largely because the third note is from the major diatonic scale. The opening notes of Beethoven's Moonlight Sonata have a somber and sad quality, however, largely because the third note of the scale in which it was written is flattened. If one were to flatten the third in Happy Birthday, it might lead listeners to conclude that the song was wrongly named.

One other term that becomes relevant in the text to follow is that of 'key'. When a song is written which uses notes from a particular diatonic scale, the song is said to be in the 'key' of that scale. So, a melody written using the C-major diatonic scale is said to be written in the key of C-major. If it were written using the D-minor scale, it would be in the key of D-minor.

This concludes the introduction to the music theory concepts which are relevant to the work presented below. Further attributes and features of melodies will be introduced as necessary, but at this point, a basic understanding of what music is made of, hopefully, has been conveyed. Of interest now is the music itself, or more specifically, the perception of melodies. An appropriate starting point would be to define what is meant by that term.

1.2 A Definition of 'Melody'

An unfortunate situation exists in the area of music perception: Currently, there is no accepted definition of the term melody. This is unusual, since the perception of 'melody' has been of interest to researchers for quite some time, even though these same researchers skirt the issue of what features are involved in determining if a note stream is considered a melody. Usually, this issue is avoided by referring to the stimuli as 'note-streams', but doing so leaves open the question of whether listeners consider these stimuli to be music or not, and calls into question the external validity of the research. In particular, there is no reason to believe that listeners will process isolated acoustic events in the same way as when those events are incorporated into more realistic melodies. This has long been a criticism of music psychology in general, by musicians and music theorists: music psychology isn't using 'music' as stimuli. While a greater percentage of research reports coming out are now using excerpts from published music, it remains that much research in music perception cannot be performed in this way, if any degree of control over the stimulus is to be exercised. Therefore, in order to construct melodies to be used as stimuli, an adequate description of the term is necessary, but not yet forthcoming.

Perusal of both psychological and music theoretic literature reveals several different proposed descriptions of the term 'melody', all of which are inadequate, either because they are too general, or too specific. A broad definition of melody might be 'any set of serially presented tones' (Ortmann, 1926). Clearly, this is an insufficient definition of melody, since the presentation of a randomly selected set of tones will result in the perception of a randomly selected set of tones. This definition does not capture whatever properties of melodies contribute to our perception of them as such. One could resort to a behavioural definition, such as 'the melody is whatever part of a song the subject would sing, if asked to sing the melody'. Radocy and Boyle (1988) subscribe to this description: "Ultimately, only the perceiver can judge whether a tonal sequence functions as a melody: If it does, it is a melody." (p.139). Unfortunately, while this serves as an adequate operational definition, it does not further our understanding of melody construction. However, it seems that any reasonable description of melody will take into account two separate factors: those structural characteristics of melody that seem to recur in written music, and the listener's response to those characteristics. A simple way of putting this, would be to define melody as "a series of tones which makes sense [to the listener]" (Zuckerkandl, 1956). To this end, rather than attempt to develop an all-encompassing definition of melody, it is necessary to examine these two different perspectives which together contribute to the perception of a sequence of tones as being a melody. The end result will be to have a set of features and characteristics that will be used in the construction of more externally valid melody stimuli, while retaining control over extraneous factors.

1.3 Structural Characteristics of Melody: The Musicological Viewpoint

Lundin (1967) identifies three attributes which most tonal melodies share in common. First, there is propinquity, which refers to the finding that most melodies are constructed using small successive intervals between component tones. One can find ample evidence of this feature of melodies by examining the occurrence frequencies of various sized intervals in large bodies of musical works (Ortmann, 1926; Radocy, 1977; Jeffries, 1971). A second attribute is repetition, where certain scale degrees tend to be sounded much more often than others. In particular, Radocy (1977) found that upon examination of two large collections of popular music (The Norton Scores, and 357 Songs We Love to Sing) the tonic ("Do", or first), mediant ("Me", or third), and dominant ("Sol", or fifth) degrees of the scale accounted for a majority of the tones sounded. This repetition factor will become important later, when psychological representations of musical pitch are considered. A third melodic attribute identified by Lundin is that of finality. That is, melodies tend towards a final resting point, where the listener is left with the impression that the end has been reached. This end point usually consists of the tonic, third, or fifth (which, collectively, make up what is called the root chord, or root triad).

So, by merely examining the structural attributes written melodies have in common, it is possible to conclude that melodies seem to consist of tone sequences with small successive intervals, a high density of root-chord notes, and some sort of resolution, or common ending (usually the tonic). As will be seen further on, many researchers disregard these features in the construction of their stimuli, calling into question the conclusions drawn from that work.

1.4 Psychological Viewpoint: Data-driven Conclusions

Research in melody perception can be put into two broad categories. First, there is that class of research that attempts to test the predictions made by musicological and music theoretic descriptions of melodies. For example, researchers have investigated whether or not listeners actually perceive the uppermost register as being the melody line. In an early study of this, Farnsworth (1938) presented listeners with dyads (simultaneously sounded pairs of tones) and asked them to state which component tone more resembled the dyad. Subjects showed a preference for the higher of the two component tones, a finding which Farnsworth labelled 'melody hunting'. Of course, a pair of notes is not what most people would call a song, and Farnsworth was not directly testing which note listeners considered to be the melody. In order to investigate this phenomenon with more musical stimuli, an unpublished study was carried out at McMaster University. Listeners heard a polyphonic phrase consisting of two different ten-note streams. The streams were presented an octave apart, and were sounded using different timbres, in order to facilitate perceptual grouping into two distinct streams (Bregman & Campbell, 1971; Miller & Heise, 1950). Upon analysis, it was found that subjects demonstrated a significant tendency to regard as the melody the stream being played in the higher of the two octave levels.

Another example of this type of work would be Meyer's examination of the perception of melodic schema. He noted that there are certain compositional regularities that recur in tonal melodies. One such regularity, a gap-fill melody, consists of a large interval jump (the gap), which is immediately followed by a series of small intervals in the opposite direction (the fill). Rosner and Meyer (1986) presented melodies in which this structural regularity was either maintained or violated, and found that listeners were quite adept at distinguishing between the two, thus supporting the perceptual reality of this music-theoretic concept.

Still other research has sought evidence of the influence on perception of various other music-theoretic notions (Cook, 1987; Rosner & Narmour, 1992), the results of which do not support the predictions made by music theorists. For example, music theorists and musicians alike would stipulate that the order of presentation of particular sections of a piece of music play a large role in contributing to aesthetic quality, or pleasingness. Cook (1987), however, found that if such altered pieces are presented to listeners, they sometimes consider the altered piece to be more aesthetically pleasing than the unaltered piece. Similarly, Rosner and Narmour (1992) found listeners' ratings of closure, or finality, for two-chord cadences were not influenced by several factors commonly believed by music theorists to influence that perception. In summary, then, this class of research has given psychologists and musicologists insight into the psychological validity of compositional rules and music-theoretic edicts as set out by theorists and composers.

A second class of research, which appears to be more popular in the psychological literature, is concerned with the study of the primary structural characteristics of a melody which are influential in their perception and storage by listeners. Features such as contour (general direction of the melody in terms of 'up' or 'down' in pitch), melodic intervals (the scale-step distance between successive notes), and tone (log-frequency of a particular note) have received a great deal of research attention, and are briefly examined below.

Contour appears to be a highly salient feature of melodies, one which is equally well-represented in musical and non musical listeners (Dowling, 1978). In general, one can represent a melody in terms of its distinctive contour, and evidence has been obtained which shows that this information allows for the recognition of a melody. Werner (1925) distorted familiar melodies by transforming them onto microtonal scale systems. The result was that listeners could still identify the melodies upon which these distortions were based. White

(1960) extended this finding, by compressing the melodies such that all of the intervals were set to be a semitone (two adjacent notes on the piano) in size, and found that listeners could still identify the melodies, despite their being recognisable only on the basis of contour. It appears, then, that contour is a feature which is recognised and stored by listeners, and this feature can by itself be used in the recognition of a familiar melody.

Contour is also considered a salient feature of novel melodies, as evidenced by studies employing short-term recognition tasks. Early research demonstrated that listeners tended to confuse transposed melodies in which the intervals were altered but contour was maintained, such that these melodies were not distinguishable from exact transpositions (Dowling, 1978). Likewise, if the contour was violated, musicians and non-musicians alike were quite adept at detecting the alteration, provided the ISI between standard and comparison stimuli is relatively small. However, contour violations become increasingly difficult to detect over the long-term, either by increasing the length of the melody (Edworthy, 1985), or by increasing the retention interval (Dowling & Bartlett, 1981).

On the basis of this early work, researchers concluded that contour was a perceptually separable and independent feature of melody, since other features such as the interval structure or tonality did not appear to perceptually interact with it (Dowling, 1978). That is, one could severely distort a melody on other dimensions, but with contour intact the melody would remain recognisable, since the listener apparently extracted and stored the contour information independently of those other features. If this conclusion were true, then one could conceptualise a listener's representation of a melody as consisting of nothing more than a tonal scheme being shaped by a contour. However, Dowling (1991) came to the conclusion that contour and tonality interact, producing a perception of an integrated whole. His study required listeners to distinguish between exact transpositions and contour violations of short melodies, which varied in terms of the degree to which they conformed to a diatonic framework. It was found that listeners' ability to detect contour violations depended on the tonality of the piece, such that contour violations for less-tonal melodies were much more difficult to detect than was the case when the melodies were highly tonal. As a result of this finding, the notion of contour being a separable feature fell out of favour in the literature. More recent thinking stresses that a given feature of a melody, although independently manipulable from other features, can give rise to changes in the perception of those other features.

The set of intervals formed by adjacent tones in a melody is also considered to be an important feature of melodies. In particular, it appears that well-learned melodies are stored as a series of successive intervals (Attneave & Olsen, 1971; Bartlett & Dowling, 1981), in both musically trained and untrained subjects. While this appears to disagree with the earlier statement with respect to contour, the reader should understand that the set of intervals of which a melody is constructed also contains contour information. That is, the intervals consist of both direction (contour) and distance information, while contour consists solely of direction, irrespective of distance. Thus, while it is possible to manipulate interval independently of contour for a given melody, it is not possible to do the reverse. So, when a melody becomes highly familiar to a listener, the interval information must also contain contour information. This is apparently what enables listeners to recognise a familiar melody, regardless of the key in which it is presented.

Several studies, however, have shown that for short-term recognition, the processing of melodic intervals varies over time, and the temporal influence on retention is opposite to that obtained with contour. Specifically, listeners appear to better detect an interval violation following longer delays rather than shorter ones. This delay can come about either by lengthening the melody itself (Edworthy, 1985), or by lengthening the ISI between the standard and comparison melodies (Dowling & Bartlett, 1981). The explanation for this seemingly paradoxical result hinges on two mechanisms. First, the interval information is blocked from access in the short-term by the salience of the contour information. Second, the interval information presumably takes more time to store effectively, requiring either repetition of the melody to afford consolidation, or a delay between presentation and test conditions. As a result, the interval information will only become useful to a listener after a period of time has passed.

Finally, in pilot research conducted at McMaster, a clear relationship was obtained between ratings of melodiousness and overall successive interval sizes of brief (8-12 note) diatonic melodies. In addition, average interval size of the melody also played a role in subjects' ability to detect an alteration, indicating that melodies comprising large intervals are less well-remembered or represented. The most likely explanation for these results is that the preponderance of larger intervals in the melody was not conforming to listeners' expectations (Lundin's 'propinquity' feature), resulting in a breakdown in their ability to accurately predict upcoming events (Schmuckler, 1989). Ample evidence indicates that listeners tend to expect small melodic intervals to appear in a melody (Carlson, 1981; Unyk & Carlsen, 1987), and some evidence is emerging which indicates that these expectancies are based on certain musical regularities such as melodic process (Meyer, 1973) and propinquity (Lundin, 1967).

So, it is clearly not the case that people perceive and store a novel melody as a series of discrete pitches. Instead, they seem to store an amalgamation of contour and melodic intervals, in addition to other features not yet discussed (mode, tonal structure). In summary, it appears that listeners focus on the relational aspects of the individual tones, either in direction (contour) or distance (interval), and their ability to accurately code these aspects might be influenced by expectations concerning upcoming events (Jones, 1981; 1982). To this point, however, very little has been said about the actual tonal information itself. To this end, psychologists have become increasingly interested in modelling listeners' encoding of musical pitch, with varying degrees of success.

1.5 Psychological Representations of Pitch

A. Psychophysical Account

This seems to be a trite statement, but there are only two types of sound in the world. One type of sound consists of a single frequency component, which graphically appears as a sine wave (when recording amplitude fluctuations over time). These sounds are called simple or pure tones, due to their simplicity to describe, and their predictable effect on the auditory system. These sounds do not occur naturally in the environment, requiring instead an electronic sound generator. Phenomenologically, the pure tone timbre roughly resembles a flute.

The other type of sound consists of combinations of pure tone frequency components. These sounds are, appropriately, called complex tones, presumably due to the complexity of the waveform when graphically represented, as well as the sometimes unpredictable effect they have on the auditory system. Each pure tone component is known as a "partial" of the complex wave, and the interaction of the pure tone components gives rise to the waveform of the complex sound. In fact, it is possible to deconstruct any complex sound into its pure tone components. Much work has gone into studying the perception of complex waves, giving rise to several different models of sound perception (Moore, 1982). These models are beyond the scope of this paper, however. Of interest to us is how people perceive these frequencies when they are presented in an organized sequential fashion.

Prior to going any further, a distinction must be drawn between the terms 'frequency' and 'pitch'. Frequency refers to the number of wave amplitude repetitions passing a point in space, per unit time. In other words, it is a physical property of sound, referring to the changes in air pressure that constitute the sound. Pitch, on the other hand, refers to the listener's perception of frequency. These two properties are related, in that low frequencies generally result in the perception of a 'low' pitch, such as that heard coming from a tuba, while high frequencies result in the perception of high pitch, such as the sound of a flute (the terms 'low' and 'high' are arbitrary, but universally accepted as descriptors along the frequency continuum). The reader should note that frequency of the wave as a whole is dependent upon the interaction of whatever sine-wave frequencies comprise the sound. These partials might, in fact, consist solely of high frequency waves, but their interaction and co-interference produces a resulting wave with a much slower repetition rate. Researchers have known since the seventeenth century that pitch and frequency are related, but only recently have realized that there is not a linear mapping of frequency onto pitch, and that frequency is not the sole determining factor in arriving at a perceived pitch, especially when the tone forms part of a musical context, or consists of more than one wave component.

Early research approached the topic of frequency perception from within a psychophysical framework, by directly examining the perceived pitch of isolated tones. Using a matching procedure, Stevens and his associates created a 'mel' scale, depicting the relationship between frequency proximity and pitch similarity without a musical context (Stevens, Volkman, & Newman, 1937). The general finding was an accelerating curve, such that the perceived pitch of a comparison tone generally increased with increases in frequency distance from the standard, but higher frequencies had greater psychological distances between them than did lower frequencies. That is, as the comparison tone increased in frequency, the perceived pitch difference between that comparison and the standard increased at a higher rate.

While the mel scale is useful in describing the basic relationship between frequency and pitch, researchers today agree that this description is far too simplistic to account for the processing of tones in a melodic context (Shepard, 1982a). In general, it can be said that the way a tone is perceived depends to a great extent on the musical context (or lack thereof) in which that tone is heard, much the same way a word may be perceived differently depending on the surrounding words, and the context those words form in a sentence (Ehrlich & Rayner, 1981). In other words, there is undoubtedly a difference between the pitch perception of acoustical tones (presented in isolation) and musical tones (presented in a musical context).

B. Music-theoretic accounts

Shepard's Helix

Shepard (1964) examined the music-theoretic notion of 'octave equivalence', demonstrating that listeners are sensitive to these frequency relationships. In general, octave equivalence refers to the high degree of perceptual similarity that is shared by tones separated by a 2/1 frequency ratio. By constructing complex tones (more than one frequency component) that varied in their spectra depending on the fundamental frequency on which they are based, Shepard created the auditory equivalent of an 'Escher staircase', such that the pitch is always heard to be increasing from note to note, but the scale itself never seems to actually rise over time. Armed with this illusion, Shepard proposed that in addition to the dimension of pitch height, described by psychophysics, there also existed the dimension of pitch class, which refers to the collection of notes standing in an octave relation to one another. In fact, his early model of pitch perception was bidimensional, with pitch class being one aspect, and pitch height being a second, orthogonal property (Shepard, 1964). More recent versions of this structural representation incorporate the three components of pitch height, the chroma circle, and the circle of fifths, resulting in a five-dimensional model (Shepard, 1982a,b). It became clear from this work that the properties of tones as delineated by music theory have measurable perceptual effects in listeners. Subsequent research further reinforced this notion.

The Tonal Hierarchy

Music theorists have long upheld the position that the perception of musical notes involves much more than the simple encoding of specific frequencies. Rather, they assert that the music heard contains meaning and features beyond that which is measurable by an oscilloscope or tuning meter (Zuckerkandl, 1956). In particular, reference has been made to the 'dynamic qualities' of the tones in a piece of music. This refers to the idea that a particular tone might have a different function, depending on the context in which it is heard: The function of the note D is different, depending on whether the context is C-major or D-major. For a time, this was considered to be little more than an interesting but scientifically unsupported claim. However, work over the last twenty years has provided experimental support for the perceptual existence of dynamic tone quality. The starting point for this research was a seminal paper by Krumhansl & Shepard (1979), in which they outlined a new procedural tool to investigate pitch perception, and discovered some interesting new properties concerning that perception. The procedure has been commonly referred to as the 'probe tone' technique, and warrants brief elaboration.

The essential features of the probe tone technique are reminiscent of psychophysical procedures: Listeners hear a context of some sort (a chord, scale, chord sequence, etc.) followed by the presentation of a single tone. The task required of subjects was to rate how well the single tone either completed, fit with, or belonged to the preceding context. The outcome from these studies was that tones were consistently given different ratings, depending on their music-theoretic position in the preceding scale. In particular, the tonic was given the highest fitness rating, closely followed by mediant and dominant tones, the remaining diatonic tones, and finally the remaining (non-diatonic) tones. While this effect was most strongly obtained with trained musicians, a similar, albeit much weaker pattern in non-musicians was also obtained. Upon further examination, it was seen that the less-trained listeners were much more influenced by the absolute pitch-height of the probe-tone relative to the preceding context, such that tones near in frequency to the final tone of the context were given higher ratings than more distant probes.

This differential set of ratings for members of the chromatic scale has been termed a 'tone-profile', a pattern apparently reflective of the internalization of a tonal hierarchy in the listener. Subsequent work identified tone-profiles for all of the major and minor keys (Krumhansl & Kessler, 1982), the result being that although the third and sixth scale degrees were associated with much higher ratings in major contexts relative to minor contexts, in other respects the profiles were essentially the same.

In addition, Krumhansl (1979) performed a more elaborate study in which a diatonic scale context was followed by two notes. All possible two-note combinations were presented to subjects, whose task was to rate their similarity to one another. The resulting set of ratings was subjected to a multidimensional scaling analysis. The outcome of that analysis was a three-dimensional conical representation, with perceived similarity being depicted as the relative distance between two nodes. The structure generally consisted of three tiers of clustered notes, with the root-chord notes clustered around the vertex of the cone, the remaining diatonic notes grouped on a second level further away from the vertex, and non-diatonic notes on a third, more distant level. Her conclusion alluded to the possibility that this structure was a graphic representation of an internalized hierarchy possessed by musicians, and that it was used in the organization and structuring of incoming musical information. In other words, the conical structure is apparently a graphical representation of the 'tonal hierarchy', although Krumhansl later stressed that it is merely a graphical representation of listeners' responses, and as such was obtained without any a priori theoretical considerations or constraints (Krumhansl, 1990). She did, however, claim that it represents a manifestation of listeners' sensitivity to the structure and regularities found in most musical compositions, and described by music theorists. So, when hearing a song, listeners apply their abstract knowledge of the regularities found in music to make sense of the song, and one aspect of that knowledge is an understanding of the hierarchical (or dynamic) nature of musical tones. It has been shown that much written music does conform to a tonal hierarchy of sorts, as evidenced by the tone frequencies employed in their construction (Youngblood, 1958; Knopoff & Hutchinson, 1983). One finds that the root, third, and fifth note of the scale tend to be used by composers much more often than either the remaining diatonic tones (which are the second-most frequent) or nondiatonic tones (least frequent). Krumhansl's claim is that this stylistic regularity in written music is what gives rise to the pattern of obtained ratings, in that listeners will eventually learn to hear the frequently-presented events as being the most stable in a melody. In other words, the root-chord notes will tend to serve as cognitive reference points (a concept borrowed from Rosch, 1975). When a new melody is presented, listeners will then attempt to apply this knowledge in organizing the elements of that melody into important versus subordinate events. Presumably, musicians would have a much stronger conceptualization of the tonal hierarchy, partly as a result of exposure to a significantly greater amount of tonally rich music (relative to a nonmusician), and partly as a result of having been trained to identify theoretically significant intervals and chords in the diatonic framework. That is, part of traditional music training involves learning to identify root chords, singing the root note of a brief melody, as well as interval relationships in isolation. It seems likely that this training, if early enough, would give rise to a better internal representation of music.

Critics of Krumhansl's work have pointed out that she uses the term 'tonal hierarchy' both as a representation of data, and as a theoretical construct (Butler, 1990). This is not an unusual situation, however. If it were the case that subjects have an internalized tonal hierarchy, then such a structure would be expected to manifest in the data, although the circularity of such an argument prohibits its usage. An indisputable fact, however, is that listeners' similarity ratings were structured in a highly regular fashion, as depicted by the conical outcome of the scaling solution. On the one hand, distance between elements represents similarity, while on the other hand distance from the vertex appears to represent the stability of an element within the framework depicted by the structure as a whole. However, this structure is not itself the tonal hierarchy. Rather, it is a graphic representation of listeners' patterns of similarity ratings, which have apparently been influenced by an internalized understanding of music structure being hierarchically organized. In other words, it is an index of the tonal hierarchy, and is useful as a short-handed method of identifying the psychological similarity relations between constituent notes and the stability/representativeness of those notes in a defined tonal context.

Since Krumhansl's original finding, evidence of an internalized tonal hierarchy has been obtained under a variety of conditions, and with several different classes of subjects. The use of broken triads, diatonic scales, and successively presented harmonic chords organized in a standard musical progression have all been successful in leading subjects' ratings of 'belongingness' to conform to that of a tonal hierarchy (Cuddy & Badertscher, 1987; Krumhansl

& Kessler, 1982). Evidence of the use of a tonal hierarchy is apparent in other styles of music besides those conforming to the classical Western harmonic structure described at length earlier. When these novel melodies are presented to listeners unfamiliar with the styles, it appears that the listeners will attempt to apply their knowledge of tonal structure, or the tonal hierarchy, in an attempt to make sense of the piece (Castellano, Bharucha, & Krumhansl, 1984). That is, they resort to summing frequency of occurrences of the notes, with the more frequent notes being heard as more stable. Evidence of the development of the tonal hierarchy in children has also been obtained, such that with age and (assumedly) exposure to music, children's representations of the structuring of music becomes more finely tuned, starting with scale-note/non-scale-note distinctions, and leading to finer distinctions within the diatonic scale (Krumhansl & Keil, 1982). Most recently, analysis of jazz improvisations has demonstrated that the choice of tones employed by musicians when performing a jazz solo conform to Krumhansl and Kessler's (1982) tone profiles, indicating that at the melody-production stage, hierarchic structure is prevalent (Jarvinen, 1995).

Music theorists and musicologists have long embraced the idea that music is structured hierarchically. Schenker (1935/1979), for example, proposed that tonal music is structured hierarchically, ranging from the surface structure of the actual notes themselves (foreground), to the overall implied harmonic structure (background). According to Schenker, an entire piece of music will, over time, point to or imply the tonic triad. That is, one could take any tonal melody, and over a series of transformations reduce that composition to its underlying background structure of the tonic triad. So, Schenker proposed that music is structured hierarchically, with tones being subsumed into more general structures of harmonic implication, while Krumhansl proposed that the tones of which music is composed can themselves be mapped in terms of a hierarchy of stability or importance. The implication from Schenker's view is that the tonic triad is the most important or stable element of any tonal composition. Any other tonal elements in the piece are transient points employed by the composer to prolong or elaborate a piece in the process of achieving the root chord. So, although Schenker's usage of the term hierarchy is different from that of Krumhansl's, the ends are both essentially the same: Certain pitch classes subserve certain others in promoting the structure of a piece of music. In the theoretical domain, Deutsch & Feroe's (1981) attempt to model the structuring of music resembles a formalization of Schenker's ideas, such that a set of hierarchically related musical operators may be internalized and used by listeners in organizing musical materials. Although their model is somewhat limited in terms of predictive power, it does offer a psychological interpretation of music-theoretic notions that comes close to describing perception in terms of tonal hierarchies.

One can find references to tonal hierarchies in other music theoretic work, although they sometimes refer to it as the 'dynamic quality' of tones in music, or the 'meaning' of a tone. The main ideas, however, are identical to that of the tonal hierarchy: A tone will convey a different perceptual effect on the listener in different musical contexts, and each tone in a given context will convey a different perceptual effect relative to each other tone.

Intervallic rivalry

Although the existence of the tonal hierarchy as reflected in listeners' response patterns has been confirmed in a wide variety of studies, conditions, and subjects, there is some reason to question the validity of the conclusions due to the methodology employed. In particular, Butler (1990) outlined four possible confounding factors when employing the probe-tone technique, confounds which could readily explain the pattern of results Krumhansl obtained. First, the original work done by Krumhansl used context stimuli designed to clearly delineate a tonal centre. Her stimuli usually consisted of either an ascending or descending diatonic scale, or the root chord of a particular key. The problem with using these stimuli is that of differing frequencies of occurrence of each of the twelve possible pitch classes subsequently tested. This is most evident when examining chordal context stimuli. As an example, if the intended tonal context is to be C-major, the chord employed would consist of the notes C-E-G, and possibly a second C one octave above the first. These tones happen to be the same ones which were subsequently considered by listeners to be the most 'stable', or fitted best, in relation to the context. In fact, Butler (1989) compared the stimulus profiles obtained by summing the frequency and duration of occurrence of the tones in the chromatic scale with the tone profile obtained using the probe-tone technique, and discovered that the two patterns were nearly identical. The likely explanation of those data, then, is simply that listeners rated frequently heard tones from the preceding context as more stable than tones not appearing in the context. Of course, this does not explain the difference in perceived fittingness obtained between the other (non-root chord) diatonic tones, and non-diatonic tones which emerged from the analysis. It also does not explain the patterns obtained when a diatonic scale was used as the context, since (aside from the root note) the tones occurred with equal frequency.

Still another contention of Butler's was that of primacy and recency effects giving rise to the obtained profiles. In particular, a diatonic scale begins and ends on the tonic. While this might explain the higher ratings of the tonic note over others, it does not explain the pattern of similarity ratings obtained for the other notes in the scale. For example, the tones G and A occurred successively near the middle of the context sequence, such that according to Butler's contention, minimal difference in perceived fittingness should occur for these two tones. However, a difference was observed in the ratings, and in the direction predicted if listeners were engaged in rating according to a tonal hierarchy.

Another argument brought up by Butler and others (Cross, West, and Howell, 1991) is that diatonic scales are very particular musical stimuli, in which a variety of cues are being presented simultaneously. The argument can be exemplified by considering that it is not the notes in the scale per se which give rise to the perceived tonal centre, but the relationships between the notes, and their temporal ordering. If one were to scramble the temporal order of the notes, Krumhansl would predict the same pattern of ratings should emerge (assuming the tonal hierarchy was the sole operating factor influencing ratings, and that hierarchy was activated via the summing of note frequencies), while the intervallic rivalry theory would state that differences should emerge depending on the particular ordering. Brown (1988) attempted such a study, and found that listeners' judgments of key were, in fact, influenced by the temporal ordering of the context sequence tones, in terms of the intervals which were made prominent by a particular ordering. Brown's study, however, differed markedly from those investigating the tonal hierarchy. Her stimuli consisted of fragments of melodies, varying from three to ten notes in length. As well, subjects were required to sing the note which they believed represented the tonic. It is possible that these differences in methodology contributed to the differences in results. West and Fryer (1990), however, performed a study which more closely matched that of Krumhansl's original work than did Brown's. Listeners performed a probe-tone task, with the context sequences consisting of various random orderings of the notes in a particular diatonic scale. They found that even highly musically trained individuals did not consistently choose the 'true' tonic of the context sequence (as defined by the scale from which the notes were drawn) as being the tonic, indicating that temporal ordering was a factor in key identification. However, an interesting finding was that the third and fifth scale-notes, which were given high ratings according to Krumhansl's own work, were the most likely alternative 'tonics' chosen. So, while temporal ordering did appear to exert an influence on key identification, it did so only to a limited extent, since even with random orderings, listeners still differentiated the root-chord notes from other diatonic notes.

These criticisms do not necessarily invalidate the idea of a tonal hierarchy existing in listeners (although Butler and others seem to believe otherwise). Rather, it merely casts doubt on the activation mechanism proposed by Krumhansl. According to her, the hierarchy may be activated by the relative distributions of note frequencies in a piece of music. Brown's study simply indicates that Krumhansl may have been wrong in terms of the way the putative hierarchy is activated.

Another attack on the approach of Krumhansl and Shepard is that they do not provide any "extra-musical rationale" for why the tonal hierarchy should exist, and how it came into existence (Butler, 1990). Rather, it is founded on music theory, and does not provide any information beyond that already stipulated by music theorists. In other words, there is a circularity to the notion that the tonal hierarchy both developed from exposure to tonal music and the structure implicit in it, as well as giving rise to that same music. The key question is, how did tonal music come about in the first place? The cognitive structuralists do not answer this question, and in fact, don't seem to even address it. Their concern is with how people perceive music and whether that perception has any underlying organizational properties. Their conclusion is that people are sensitive to the structure of melodies as delineated by music theory, and provide evidence to that end, in the pattern of responses obtained in probe-tone tasks.

Possibly the strongest argument against the existence of a tonal hierarchy, paradoxically, came from Krumhansl's own data. She made the observation that the order of presentation of two notes for comparison will affect the resulting similarity perception. If the first note is more stable than the second, subjects' ratings will indicate this pair is less similar than if the order of presentation is reversed. Given that the same two notes will have different perceived similarities depending on the order of presentation, it doesn't seem possible to map out similarity with a static, spatial model, since it will not capture the dynamic quality apparently possessed by musical tones. The position espoused here accepts the existence of the tonal hierarchy is existent and functioning in listeners, and maintains that the ordering effect is a result of the usage of that hierarchy in performing these ratings tasks. A similar view was hinted at in Cross, West & Howell (1991): "Neither of these asymmetries are easily reconcilable with Shepard's model (although this could be because [Krumhansl's] subjects performed similarity rather than 'fittingness' judgments)" (p.215). In other words, subjects are forced to use the hierarchy in a way they wouldn't normally use it, such that the asymmetrical similarity ratings are by-products of a task requiring a dynamic usage of the hierarchy. This possibility will be further described in later sections, and forms the basis for the experiments reported.

As an alternative view, Butler (1989) claimed that listeners arrive at a sense of key on the basis of elements of a tone series that subjects arrive at a sense of key. More specifically, he contended that it is the less common intervals, the minor second, major seventh, and tritone, which offer unambiguous information concerning the key of a piece (Butler & Brown, 1984). The rationale here is that these so-called rare intervals, which occur infrequently as natural intervals between notes in a diatonic scale, point to a particular key as being the correct one. One should note here that Butler does not explicitly state that there is no such thing as a tonal hierarchy. He only says that in trying to determine the tonal centre of a melody, the intervallic relationships between tonal elements (and their commonality) are used rather than the summed frequency of occurrences of each individual tonal element. For example, an examination of the various intervals that may be formed between elements in a particular diatonic scale reveals that a tritone (notes separated by six semitones) can be formed only once in that scale (for example, B-F forms the only tritone in the C-major diatonic scale; no two other elements of that scale are in a tritone relationship). Furthermore, each diatonic scale has within it a unique tritone relationship, in that the actual notes will differ from scale to scale. According to Butler, if one were to hear a tritone occurring in a melody (e.g. the notes F and B occurring in a melody line), there exists only one possible diatonic scale (C-major, in this case) from which those two notes could have arisen, and that is the key subjects should infer. This assumes, of course, that listeners are not sensitive to other cues from which they might infer that a tritone is actually the result of a grace note or accidental. This issue will be pursued further when criticisms of Butler's model are discussed.

The intervallic rivalry approach has an advantage over Krumhansl's notion due to its direct reference to the temporal nature of music, in that listeners are assumed to be making comparisons between tone elements over time, and not merely summing and comparing overall durationally-weighted frequencies of occurrence. As a possible activating mechanism for the tonal hierarchy, Krumhansl stated that subjects might 'count' the occurrences of notes, giving extra weight to notes sounded for a longer duration, since frequency of presentation appears to be highly correlated with its position on the tonal hierarchy in real music. That is, frequently presented notes are indices of highly stable elements in the hierarchy, and it is in performing this tacit counting task that subjects arrive at a sense of key. Butler pointed out that this system is not sensitive to ongoing shifts in local key, such as when a melody line modulates to a new, albeit temporary key. In this case, Krumhansl might have to resort to a local-global key assignment system, where subjects would take samples of some duration, perform a count, and assign a key, based on the given information concerning local note frequencies. A potential difficulty with this is in determining what size these sampling chunks might be. One might have to compress the size of the chunks to the point that the key begin to resemble those proposed by Butler, if the chunks are, for example, two notes in size. Against the necessity of such a drastic modification is the evidence that when such a melody, consisting of rapid changes in key (or scale structure) is presented for pleasingness ratings, a corresponding drop in ratings occurs with increasing key alterations (Cross, Howell, & West, 1983). So, in order for a sequence of notes to be considered a coherent melody at all, it appears that conformance to a particular scale structure (with a particular tonal centre) over time is a necessary constraint. In fact, direct tests of "melodiousness" of a melody with varying degrees of conformity to diatonicity support this claim (Cuddy, Cohen, and Mewhort; 1981).

So far, there have been few published criticisms of Butler's model, but several are possible. For example, Butler's claim that his model doesn't require listening to the entire piece of music, as Krumhansl's does, is questionable. As a thought experiment, consider a melody consisting primarily of root-chord notes, and employing only common intervals as delineated by intervallic rivalry theory. Using Butler's strategy would necessitate hearing the entire melody, since, in Butler's own words, "Any tone will suffice as a perceptual anchor - a tonal center - until a better candidate defeats it." (Butler, 1989; p.238). The listener will not be sure of what the absolute best candidate for the tonal centre is going to be, unless they hear the entire melody. In that respect, then, this criticism applies equally well to intervallic rivalry theory as to the theory of the tonal hierarchy. In addition, Krumhansl & Schmuckler (1986) developed an algorithm based on the use of note frequencies to activate the tonal hierarchy. The result was a high rate of success at detecting the 'true', or written key of the piece, using as few as the first three or four sounded tones, indicating that Butler's contention is a weak one.

The issue, though, shouldn't be when and whether listeners are absolutely positive about the key of a melody. Rather, the success of any given model at predicting the perceived key of a stimulus should be the primary index of validity. In this sense, it appears that the tonal hierarchy and intervallic rivalry models are equally valid, under certain conditions. With minimal tonal information (such as a three-note passage), hierarchic structure might be overshadowed by the interval structure. If, however, the stimulus conforms to the rules and regularities found with most melodies (repetition, begin/end on tonic) then it is very possible that the tonal hierarchy plays a more significant role, possibly overshadowing that of rare-interval information. Such would appear to be the case for any piece of music employing grace notes or accidentals. Although this prediction has not yet been directly tested, there are some supportive hints in the available literature.

Krumhansl (1990) has pointed out the fact that 'rare intervals' are intervals that rarely appear in music. In written works by the likes of Bach, Chopin, and Shostakovich, the tritone occurs extremely rarely, and its appearance is relatively late in the melody, presumably long after listeners have identified the key. If subjects must wait for a rare interval before deciding on a definitive key, they might have to postpone their decision for at least as long as would be the case if they were summing note occurrences in building up (or activating) a tonal hierarchy. In fact, empirical evidence of intervallic rivalry playing a role in determination of the tonal centre has primarily been obtained in those situations where the 'melodies' are two- or three-note sequences. If exposed to a melody with no rare intervals, listeners employing a key-finding algorithm proposed by Butler should have a great deal of difficulty identifying the tonal centre of the piece. For example, the melody "Mary Had a Little Lamb" consists of five intervals, none of which are considered "rare" (major seconds, minor thirds, major thirds, perfect fourths, perfect fifths). Therefore, when presented with this melody, listeners relying on a rare-interval algorithm should not be able to definitively identify the tonal centre of the piece.

Another potential concern with Butler's model has to do with its "dynamic nature". Butler repeatedly contends that his algorithm is dynamic, with listeners using note information over time, whereas Krumhansl's is static, employing no temporal information whatsoever in obtaining a key center. Although both descriptions are technically true, further examination of Butler's model reveals it to be as static as Krumhansl's. Butler claims that listeners are sensitive to the rare intervals, and makes explicit the proposition that they possess an interval hierarchy of sorts (Butler, 1989; Butler, 1990), and must hear a rare interval in order to be relatively certain of the key. However he does not make any definite predictions concerning how the temporal ordering of notes in a melody might alter the effect of rare intervals, stating only that order does play a central role (Butler & Brown, 1984; Butler, 1988). His sole formalization of this is to point out that a rare interval in a melody might be highlighted or obscured, depending on the relative proximity of the component tones (Butler, 1990). However, his claim of listeners possessing an interval hierarchy implies that listeners, instead of 'sitting around counting notes', are 'sitting around waiting for a rare interval'. The main differences between Butler's and Krumhansl's positions appears to be whether the listener is using one or two notes at a time in key identification, and whether a rare interval or frequently heard tones influence key identification. In other words, Krumhansl's algorithm samples each note event individually and adds that single event to the 'frequency of occurrence' database, while Butler's algorithm compares adjacent notes for their rarity in diatonic scales. The rare-interval hypothesis, therefore, is dynamic only in the sense that the stimulus contributing to key identification consists of two notes presented sequentially.

An examination of composed music and its perception reveals another possible problem with intervallic rivalry, the widespread use of musical forms known as 'accidentals' and 'grace notes'. These are notes which are not part of the diatonic scale of the melody in question, and which usually resolve to (followed immediately by) a diatonic note. Recall that in describing Butler's notion, the claim was made that the key information is provided by the occurrence of rare intervals between elements in a melody. This claim is weakened by the existence of accidentals. Presented with a tonal melody containing an accidental, listeners using Butler's mechanism would be left in a state of confusion concerning the tonal centre of a piece when the accidental appears, since that note will likely form a tritone or a semitone with another element in the melody, a tritone which would be from a key quite different from that intended by the composer. For example, listeners might mistakenly assume from the opening notes of Beethoven's "Fur Elise" that the song is actually written in the key of D-flat minor, or B-major, instead of the key notated by the composer (A-minor), since the first five notes consist of the repeated presentation of a minor-second interval found in those 'false' keys. One might argue that listeners could discern grace notes from melody notes based on other features such as duration or loudness, according them 'nonmelody-note' status. While this might be true for some situations, it does not address the issue when these cues either do not vary between accidentals/nonaccidentals, or when the cues are in direct opposition, such that the accidental is held for longer than the surrounding diatonic tones.

To more clearly illustrate this possibility, let us perform a thought experiment based on the brief melody C-G-F#-G-E-D-E-C. In this case, both the presence of a tritone (between F# and C) and a semitone (between F# and G) should, according to the intervallic rivalry method of key identification, lead subjects to the unambiguous impression that the melody is in the key of G-major. Although Krumhansl's method would also fail to point to a definitive key, her prediction would be towards C-Major and away from G-Major. Looking at relative frequencies of occurrences, there are two C's, two G's and a single D, A, and F#.

According to Krumhansl's tonal hierarchy, the tones making up the unison, major third, and perfect fifth are most stable, the major second, perfect fourth, major sixth and major seventh are somewhat less stable, and the remaining elements least stable. Given this, the elements can be summarized in terms of the frequency distributions of notes from the three levels of the hierarchy:

High stability	Moderate stability	Low stability	
6	1	1	(C-major hierarchy)
3	5	0	(G-major hierarchy)

So, Krumhansl's model predicts that subjects would be more likely to hear the melody in C-major than in G-major, if forced to choose. This is based on the fact that when the perceived key is C-major, there are consistently higher relative frequencies of occurrences in the 'most stable' category, which carry the most weight and are the focal notes in determining the key. Butler, on the other hand, would more strongly predict G-major as being the most likely candidate for key, on the basis of the presence of either the tritone of that key (C-F#), or the semitone (F#-G). In fact, informal testing revealed that all of the dozen listeners sampled heard the sequence in C-major, when asked either to state whether the C-major or G-major root-chord better represented the key, to sing the root note of the melody, or state whether the melody sounded 'finished'.

One might justifiably argue that the stimulus in question is heavily biased toward listeners hearing it in the key of C-major, since it begins and ends on the note C. That is, were a different temporal arrangement employed, listeners might report the stimulus as being in a different key. While the order of presentation does have a demonstrable influence on key perception (Brown, 1988), it is likely that part of this influence has to do with a listener's sensitivity to musical regularities (most songs begin and end on the root note) rather than any underlying key-determining mechanism based on constituent tonal elements. Furthermore, it could also be argued that the close proximity of the semitone in the example provided above should serve to bias listeners toward choosing the note G as the tonic. Were this the case, the melody should have sounded 'unfinished' to listeners, ending on the fourth degree of the G-Major scale.

The point of this thought experiment is that while Butler contends rare intervals point to unambiguous key information, it is possible that information can be overshadowed by tonal stability functions based on frequency of presentation in a more natural musical setting than that employed by Butler.

One final critique of the intervallic rivalry hypothesis is of a more theoretical nature. Butler maintains that tonality apperception is a result of sensitivity to time-order dependent qualities between elements in the musical stimulus. The question that arises at this point is, 'to what end'? In particular, what happens perceptually, once the tonality of the piece is decided? Butler neatly side-steps the issue, by defining tonality in terms of the proposed methods by which it might be perceived. His only comment with regard to this issue is that "...tonality usually seems to play the role of perceptual ground rather than figure" (Butler, 1989, p.238), a statement implying that once the key has been decided upon, it does little else to influence perception. Conversely, he might be alluding to the system proposed here, which is simply that tonality affords the listener the opportunity to use the tonal hierarchy in organizing and structuring the musical stimulus. If not, then it is unclear to this author what purpose or function tonality serves. It is known that listeners perceive atonal passages as disjointed, disconnected sets of tones as indexed by 'goodness' and liking ratings, while tonal passages are perceived as self-contained, connected, goal-directed melodies (Cross, Howell, & West, 1983). If one chooses to ignore these points, then it becomes necessary to derive new explanations for the purpose of tonality, as well as for the bulk of data accumulated by Krumhansl and others, pointing out the existence of these stability functions. Of course, it might be that tonality per se serves no purpose to the listener, but this seems obviously wrong in light of the available evidence of the appreciation/perception of 'atonal' works (pieces with no defined tonal centre) by a majority of listeners.

C. Group-theoretic accounts

Balzano (1980), as well as several other music psychologists (Browne, 1981; Cross, West and Howell, 1991) take a different approach to the entire issue of what factors play a role in perceptual structuring. They contend that it is the features particular to diatonic interval scale structure which provide the necessary information about key, and not anything to do with a tonal hierarchy. According to them, Krumhansl's context stimuli were confounded, in that the scales or broken chords were specific musical stimuli, with their own properties and features above and beyond the invocation of stability functions. For example, a diatonic scale consists of tonal elements in unique relations to one another, in terms of the intervals formed between them. When a diatonic scale is used as a context stimulus, then, it is the structure of the stimulus which provides key information, and not the specific durations or frequency of occurrences. That is, according to this account, it shouldn't matter so much how often a particular note is sounded in a stimulus, when determining the key of the stimulus. Rather, what matters is the interval relations formed by all the notes, which in turn dictates from what pitch-set the notes might have been drawn (diatonic, chromatic, etc.). This is similar to Butler's contention of rare intervals being important in determining key, however here it is not stipulated that 'rare intervals' are necessarily the operating factor in determining key. Upon examination of the probability of occurrence of three-tone pitch-class sets (PC-sets) in the diatonic scale, it was found that subjects' accuracies for identifying 'wrong' (non-diatonic) probe tones increased with increasing 'commonality' (number of distinct ways the set can be constructed from the notes in the scale) of the PC-set in the diatonic scale (Cross et al, 1991). While the 'major-triad' set (tonic, mediant, dominant, for example) was associated with high detection accuracy, the highest level was achieved with the set consisting of tonic, supertonic, and subdominant (first, second, and fourth scale degrees). This finding would not be predicted by either Krumhansl or Butler, since Krumhansl would have predicted the highest accuracy to occur for the PC-set consisting of the tonic, mediant, and dominant (root chord), while Butler would have predicted highest accuracy for PC-sets with rare intervals (tritone, minor second). In light of this evidence, Cross et al (1991) state, somewhat tentatively, " It may be that the tonal hierarchy is not relevant to pattern-matching tasks...." (p.229). What they neglect to point out is that rare intervals don't appear to be relevant to pattern matching tasks either, at least according to their data.

After reading this literature, however, one cannot help but wonder if this approach shares the weaknesses of the psychophysical approach. The stimuli were highly impoverished, such that one has a hard time generalizing from that situation to the real world of cue-rich music. The group-theoretic account for determining key may have been reflected in the data simply because stronger musical cues to indicate key were lacking. In other words, presenting three notes to subjects is not the same as presenting a full-fledged, multi-phrase melody. Although such studies have been performed (Cross, Howell, and West, 1983), they did not specifically require subjects to identify the key of the piece. Rather, they asked for preference or musicality ratings when scale structure was manipulated. The conclusions drawn, then, pertain to scale structure, independent of mode, being an operating factor in listeners' perception of music. They cannot, however, draw conclusions concerning the influence of scale structure on key identification or tonality perception, unless one assumes that preference ratings are indices of such perception. The group theoretic account, then, may have limited generalizing capabilities; subjects had to latch onto something to determine key, and these experiments made sure that the PC-set was the sole possibility. While not a documented finding, it is intuitively obvious to most researchers that subjects will actively search for any cue which will enhance performance on tasks with a potentially correct/incorrect response requirement. Showing that subjects are sensitive to a factor in a situation when it provides the only available information does not mean that factor will exert an influence when the listener is faced with a more complex, multi-varying stimulus. When hearing streams designed to be more externally valid, the role of PC-set information on perception is unclear, and might be overridden by a more salient cue or set of cues (such as element stability, or frequency of occurrences).

In spite of its limited explanatory power in a cue-rich musical setting, the group-theoretic viewpoint does make a rather significant contribution to the understanding of the origins of diatonic structure in music, and is reconcilable with the ideas expressed by cognitive-structuralists. This is made possible if one assumes that both are part of a larger picture of music perception, each making a contribution to a different portion.

Balzano (1982), in examining the mathematical properties and regularities of pc-sets independently of any music-theoretic constraints, concluded that the diatonic scale, or 'diatonic set', embodies certain characteristics which no other sets of tones possess. In particular, three properties, uniqueness, coherence, and simplicity, are captured by the diatonic set, and only the diatonic set. Uniqueness refers to the fact that each element in the set is related to the remaining elements by a different set of intervals. That is, for the diatonic set represented by the C-major scale, the note C forms a unique set of intervals with the other members of the scale, relative to that set of intervals formed between D and the other tones of the scale, E, and so on. Coherence refers to the notion that the scale-step distance (number of intervening semitones) between any two adjacent members of the set will be smaller than the scale-step distance between any three adjacent members of the same set. For example, again in the C-major scale, the scale-step distance between the tones C and D is smaller than the distance between C and E. Finally, simplicity refers to the fact that by changing only a single element of a particular member of the diatonic set, one transforms the set into a new member. For example, by changing the fourth scale degree, in this example the tone F in the C-major scale, to F-sharp, one creates a new diatonic scale, this time in the key of G-major. To change from G-major to a new member of the diatonic set, one need only once again sharpen the fourth element in the ordered set (the note C, in this case), resulting in the scale of D-major, and so on.

Based on this information, it seems plausible to conclude that the reason the diatonic scale was so forcefully embraced by Western composers, and has endured for so long, was as a direct result of being the only subset of the chromatic scale containing these special properties. Furthermore, it is possible that once the diatonic set is employed to write melodies, it gives rise to the hierarchic qualities as depicted by Krumhansl, with specific tones relegated positions of prominence or subordinance, depending on their position within that set. Presumably, listeners come to incorporate the hierarchic structure as a result of enculturation and exposure to music written from within this set, the result being that the structure itself has an impact on perception of musical tones (one finds a similar argument in Cross, West, & Howell, 1991).

Parncutt (1989) correctly points out that this explanation assumes the existence of the chromatic set as its starting point, and does not offer any explanation as to how this set came about, and how it came to be accepted as a musical standard in the first place. He mentions that the JND for frequency is a great deal smaller than the adjacent tones in the chromatic scale, and argues that other divisions of the octave were equally likely. However, he conveniently goes on to demonstrate that the existence of chromaticism is nicely handled by examination of the psychophysical properties of complex tones, thus relieving Balzano of that burden. Insofar as answering the question of why the chromatic set was chosen over other possibilities, one might appeal to the musicological consideration of utility of a tuning system versus its complexity. Ideally one would like a tone set which is diverse enough to allow for a wide variety of scale-systems, while keeping complexity at a minimum to ensure listeners will be able to understand what is going on. Pursuing this line of thought, however, would extend the discussion far beyond the topic at hand, and will be left to music theorists.

1.6 Interim Summary, and a Proposal

The multidimensional nature of musical stimuli has given rise to several multidimensional models of pitch representation. At one level, features of melodies that might be independently manipulable relative to other features are not necessarily represented independently. At another level, it appears that tonal information in a non-musical setting is perceived differently from when it appears in a musical context. The tonal hierarchy, contrary to the other models described earlier, has as an implicit assumption that a musical context is necessary for its activation. The more complex and rich the stimulus, the more strongly the tonal hierarchy should be activated. It is primarily when impoverished stimuli are presented, such as those employed in Butler's and Balzano's work that the hierarchic nature of each tonal element is not evident, or at least that the possible effects are hidden or overshadowed by other cues. A two- or three-note fragment, then, is simply not enough of a tonal context to invoke a stable tonal hierarchy. It must also be stressed that Krumhansl did not intend to demonstrate that the tonal hierarchy was the sole factor responsible for pitch perception in a tonal context, only that it was one of several which must be taken into account. This was explicitly stated at the outset of her reply to Butler "...the tonal hierarchy is just one component of experienced listeners' abstract knowledge of relations among tones, chords, and keys demonstrated by experimental studies." (Krumhansl, 1990; p.309).

To date, the debates concerning key identification and pitch representations do not appear to be settled to anyone's liking, such that no one model appears to have a real advantage over the others in terms of parsimony and explanatory power. Although Butler contends that rare intervals point to key-identification, and virtually denies the existence of a tonal hierarchy altogether, common pc-sets (which do not have rare intervals between elements) have been associated with higher detection rates of out-of-tune notes. Tonal hierarchy models, on the other hand, are inextricably linked with the way music itself is organized, and offer no explanation for why the music is organized in that way, although it seems clear from the data that people are sensitive to that organization. Finally, although group-theoretic accounts are useful in explaining why the diatonic set is so enduring, to date, the research confirming this model does not convincingly demonstrate that pc-set information plays the main role in musical pitch perception.

Browne (1981) alluded to a possible resolution to this controversy. Rather than focus solely on the stimulus features as the operating factor in music perception, Browne also took into account the particular task required of the subjects, proposing that the latter dictates which subset of the former will be predominantly influencing perception. He discussed the possibility that there are two processes which subjects can engage in when performing a listening task: pattern matching and position finding.

Pattern matching, according to Browne, is the process engaged when subjects attempt to answer the question 'is this tone sequence the same as that?', while position finding addresses the question 'what is the key of this tone sequence?'. According to Browne, these processes are independent of one another, and are sensitive to different aspects of musical stimuli: Pattern matching primarily makes use of the intervallic information in a melody, while position finding makes use of a tonal hierarchy. For example, if a listener was presented with two versions of "Happy Birthday" in different keys, s/he would be engaging in a pattern-matching task, and therefore should not be influenced by any key difference in comparing them for similarity (which would set up a different tonal hierarchy, but leave the intervallic relationships intact). If a listener is presented with a single version of Happy Birthday, and asked to judge the key, s/he would be engaging in position finding, and would be influenced by the tonal hierarchy. He further proposes that rare intervals are more useful for position finding tasks, while common intervals are more useful for pattern matching. The appeal of Browne's proposal is that it takes into account both the stimulus to which a subject is exposed, and the task the subject is required to perform.

One criticism of Browne's notion concerns his statement that the tonal hierarchy does not have an effect when a listener is engaged in a pattern-matching task. The problem with this statement, pointed out earlier in the discussion of intervallic rivalry, is that it assumes that even if a melodic stimulus were to activate a tonal hierarchy in a listener, the subsequent tonal events would still be processed equally, and equally well, relative to one another. This may not be the case at all. Furthermore, Browne's statement assumes that subjects must be sensitive to one or the other features of melodies, but not both, while engaging in a specific listening task. This ignores the fact that human beings are constantly seeking order and patterns in their environment, whether or not those patterns truly exist (Gould, 1991). The results from the study of human perception of randomness indicate that, in general, we are not very effective at either detecting or producing random sequences. One rather important orderly feature of Western music is that it can be heard from within a structured tonal framework, complete with primary reference tone and subordinates, as described in music theory. In his defence, Browne does not deny that listeners will have access to a tonal hierarchy when engaged in a pattern-matching task. He only denies its utility to a listener when comparing whether 'this' is the same as 'that'.

Contrary to Browne's strong stance, it is possible that both intervallic and tonal hierarchy information are being used by subjects, but at different times in a listening task, and depending on the information available in the stimulus array. When faced with the problem of rating a melody for pleasingness, or rating the similarity of two melodies, subjects might first use Butler's (or Balzano's) strategy to arrive at a tonal centre (in their discussion of Browne's ideas, a procedure akin to this is alluded to in Cross, West & Howell, 1991), or, they might use frequency of occurrences, if the melody contains a high degree of repetition. Either way, once that tonal centre is activated, the tonal hierarchy thereafter directly provides information concerning salience and relative importance ascribed to each incoming note event. Janata and Reisberg (1988) entertain a similar notion, by proposing that the tonal hierarchy guides expectations concerning upcoming events, such that some notes are more or less expected than others, depending on their position in the hierarchy. Given that a musical sequence will contain a variety of rare and common intervals (unless specifically designed to avoid this), it follows that subjects exposed to such a sequence will be using both types of information (intervallic, and tonal hierarchy), possibly in a manner described above. Butler himself contends that listeners will select "...a most-plausible tonal structure based on the evidence at hand." (Butler, 1990; p.16), although he is reluctant to grant that event hierarchies have anything at all to do with it. In their excellent review of the cognitive correlates of tonality, Cross et al (1991) arrive at a similar conclusion, although they don't explicitly state how these two facets of melody interact:

> "In a paradoxical way, the dynamic qualities of atemporal diatonic interval structure appear to be involved in a fruitful interaction with the static qualities conferred by the...tonal hierarchy. One can speculate that the cognitive representation of musical pitch is best construed as being multi-levelled and that any formal and computable model of the cognitive organization of music pitch would have to take all of these possible levels into account." (p.238-9).

Placing the tonal hierarchy in the sort of post-key identification role mentioned above eliminates many of the criticisms pointed out by Butler. Most of his contentions deal with the idea that the tonal hierarchy is not necessarily used to identify the key of a piece. If we grant that it may not function at that point in melody encoding, but that its utility becomes prevalent following key identification (in the form of providing the listener with an algorithm to recognize the importance of each incoming note event and organize the sequence into structurally relevant or irrelevant events), Butler's criticisms lose much of their strength, leaving intact the idea that the tonal hierarchy is functional in the perceptual organization of melodies.

Such a notion of multiple or mixed models being necessary to explain a perceptual phenomenon is not new to psychology. Both external to and within the auditory domain, sensory processes have been seen to proceed in more than one way, depending on the conditions of the situation. For example, the idea that more than one mechanism is necessary to explain something as fundamental as the neural code for pitch (place or temporal coding, depending on the tone's frequency) has received widespread agreement, following much debate over one or the other model being 'the' correct one (Moore, 1982). Such a compromise is being proposed here, wherein both the tonal hierarchy and the nature of the intervallic relationships formed by constituent elements in a melody may be functional in the perception of that melody. Which model has more of an influence is presumably determined by stimulus factors (repetition, number/salience of rare intervals), and listening orientation (same-different comparison, pleasingness/keynote ratings). If there is a high degree of repetition, then this might directly activate a tonal hierarchy in listeners. If there were less repetition, the interval structure (and possibly rarity) would provide information concerning the specific key, which would thereafter give rise to the same hierarchic perception.

In summary, the proposal outlined above stipulates that listeners will attend to information from several different sources in arriving at an understanding of what they are hearing. Initially, stylistic regularities would provide the first hints as to the key of a piece (beginning/ending tones). Once more information is provided via the introduction of more tones, the listener may or may not alter that perception, depending on the distribution of occurrences of tones (Krumhansl's proposal), as well as the interval relationships that are formed between elements (Butler's proposal). Upon deciding on the key of the melody, the listener's knowledge of the dynamic quality of tones in that tonal framework (in terms of their relation to the tonic) would thereafter influence perception of those tones, allowing listeners to organize the information into structurally important or subordinate positions.

One potentially damaging criticism of geometric modelling of pitch representations still remains, however: The asymmetrical similarity ratings, which Krumhansl herself obtained.

1.7. Asymmetrical similarity perception of tones/melodies

A perceptual phenomenon first pointed out by Krumhansl (1979) has received relatively little direct attention in the research literature, which is surprising considering the theoretical implications to which it gives rise. Specifically, when subjects are engaged in a probe-tone task, and are asked to rate the similarity of two tones which follow a context sequence, the similarity rating given by subjects depends to a significant extent on the order in which the two tones are presented. If the first tone is from a more stable level of the tonal hierarchy than the second, the two tones are considered less similar to one another than if the order of presentation is reversed (less stable followed by more stable).

At this point, one could argue that if similarity perception is influenced by order of presentation, it obviates the existence of any static hierarchical structure with fixed distances between nodes (Butler, 1989; Cross, West, & Howell, 1991). Such a contention is not new to cognitive psychology, where geometric models representing underlying similarity in observers have been shown to be inadequate, due to the emergence of asymmetric relations (Tversky, 1977). However, with respect to the 'geometric model' presented by Krumhansl (1979), such an argument is based on the incorrect notion that the conical structure is itself the tonal hierarchy. If, instead, it is made clear that the structure is a reflection of differing stability weightings assigned to incoming tonal elements, the argument becomes invalid. That is, the structure is the result of listeners' knowledge of the dynamic relations between constituent tones and the tonal framework in which they appear. The structure does not represent the mechanism directly; rather, the obtained pattern of ratings is merely the outcome of the mechanism's operation in this task.

A probe-tone task is not comparable to the act of simply listening to a melody. Subjects are forced to perform a similarity comparison between two isolated tones, albeit tones which follow a discrete tonal context. When put into this situation, listeners more than likely will engage in a form of processing that is not required when rating a melody for its pleasantness, or even when simply listening to music on the radio. Of interest, then, are those aspects of the task that might be producing the obtained asymmetrical similarity perception.

One likely possibility is that the task itself, in which two stimuli are presented for comparison, was influential in the result. If it is assumed that listeners were making use of the hierarchical structure, then the way in which that structure was accessed might be responsible for the obtained pattern of ratings. The subject first hears a tonal context, which apparently activates a tonal hierarchy for that key. Once that hierarchy is activated, however, the listener is now comparing two elements of that hierarchy. It is here that the task departs from that of a simple listening task, in that the comparison process involves a dynamic usage of the hierarchy.

When hearing a melody, subjects supposedly make use of the hierarchy by identifying incoming note information as either important to the structure of the melody, or not important, depending on that note's position on the activated hierarchy, regardless of how it came to be activated (possibly through Butler's intervallic rivalry system). The tonal hierarchy would thereafter serve the function of 'separating the wheat from the chaff' so to speak, in that the information is weighted in terms of its stability in the given key (Shepard, 1982). According to this notion, then, listeners are actively processing the incoming note information in an effort to sort out and hear the structure and organization of the melody. If faced with the somewhat artificial task set up by Krumhansl, however, the effect of the tonal hierarchy is different from this, because listeners would now be forced to compare particular elements to one another, rather than simply perceive and interpret each element in terms of its stability in the tonal context. Since these elements have an associated stability weighting attached to them, it is possible that this information factors into the resulting similarity percept. In other words, listeners engaging in Krumhansl's study may have been performing their similarity ratings based on relative similarities of the two elements to the context key. As an analogy to this, suppose an observer were asked to rate the similarity of two objects, an apple and a baseball. In the context of 'round things', these two objects would receive a higher similarity rating than if the context were 'sporting equipment'. The context defines what qualities of the object are important, thus influencing their perceived similarity to one another. In a similar fashion, the musical context does not simply define, but rather determines the dynamic quality of the tones, in terms of their relative positions in the tonal hierarchy. It is likely that this dimension had a significant influence on the similarity ratings that were obtained, such that the relative stabilities of the tones in the context influenced their perceived similarities. As far as how these perceived stabilities specifically had the asymmetric influence described earlier, one can find reference to several models developed from within cognitive psychology, to explain similar asymmetries observed in the visual domain.

In an early study of asymmetric relations, Rosch (1975) proposed that within a general category, there exist certain referent members to which other, non-referent members of the category are compared. She called these referent members 'cognitive reference points', and set out to demonstrate that they serve as prototypes of the category. These prototypes, in turn, should lead to asymmetric perceived similarity when presented with a non-referent member for similarity judgments. On a variety of tasks (completion of verbal 'hedge' statements such as "______ is essentially _______", or physical placement of two objects so as to reflect psychological similarity) and stimuli (numbers, colours, line orientation), Rosch found that asymmetric similarity emerged. For example, if faced with the task of filling in the hedge statement with the numbers 100 and 103, subjects showed a significant tendency to consider the latter number as being part of the former, such that "103 is essentially 100", rather than the other way around. As well, if asked to place the numbers on a board in order to reflect psychological distance, subjects put the 103 nearer the 100 when the latter was already fixed to the board, relative to when the 103 is fixed. If neither member (or both) were reference point stimuli, no asymmetries emerged.

Rosch argued that categories within a general stimulus set (colour, number) emerge as a result of individuals first recognizing that there are cognitive reference points, and thereafter seeing nonreference stimuli as deviations from them. Categories, then, result from a clustering of nonreferents around a central prototypical stimulus. Music theory tells us that the tones constituting the root chord serve as central indicators of the key of the piece, while the other non root-chord tones are less indicative. If it is assumed that the root-chord tones are cognitive reference points serving to strengthen the impression of key, then the asymmetric ratings obtained by Krumhansl might be explained using Rosch's proposal. The presentation of the two tones in a particular temporal ordering is an equivalent manipulation to asking subjects to mentally 'place' one tone in psychological space relative to the other. Root-chord – non root-chord tone orderings would then be associated with lower perceived similarity than when that ordering is reversed, largely as a result of listeners perceiving root-chord tones as prototypical members of the diatonic framework specified by the context.

Tversky (1977) provided a slightly different explanation for these asymmetric patterns. He believed that asymmetric relations are not dependent on one member of the to-be-compared stimuli being a cognitive reference point. Instead, he claimed that the asymmetry is a result of salience differences between the two stimuli, such that if a less salient member of a category is followed by a more salient member, perceived similarity is higher than when the reverse occurs. He explained Rosch's data by stating that the cognitive reference point stimuli were more salient members of the categories in question than the non-referent stimuli, and that this was the operating factor in producing the asymmetry. Salience, in turn, is determined by examination of the set of distinctive versus common features, which is enhanced by the observer's focal attention, or which object they see as the 'subject' versus the 'referent' in the comparison. In other words, when asked to 'compare object X to object Y', observers will recognize X as the subject, and Y as the referent. Consequently, the perceived similarity between the two objects will be more influenced by the distinctive features of X, than by any distinctive features of Y.

In the case of the tonal hierarchy, it was earlier proposed that the tones each receive a weighting according to their relation to the tonic of the melody. Tone profile studies have shown that tones closely related to the tonic ('near' tones) receive a higher weighting value (as indexed by 'belongingness' ratings) relative to tones more distant from the tonic ('far' tones). Music theory dictates that these same near tones are considered the most important/prominent for tonal music. In Tversky's terminology, then, the near tones might be considered highly salient, whereas the far tones would have low associated salience. Also from Tversky, when two tones are presented for similarity comparison, the fact that the two tones must be presented serially gives rise to a directional task, such that the tones would be 'X' and 'Y' from the example above. So, the first tone's salience should have more of an impact than the second. When the presentation order is near-far, the high salience of the near tone would be further augmented by it being seen as the subject, resulting in lower similarity ratings relative to the reverse presentation ordering.

Unfortunately, neither of these accounts of asymmetric relations can fully explain Krumhansl's data. Krumhansl herself pointed out that Tversky's account would predict that the degree of similarity between two elements should be more influenced by the distinctive/common features of the first stimulus than the second, due to that stimulus being seen as the 'subject' of the comparison. However, her analysis indicated that across pairs of tones, the second tone's identity accounted for more variability in the ratings than did the first.

A more serious difficulty with applying Tversky's model to these data is that one cannot be certain that listeners in Krumhansl's task actually did see the first element as the subject, and the second element as the referent. In fact, one could construe the situation as the reverse. If presented with two tones in succession, it is very possible that listeners considered the first tone the referent, to which the second tone was compared (the subject), in which case Krumhansl's data do not conform to Tversky's predictions. A similar objection can be raised with respect to Rosch's account of asymmetric similarity. The listeners, according to Rosch, are mentally 'placing' one tone in psychological space relative to the other. However, the ordering of the two tones appears to imply that the first tone is 'already on the board', and the second tone is being mentally placed in relation to it. If that is so, then in the case of near-far stimuli, the far tone should have been placed closer to the near tone than if the reverse ordering were presented. Since the opposite result was obtained by Krumhansl, the only conclusion which can salvage Rosch's proposal intact would be to assume that subjects were actually placing the first tone in psychological space relative to the second, and not the other way around. The only evidence we have that listeners interpreted the stimuli as subject--referent (in Tversky's case), or placing the first in relation to the second (in Rosch's case) and not the reverse is the obtained pattern of ratings which each model was supposed to explain. The circularity of this makes both Rosch's and Tversky's proposals unappealing.

Finally, Rosch's notion of 'cognitive reference points' does not seem well suited to account for Krumhansl's result. In particular, it is not clear that root chord tones conform to Rosch's description of cognitive reference points. While root chord tones might serve as referents for the key of a melody, and the other diatonic tones are somewhat less referent, it does not follow that the relation between these two sets is in the form specified by Rosch. That is, while music theory dictates that root chord and non-root chord tones differ in terms of their representativeness or stability within a given key, there is nothing in the music theoretic description of these sets of tones which indicates that one set serves as a deviation from the other. A cognitive reference point is considered a prototypical stimulus around which cluster other, less prototypical elements. There is no reason to believe that root chord tones are actually seen as 'prototypes', and the remaining diatonic tones are considered deviations from that set. Music theory dictates that this is the case, but stating something does not make it a scientific fact. In short, it appears that the prevailing cognitive accounts of asymmetric similarity do not adequately explain the results Krumhansl obtained. However, there exists a potential explanation for her asymmetry, which does not rely on any underlying constructs.

It may be that her asymmetric ratings were in part due to an effect described by Bharucha (1984), known as anchoring. When a nondiatonic note is sounded in a piece of music, the immediate impression on the listener is that the note is 'jarring', or wrong. If, however, that nondiatonic note is immediately followed by a diatonic note of close frequency proximity, a retroactive 'smoothing over' takes place, and the nondiatonic note is perceived to be more consonant within the tonal context than if it was not followed by the diatonic note. Krumhansl's study (from which the conical structure was obtained) was set up such that following the context sequence, two tones were presented for comparison. It is possible that part of the reason the asymmetry was obtained was because anchoring was taking place in one case (nondiatonic-diatonic), and not the other (diatonic-nondiatonic). Since, in the first case, the nondiatonic tone is anchored, it would sound less jarring, and hence more similar to the diatonic tone. Of course, this explanation only holds for those comparisons involving two tones close together in frequency proximity, and cannot account for the full range of asymmetric relations Krumhansl obtained. However, a different testing method is necessary to avoid this possible confound. A likely candidate is to present entire melodies for comparison.

There has been, to date, only one systematic study appearing in the literature that employs melodies as stimuli when investigating asymmetric relations. Bartlett and Dowling (1988) presented pairs of melodies differing in terms of their diatonic scale structure by a single note. There were two types of melodies presented: melodies in which all notes were drawn from a single diatonic scale (S, or scalar melodies), and melodies in which a single note violated the scalar structure set up by the other notes in the melody (N, or nonscalar melodies). These two melodies were then paired in the four possible combinations, resulting in the trial types SS, SN, NS, and NN. These four trial types were presented to subjects for similarity comparison. The important result was that SN melodies were considered less similar to one another than NS melodies, analogous to Krumhansl's finding at the level of individual note comparisons.

Their explanation of this asymmetry is known as the 'perceived alternatives' hypothesis. Based in part on Garner's (1970) notion that "good patterns have few alternatives", they suggest that when one is exposed to a melody, it invokes a hypothetical set of alternative melodies to which the current one belongs. One could simplify this by saying the perceived alternatives are those melodies which could reasonably substitute for the presented one. The prime constraint determining membership to a set is that the alternatives consist of notes drawn from the scalar structure of the original melody. If the scalar structure is diatonic, then the set of alternatives are also only diatonic. If the scalar structure is chromatic, then the set of alternatives consist of both chromatic and diatonic melodies, since the latter is a smaller subset of the former. Given this constraint, they state that the asymmetry is a result of comparing the second melody in the pair to the set of alternatives invoked by the first. If the second melody is part of the set, similarity ratings will be higher than if it isn't part of that set. In other words, SN pairs should have lower similarity ratings associated with them than NS pairs, since in the former case, the second melody is not part of the set invoked by the first, while in the latter case, it is.

There are several potential problems with the perceived alternatives hypothesis which deserve consideration. First, the data obtained by Bartlett and Dowling do not always conform to the predictions made from the perceived alternatives hypothesis. In particular, they found that NN pairs received significantly lower similarity ratings than did SS or NS pairs. According to the perceived alternatives hypothesis, this should not occur, since in each case the second melody is part of the set of alternatives invoked by the first, such that the similarity ratings ought to be equal. In their discussion, Bartlett and Dowling attempt to deal with this issue by proposing that during the NN comparison, listeners might have perceived the nonscalar note in the first melody as a grace note of sorts, such that the melody was still considered to be scalar in the observer's mind. The second melody's nonscalar note, however, was perceived as a truly nonscalar element, resulting in the melody as a whole becoming nonscalar.

This addendum to the original perceived alternatives hypothesis does not adequately solve their problem. Remaining is the issue of why listeners should see the nonscalar note as a grace note in the first melody, but not the second. One could convincingly argue that the repetition of the nonscalar tone in the second melody promoted it to 'grace-note' status in the minds of the listeners. Consider, for example, a melody which repeats itself several times. If, in the first presentation of the melody, a nonscalar note appears, which does not appear in subsequent repetitions, listeners would likely consider that note's appearance to be a mistake on the part of the performer, and retroactively perceive that note as distinct from the rest of the melody (not a grace note). If, however, the nonscalar note appears in subsequent repetitions, listeners might then judge it to be a grace note, and part of the melody (giving rise to an NS, or even an SS-like representation, if both out-of-key notes were retroactively considered grace notes, and not out-of-key mistakes). Little is said by Bartlett and Dowling concerning just when or why a listener considers a nonscalar note to be a 'sort of scalar' grace note, as opposed to 'truly nonscalar' performance error.

A second and more drastic problem with their work concerns the differing saliences of scalar and nonscalar melodies. It is possible that the nonscalar melodies, with a jarring, "sour", nonscalar note in them, had a higher associated salience relative to the scalar melodies. In addition, this salience value might be attenuated somewhat by the passage of time and intervening stimuli. Given this, it's possible that the asymmetry they obtained was a result of differing saliences of the nonscalar melodies, depending on whether they appeared first or second. When the nonscalar melody was the first member in the pair, the salience of the melody would be attenuated following the passage of time with hearing the second melody. When the nonscalar melody is the second member of the pair, however, very little time passes between exposure to it and the opportunity to respond, meaning the associated salience would be maximal. While Bartlett and Dowling acknowledge the jarring quality conferred by a nonscalar note in an otherwise scalar melody, they do not believe it to be a main contributing factor to asymmetric similarity. Interestingly, however, it is possible to develop a simple model relying solely on this quality, which accounts for their data quite efficiently.

In terms of memory studies, it has been seen that the ability to detect an alteration in a melody is, in part, dependent on the degree to which that melody conforms to a diatonic scale (Dowling & Fujitani, 1970). What this means, then, is that accurate encoding of each note event in a melody seems to be facilitated by diatonic structure, and interfered with by deviations from that structure. However, to perform the task laid out by Bartlett and Dowling (1988), subjects would not necessarily have to attend to and remember each note event. Rather, if any feature of one melody stood out relative to the other melodies, and the psychological effect of that feature dissipated over time, then subjects could have been sensitive only to that, without resorting to memorizing each note event. A prime candidate as the feature in question would be the jarring (salient) quality of a non-scalar note in an otherwise scalar melody. If no nonscalar notes appear, the melody would have a low salience associated with it immediately upon listening. If a nonscalar note does appear, however, the immediate sensation would be high salience, but that would dissipate toward low salience over the passage of time and intervening stimuli.

The above mechanism better explains Bartlett and Dowling's data than does the perceived alternatives hypothesis. The pattern of results was that SN melody pairs were rated as less similar than were NS pairs. Furthermore, the SS melody pairs were given higher similarity ratings than were the NN pairs, a result not predicted by the perceived alternatives hypothesis. But consider for a moment the possibility that subjects' ratings were based on relative saliences. As a rough guide, since there are no jarring notes in the scalar melodies, they would have an associated salience rating of zero (0). The nonscalar melodies would have an initial salience rating of one (1) due to the jarring quality of the nonscalar note, but over time that salience would approach zero, and become some number greater than zero but less than one (for the sake of argument, we'll use 0.2). This means that after hearing the SN melody pair, the saliences of the scalar and nonscalar melodies would be 0 and 1 respectively, which would result in low similarity ratings. After hearing the NS pair, however, the relative saliences would be 0.2 and 0, meaning similarity ratings would be higher for these pairs than for SN pairs. More important, though, is the fact that the SS/NN difference in similarity ratings is also explainable using this algorithm. After hearing the SS melody pair, associated ratings would be 0-0, which should result in high similarity ratings. Following listening of the NN melody pair, however, associated ratings would be 0.2-1, which should result in lower similarity ratings relative to the SS pairs, which is exactly what occurred. Although this is a post hoc hypothesis, its explanatory power exceeds that of Bartlett and Dowling, and requires no tailoring to handle any particular stimulus condition (i.e. the NN ratings). The key to their interaction then, might very well have been the contrasting saliences of scalar and nonscalar melodies arising from the inclusion of nonscalar elements.

Dowling (1984) contends that the above mechanism is countered by evidence that when the interstimulus interval is increased between the standard and comparison, one obtains a weaker rather than stronger asymmetry, counter to what the above model would predict. However, in the task to which he refers, listeners were instructed to explicitly rehearse the standard melody in that intervening period. It stands to reason that explicit rehearsal would give rise to a maintenance of the 'salience' of the sour note, since listeners would be repeatedly reminding themselves of it during rehearsal. Dowling points out that assimilation of the nonscalar element into the tonal structure of the rest of the melody did not occur under these conditions, and he is absolutely correct. There is no reason to assume that mentally rehearsing a piece of music would give rise to assimilation of sour notes into the prevailing key; the drop in salience might only come about as a result of a passive decay, which is interfered with by active rehearsal (in the Discussion, a second possible mechanism not relying on the influence of passive decay will be proposed).

One could extend the 'salience hypothesis' outlined above to include diatonic notes from different levels of the tonal hierarchy, such that salience would be determined by the relative position of each note in the hierarchy. In other words, notes from highly stable positions would have lower associated salience ratings relative to notes from less stable positions (note that Tversky's model gives rise to the opposite relationship between hierarchic position and salience, largely as a result of a different definition of the term). When two different melodies are presented for similarity comparison, the resulting rating will depend on the relative positions of the changed notes on the tonal hierarchy, and the order in which they are presented. The predicted outcome from this 'modified salience hypothesis' is different from that made by the perceived alternatives hypothesis, at least for the case when all melodies are diatonic or scalar, since, according to the perceived alternatives hypothesis, only departures from scalarity should be effective in evoking different sets. In short, then, the perceived alternatives hypothesis would state that no differences in similarity ratings should be obtained when no deviations from scalar structure occur. The modified salience hypothesis, however, would say that the asymmetry is still obtainable when scalarity is preserved, provided subjects are sensitive enough to notice a change when it occurs. These differing predictions are testable, and are the focus of most of the studies presented in this thesis.

The hypothesis described above is not the only one available, however. A second possibility might rely more on the degree to which the target note in the first melody was accurately stored in memory. That is, when the melody is first presented, and a tonal hierarchy activated, the relative importance of each incoming note event is assessed. Some tones would be considered 'important', and others subordinate, according to their relative positions on the tonal hierarchy of the piece as a whole (Bharucha, 1984). 'Important' or 'stable' tones (residing near the vertex on Krumhansl's cone), then, might have a stronger memory representation than 'subordinate' tones. When the second melody is presented for comparison, the changed tone would be compared to that memory representation. If the target was originally an 'important' tone, the comparison would be facilitated by the more accurate memory representation, meaning subjects ought to more readily notice these alterations (stable-unstable) than when the stimuli appear in the reverse ordering. Transferring this to the domain of verbal processing, it would specify that listeners would more accurately encode the important features of a sentence or story, such that if those features were later changed, they would be highly noticeable. If, on the other hand, less important words were altered, listeners would be less likely to notice those changes, due to their weaker memory representation.

This notion is essentially identical to Bartlett and Dowling's memorability hypothesis, with one added feature: It does not apply when the alteration involves a nonscalar note. The nonscalar element has the additional confound of being jarring, and hence highly noticeable to listeners. As a result, the sourness of the tone gives rise to a confound with memorability: listeners might not have an accurate memory representation of the tone's identity, but would have a rather strong memory impression of there being a jarring element in the melody. So, although these nonscalar tones are considered 'subordinate' to more scalar elements according to Krumhansl (implying a weak memory representation), they are nonetheless highly noticeable to listeners, thus introducing a confound between noticeability on the one hand, and encoding accuracy on the other. Again, the verbal equivalent to this would be to include a word that does not fit the semantic or grammatical structure of the rest of the sentence. Such a word would be highly noticeable, for reasons different from its being important to the structure of the sentence as a whole.

The work described in this thesis was carried out largely with the intention of demonstrating that the perceived alternatives hypothesis is limited in its generalizing capabilities. In a typical experiment reported below, listeners heard all-scalar melodies, and were asked to indicate whether the melodies were the same or different, on a confidence scale. If the perceived alternatives hypothesis is the functional mechanism in these tasks, then no asymmetry ought to appear. If, however, one of the other two alternative proposals described above is the correct explanation, then asymmetries should appear, depending on the relative stability of the changed notes in the melodies. Specifically, alterations resulting in departures from the tonic should be more noticeable than alterations resulting in a return to the tonic.

A same-different task was employed for several reasons. First, this technique serves to rule out potential difficulties introduced with similarity ratings or probe-tones. In particular, one is not sure on what basis subjects perform their ratings in those tasks. Asking listeners to rate how representative a particular tone is with respect to the previous tonal context in no way guarantees that they will do so. The ratings might have been a reflection of how well the tone completed the sequence, a very different task than that intended by the researchers. In a same-different task, however, the goal set out for the listener is clear: They must state whether the melodies are identical or not. It is possible, then, to analyse the data with detection methods. Second, the same-different task allows for conclusions to be drawn concerning the memorability of melodies and their constituent elements, and how the memory for those elements depends on their hierarchic position. Prior work involving same-different measures employed single-tone comparisons which either followed, or were interpolated by a tonal context. While this method provides insight into the coding of isolated tones, it would be imprudent to draw definite conclusions concerning the way tonal material is processed when that material is part of the context.

One additional modification to earlier work was introduced. Specifically, in Krumhansl's and others' work, the asymmetry was obtained by comparing two notes residing on different levels of the same tonal hierarchy. While it seems reasonable to conclude that movement toward or away from the vertex of the hierarchy was somehow involved in this asymmetry (that is, direction of movement produced differences in similarity), there is no guarantee that this is the case. It may be that some other factor produced the asymmetric ratings; A likely candidate might be the fact that the melodic intervals formed between the target tone and tones immediately adjacent would be different between the standard and comparison melodies. So, an alteration may be detected as a change in interval size, and varying interval size changes might be associated with varying degrees of noticeability. To control for this, it was necessary to manipulate the relative positions of the two notes in the tonal hierarchy, without actually changing the order of presentation of the notes themselves.

The solution to this was actually alluded to in Krumhansl's original paper.

Krumhansl (1979) reported a study in which listeners were required to state, via a confidence scale, whether two tones were identical or not, when a musical sequence was interpolated between them. The two tones were either diatonic-nondiatonic, or the other way around. She also varied the degree to which the interpolated sequence conformed to diatonicity, such that half were 'tonal', while the other half were 'atonal'. The central result was that for diatonic standards alteration detection was superior in the tonal context, while for nondiatonic standards the atonal context showed superior detection accuracy. While she did not pursue this further, she did state that it was possible that the 'atonal' melodies were actually heard in a different key, resulting in a reversal of stability functions of the two tones (assuming inverse tonal hierarchies).

There are several flaws with this study, calling into question the obtained results. First, she did not adequately control for possible tonal structure cropping up in the 'atonal' sequences. Second, she neglected to include trials where either a diatonic or a nondiatonic tone served as both the standard and comparison (akin to Bartlett and Dowling's SS and NN conditions), and were different from one another (her "same" trials were either both diatonic or both nondiatonic). This manipulation would be necessary to rule out the possibility that subjects were not comparing one tone to the other, but rather comparing the second tone to the previous context. In other words, for tonal contexts, if the second tone was nondiatonic, listeners simply might have been biased to report 'different', irrespective of the first tone's identity; for 'atonal' contexts, these same tones would be associated with a 'same' bias. Given this simple algorithm, one could obtain the pattern of results Krumhansl reported. She stated that "diatonic tones were more often correctly recognised on same trials than were nondiatonic tones" (p.369). In other words, when the second tone was diatonic, the bias to say 'same' would be correct, while the bias to say 'different' when the second tone was nondiatonic would be incorrect. She goes on to say that "...on different trials, diatonic tones were less often confused with nondiatonic tones than nondiatonic tones were confused with diatonic tones" (p.369). In other words, in the first case, where the second tone was nondiatonic, the bias to say 'different' would have been correct (less 'confusion'), whereas in the second case, in which the second tone was diatonic, the bias to say 'same' would have been incorrect (greater 'confusion'). In short, the asymmetry that Krumhansl attributed to differential stability levels of the tones might in fact have been due to a simple response bias brought on by the presence of nondiatonic tones in the second position of the paired-comparison task.

For our purposes, it was necessary to more carefully manipulate the tonal hierarchy that was instantiated, such that two notes considered stable and unstable respectively in one tonal hierarchy would reverse that ordering on another, but in either key the notes would both be considered diatonic. This type of manipulation was carried out by designing melodies that would promote a key of C-major, or D-minor. These two keys are identical in terms of their constituent scale elements, except for a single note, whose identity in D-minor is theoretically debatable. The note B is actually a B-flat in the D-minor (natural) key signature, but does appear in the Dorian minor mode, as well as the melodic minor mode. For this reason, the note B was not employed as a target in any comparison. Instead, comparisons took place between the respective root-chord tones of the two key contexts, tones which should, according to Krumhansl's data, be associated with the highest stability levels. The C-major tonal hierarchy has the notes C, E, and G at high stability levels, with D, F, and A being less stable. The D-minor hierarchy has the notes D, F, and A at the high stability level, and C, E, and G being less stable. Based on these relative stabilities, then, a single scale-step alteration associated with high sensitivity in one key should have relatively lower detection levels in the other. For example, moving from the tone G as the standard to the note A as the comparison should be highly detectable in the key of C-major, but significantly less noticeable in the key of D-minor.

In summary, these stimuli differ from Krumhansl's in three respects. First, the melodies were designed a priori to promote a particular key, and hence a particular tonal hierarchy. Second, the alterations were all diatonic in nature. Third, Krumhansl required listeners to compare two essentially isolated tones, one occurring prior to the interpolated sequence, the other following the sequence. The sequences presented here contain the target elements, thus ensuring that the focus of the research is squarely on the processing of the tonal material of melodies. It is therefore possible to test the claim put forth by Bartlett and Dowling (1988). Again, they purport that deviations from scalarity are necessary for asymmetric similarity relations to arise, such that no such relations should follow from all-diatonic/scalar stimuli. If, however, the asymmetries arise from the dynamic tone quality differences between the altered elements in the two melodies, asymmetric similarity should remain. As well, by employing a recognition task, it is possible to further explore the memorability of tonal material that is part of a musical context, and how that memorability might vary solely as a function of hierarchic position of a note forming an integral part of a melody.

Chapter II. Statistical Methodology

The memorability of tonal material was of primary importance to this thesis. To this end, a signal detection model was applied to confidence interval ratings in the analysis, in order to examine sensitivity independently of response bias. In this chapter a brief overview of the theory of signal detection, as well as the specific calculations involved in the data analysis, is presented. For more extensive coverage of these issues, the reader is urged to examine one of several excellent sources on this topic (Green & Swets, 1966; Macmillon & Creelman, 1991; Swets, 1973; 1986).

In the classic signal detection paradigm, on each experimental trial an observer is presented with either a weak signal, or nothing. The observer's task on each trial is to determine whether or not the signal was present. There are two possible stimulus types (Signal -plus-noise, or noise alone), as well as two possible responses ("Yes a signal appeared", or "No, a signal did not appear"). The result is a response matrix with four outcomes, depending on the stimulus and the response. A hit (H) is said to occur if the observer says "Yes" on a signal trial. A false alarm (FA) is said to occur if the observer says "Yes" on a noise trial. Likewise, misses and correct rejections are said to occur when the observer reports "No" on signal trials and noise trials, respectively.

At this point, a method is required for quantifying the observer's ability to detect the signal, hereafter referred to as 'detection sensitivity'. One might, for example, decide that the proportion of hits could serve as an index. This measure, however, is clearly inadequate since two observers with equal hit rates could be at completely different levels of detection ability (e.g.: H=.9, FA=.1, versus H=.9, FA=.9). In the latter case, the individual does not appear to have been sensitive at all to detecting signal trials; rather, they appear to have been responding 'yes' most of the time, regardless of whether a signal or noise trial was presented. Furthermore, if the H value were by itself considered the sole index of sensitivity, then one would conclude that an individual with H=.1, F=.1 is less sensitive to detecting the target relative to someone who gives H =.9 FA = .9, when in fact they are simply exhibiting a response bias in the opposite direction. Ideally, then, one would like to have a sensitivity index which will not vary with response bias variations. Other early sensitivity indices (H-F, [H-F]/[1-F], etc.) suffer from similar criticisms (for a detailed examination, see Swets, 1986). However, several sensitivity indexes which do not vary with response bias variations are based on the use of a plot of the observer's performance known as the ROC, or Receiver Operating Characteristic.

The basic model of signal detection first assumes that there is no such thing as a 'null' trial, from the observer's perspective. That is, on those trials where no signal is presented, observers will often report that a signal was present (FA). According to signal detection theory, the observers actually did detect something which they labelled a 'signal'. That 'something' was an internal stimulus, caused in and by the observer's own sensory system, and which the observer detected and mistook for a signal. It is further assumed that this internal noise can be represented as a probability density function of possible intensity values, such that its strength can vary from trial to trial. The signal is also represented as a probability density function, and in the simplest model of signal detection analysis is assumed to have the same shape and variance as the noise alone trials. This is based on the assumption that the noise, still present on signal trials, is the main contributing factor in determining the variance and shape of the distribution, while the signal contributes a constant value to that distribution.

Whether or not the observer reports detecting a signal depends on the intensity of the stimulus, as determined by the probability density function. The observer is assumed to have an internal criterion value, such that if the stimulus intensity surpasses that value, they will report that a signal appeared. This will also occur if the value sampled from the noise distribution surpasses the criterion. The location of the criterion is not a constant in more recent versions of signal detection theory, such that it can vary from one observer to another, as well as within a particular individual. The setting can depend on internal factors, such as an individual's overall level of conservativeness, or on external factors such as instructions or payoff matrices. For example, an individual who is told to be very sure a signal is present before responding 'signal' will set the criterion at a higher level than an observer told to try to detect all of the signals. In other words, the criterion setting is a function of the observer's response tendency, or bias.

Insofar as sensitivity is concerned, if the two distributions are largely overlapping, then for a particular trial either distribution might cause the internal sensation to fall above criterion, whereas if they are further apart from one another the likelihood of the signal-plus-noise distribution producing an above-criterion value will be greater than that for the noise distribution. As a result, with largely separated density functions, the likelihood of a false-alarm will be smaller, and the likelihood of a hit will be larger, relative to situations where the distributions are on top of one another. Since the goal of this analysis is to determine an observer's detection sensitivity, an appropriate measure would therefore be the distance between the means of the two probability density functions.

In measuring sensitivity, one does not attempt to directly map the underlying probability density functions. Rather, their relative positions can be derived from the proportion of Hits and False Alarms in a session. The proportion of Hits corresponds to that area under the signal-plus-noise distribution which falls above the criterion line, while the proportion of false alarms corresponds to that area under the noise distribution which falls above the criterion. By systematically varying the criterion location, one can see how those proportions vary, and determine the locations of the two distributions relative to one another. If one plots these (H, FA) values, with hits on one axis, and false alarms on the other, the resulting graph is known as a receiver operating characteristic, or ROC. Other terms used in the psychological literature are 'memory operating characteristic', or MOC, (Norman & Wickelgren, 1965) isosensitivity curves (Luce, 1963), and relative operating characteristic (Swets, 1973). For simplicity's sake, we will refer to it simply as the ROC, since the terms are largely interchangeable, differing only in their descriptive value, rather than the calculations involved. All the points on a given ROC curve represent responding with equal sensitivity, but varying bias. The actual shape of the ROC is assumed to depend on the shape of the underlying density functions (logistic, Gaussian, and so on). It is this curve which forms the backbone of several widely used sensitivity measures.

One of the most common parameters of the ROC used to estimate sensitivity is known as d-prime, or d'. The value of d' corresponds to the distance (in standard deviations) between the means of the N and S+N probability density functions (Swets, 1973). While this value has been used extensively in the literature to stand for an estimate of sensitivity, the validity of d' depends to a large extent on the assumptions inherent in its calculation not being violated. For example, the standard calculation of d' involves the assumption of equal-variance Gaussian probability density functions underlying the observer's response tendencies. Under these conditions, the binormal plot (using z-scores) of hits and false-alarms will correspond to a straight line with unit slope. However, if the plot is not linear, this implies underlying distributions that are not normal; likewise, a non-unit slope implies unequal variances of the underlying distributions. To handle these irregular plots, one might attempt to base their calculation on underlying distributions with shapes different from normal, which correspond to the obtained ROC. Alternatively, one could turn to using a 'nonparametric' sensitivity measure which does not make the assumptions carried by d' calculation, or at least makes fewer. Several such less-parametric measures exist, although some of the more popular ones carry with them new sets of issues and flaws.

One such sensitivity measure commonly used is that of proportion correct, or p(c). The appeal of p(c) is primarily due to the ease of its calculation (the weighted average of hits and correct rejections), as well as its purported 'nonparametric' status (Macmillan & Creelman, 1991). Unfortunately, however, its relation to an observer's actual detection sensitivity is questionable if there is a great deal of response bias. Macmillan and Creelman (1991) point out that with varying response bias, p(c) can remain constant as d' values change drastically. Therefore, p(c) is truly only a reasonable index of sensitivity when response bias remains constant. From the available data, it seems that same-different procedures inevitably lead to response bias in observers, making p(c) an unviable sensitivity measure for these designs (Macmillan & Creelman, 1991).

A second class of nonparametric measures is based on areas under the ROC curve.

The theory underlying these nonparametric indices is quite straightforward. If one were to trace out an ROC for a particular observer, while one could compute d', such a measure carries with it assumptions which, if violated, can influence the accuracy and validity of the measure. As an alternative, were one to compute the area under the ROC curve, one would have a reasonably assumption-free estimate of sensitivity. As sensitivity increases, the ROC curve moves toward the upper-left corner of the graph, thus increasing the area underneath, irrespective of response bias or underlying distributions. One such measure which is quite common in the psychological literature (largely due to its ease of calculation as opposed to its nonparametric status, according to Macmillan & Creelman, 1996) is that of A'. This measure is based on a single (H, FA) point in ROC space, making it one of the simplest to compute, but necessitating a great deal of mathematical guesswork as a result.

With only a single point in ROC space, tracing a precise curve is impossible. Theoretically, an infinite number of ROC curves pass through any given point in ROC space. So, in order to compute the 'area' under the non-existent ROC curve, some estimating is involved. One conservative estimation method would be to calculate the area under the boundaries defined by drawing lines connecting the lower-left corner of the ROC (where hits and false alarms equals zero), the obtained (H, FA) point, and the upper right corner (where H and FA equal one). Figure 2.1 depicts a single (H, FA) point, and the connecting lines described above. The area labelled "MIN" would be used as an estimate of sensitivity. Although this is almost always an underestimate of true sensitivity, it would be adequate for comparison of sensitivity levels across stimulus types, since all would be equally underestimated, assuming equal bias. Unfortunately, this area also happens to be equal to proportion correct (Macmillan & Creelman, 1991), which, as mentioned earlier, may not be a true reflection of sensitivity with unequal bias.

The measure A', developed by Pollock and Norman (1964) begins with the assumption that the area under any ROC that passes through a given (H, FA) point must include at least the area term MIN (in Figure 2.1), since the values of successive points on regular ROC's generally increase monotonically, while the slopes at these points do not increase (Norman, 1964). However, being a curve, the area would also include at least some of the space above this minimum value. The area delineated by MAX in Figure 2.1 would be excluded from the calculation, since points in this space represent performance superior to that of the point (H, F). As a result, there is a nebulous area region, labelled MID1 and MID2 in Figure 2.1, which may or may not be included, depending on the specific shape of the ROC. In determining A', Pollock and Norman (1964) proposed to average these two areas, and add the result to the minimum area value.

As noted earlier, A' has been used extensively in research for some 25 years, partly due to its ease of calculation, but also due to its purported 'nonparametric' status. This reputation has arisen largely as a result of it being developed without reference to any theoretical assumptions concerning the shape of the underlying density functions (Macmillan & Creelman, 1996 contains a tabulation of these citations). Recently, however, this proposition has been questioned by several theorists (Macmillan & Creelman, 1991, 1996; Swets, 1986). In particular, if one computes implied ROCs for different levels of A', one finds a marked resemblance to ROC's traced by density functions of specific shapes. At low sensitivity values, the implied ROC's for A' correspond to those traced with logistic distributions, while at high values the correspondence shifts to that of rectangular distributions (Macmillan & Creelman, 1996). It is therefore incorrect to justify one's use of A' on the basis of it being 'nonparametric' since its implied ROC's do, at least mathematically, make reference to specific underlying distributions. As well, because rectangular distributions are only rarely supported by the available ROC data, it would warrant against using A' for high sensitivities. However, Swets (1986) points out that as a rough estimate of sensitivity, A' appears to do the job.

Instead of A', many detection theorists advocate collecting the necessary data to trace a more complete ROC, in order to have a better picture of what true sensitivity might be. Such a task requires manipulating the observer's criterion setting within a given session, by varying payoffs for hits and false alarms, or through instructions. Alternatively, a more simple method of accomplishing this is to ask observers for a confidence rating on each trial (Egan, Schulman, & Greenberg, 1959; Murdock, 1965). The rationale underlying this is that the subject's criterion can and does change in the course of an experiment. The rating scale offers the researcher insight into the position of that criterion relative to the density functions. Each point on the confidence scale can be seen to represent a criterion location. For example, the point labelled "very sure different" would correspond to a strict criterion location, whereas the point labelled "very unsure different" would correspond to a more lax criterion. It is possible to successively break the rating scale up into different binary-decision responses, such that "everything to the right of this point was above this criterion, everything to the left was below it". So, for example, subjects might be asked to rate their confidence in telling whether two stimuli were the same or different from one another on a six-point scale (ranging from very sure different to very-sure same). A hit may be defined as those occasions when the stimuli were different and the subject reported that they were different. A false alarm may be defined as those occasions when the stimuli were not different, but the subject indicated that they were. Consider now the right-most point on the scale (the very-sure different response). If we consider this as the criterion location, then the scale as a whole becomes a binary decision space, such that responses of "very sure different" are assumed to occur when the stimulus value exceeded this criterion location (hits when the stimuli were different, but false alarms when they were the same), whereas any other responses are assumed to occur when the stimulus value did not (and therefore are misses and correct rejections). Moving down a step on the scale (to the "sure different" response), and placing the criterion here corresponds to a more lax criterion, where all responses at or above this location (the cumulative proportion) corresponds to a hit when the stimuli were different, and a false alarm when the stimuli were identical. In computing the cumulative hit and false alarm proportions for each point on the scale, one is, in essence, computing the area under the S+N and N distributions at various criterion locations. In other words, one is mapping points in ROC space, and hence tracing a more detailed ROC than would be possible from a simple binary response measure.

A more accurate area measure of sensitivity than A' may then be derived from this data, either by plotting the best-fitting curve of these points and calculating the area (denoted Az), or by calculating the area under the boundary defined by linking the points in ROC space (denoted Ag). This latter measure has also been called P(A) by some theorists (Green & Swets, 1966/1974; Swets & Pickett, 1982), but the more common usage of Ag will be employed here Both Az and Ag are more accurate assessments of sensitivity than is A', since their area calculations are based on a more accurate approximation of the true ROC of the observer than is A', being based on a single point in ROC space (Swets & Pickett, 1982). In addition, the implied ROC's for these measures do not suffer the same fate as does A', which shows a change in the shape of the implied underlying distributions as one goes from low to high detection levels.

In choosing an appropriate sensitivity measure, several factors were considered. First, one would like a measure which does not include distributional assumptions, or at least those assumptions are minimal. Second, one would like a reasonably accurate sensitivity measure, which will not be influenced by possible biasing factors.

The measure Ag fits this bill perfectly. It is more accurate and less subject to unwanted biasing factors than is A' in determining sensitivity, being based on a more complete representation of the true ROC. Second, it does not require underlying distributional assumptions or estimation procedures, as does Az (of which the calculation involves assuming Gaussian underlying distributions). Third, the calculations involved are relatively straightforward. The formal computation for Ag is a simple algebraic formula:

$$Ag = 0.5 * \sum (FA_{i+1} - FA_i)(H_{i+1} + H_i),$$

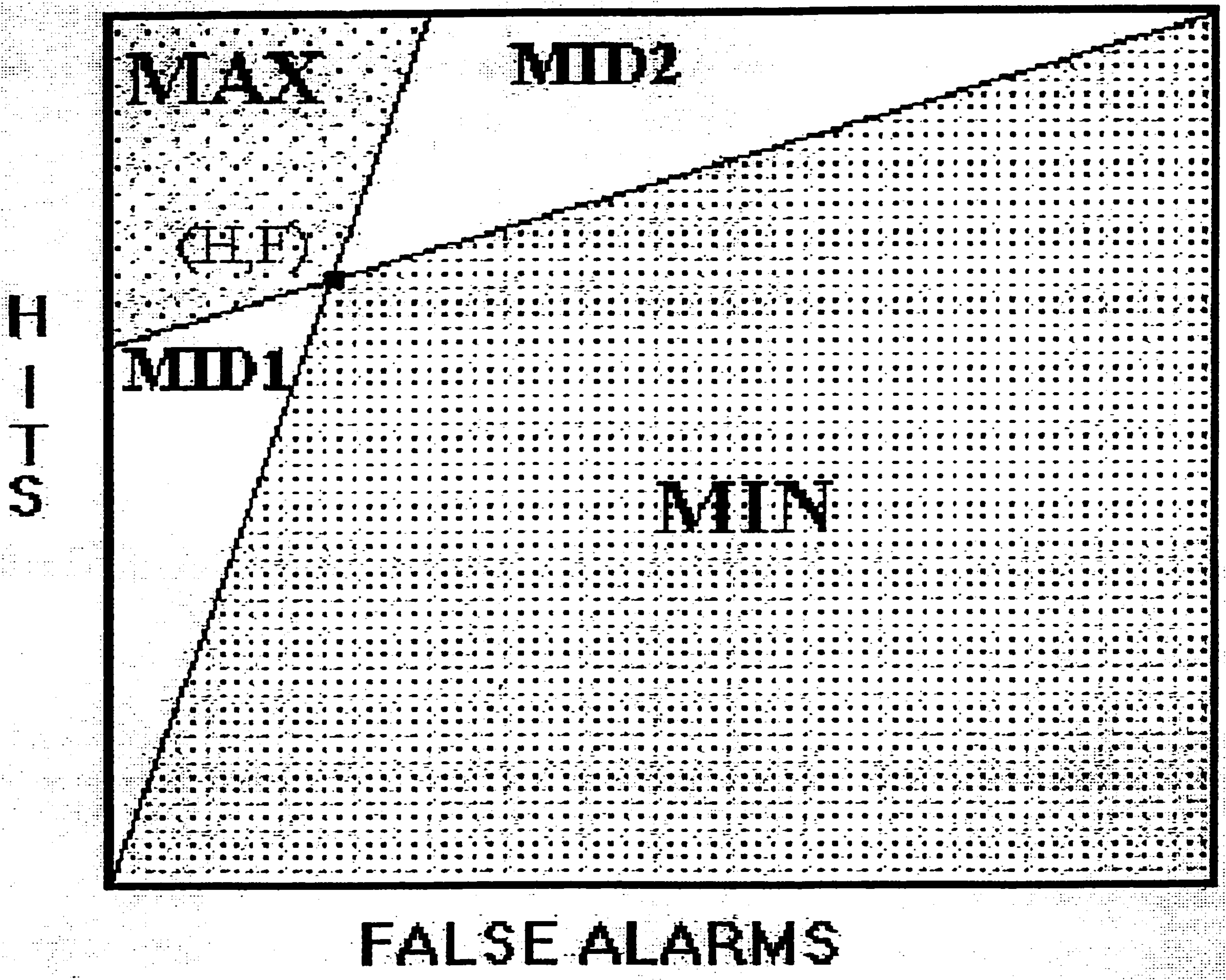
where i refers to each point on the confidence scale, H refers to hits, and FA refers to false alarms. Ideally, one would like to have as many points as possible, since fewer categories will result in an underestimate of true sensitivity (MacMillan & Creelman, 1991). To this end, the studies presented in subsequent chapters employed a six-point confidence scale.

The author developed a simple computer program that would accept as input the cumulative ratings data from each subject for each specified trial type. The program's accuracy was tested using data from other published works; the output was compared to the values obtained in those papers. In all cases, the results were identical to those obtained by the original authors, confirming that the program was doing its job. The output consisted of the calculation of Ag, A', and the slope of the best-fitting line traversing the points when mapped on binormal axes (this last measure was provided merely to examine whether the obtained ROC's were regular, and did not enter into the analysis). In order to test between sensitivity values for particular stimulus conditions, and because the experimental designs were somewhat more complex than allowed for nonparametric analysis, the subsequent Ag scores were submitted to a repeated-measures ANOVA.

Figure Legend

Figure 2.1: A single Hit/False-alarm point in ROC space, labeled (H, F). Shaded area 'MIN' corresponds to ROC space which must fall under any regular ROC curve passing through point (H, F). Shaded area 'MAX' corresponds to ROC space which cannot fall under any regular ROC curve passing through point (H, F).

Figure 2.1



Chapter III. Experimental Results

3.1 Introduction

Primarily, the experiments described below were intended to demonstrate that Bartlett and Dowling's explanation for perceived asymmetries is inadequate for purely tonal situations. That is, if the melodies used in testing are tonal (with no "sour notes"), the perceived alternatives account for asymmetric relations would predict that no asymmetries should emerge in the similarity ratings data. However, if such asymmetries do occur, then it becomes necessary to invoke a new model, based on the tonal strength of each note in the prevailing context.

Prior to delving into the data, it is necessary to discuss several of the terms that are repeatedly employed in the body of this chapter, which may be misleading to the reader.

In general, the melody pairs for each trial were constructed using tones drawn from the C-major or D-minor scales. However, at presentation time, the computer program randomly determined the actual starting point for each melody, while preserving all of the original successive interval relations. In English, this means that on each trial the program altered the key in which the melody-pair was presented. So, for example, if the standard and comparison melodies were originally in the key of C-major, but the computer altered the starting tone to be that of E-flat on a given trial, the subject heard that melody-pair in the key of E-flat-major. These alterations occurred between trials, such that the standard and comparison melodies were in the same key on each trial. It was considered necessary to alter the key from trial to trial in order for a new hierarchy of tonal relations to be instantiated each time. This was done to avoid the possibility that listeners were merely recognizing the target sequences from a previous presentation, resulting in repetitive consolidation of the target sequence component tones in memory. By altering the starting point of each melody, listeners were forced to gain a new sense of key each time, thus causing a fresh set of dynamic tone qualities to be conferred on the melody-notes. As a result, the incoming tonal information was presumably processed in terms of important versus subordinate tonal information, rather than a more long-term memory representation.

Although the use of the terms C-major and D-minor are not truly what subjects were exposed to, the abandonment of these terms for less descriptive labels (melody X and Y, for example), would undoubtedly lead to confusion on the part of the reader. These terms were retained in the body of this text, since the thrust of the research was to demonstrate that the position of a given tone in the tonal structure of a melody has an influence on its alteration detectability.

3.2. Experiment 1: Pilot Study

This study served both as an initial exploration of asymmetric relations in tonal situations, and as a guide for constructing subsequent experiments. The first issue to contend with was the instantiation of an appropriate tonal hierarchy (C-major or D-minor) on a given trial. Based on the notion that tone-repetition is one possible causal factor in hierarchy activation (Oram, 1989), it was assumed that if the tonic note was repeatedly presented early in the sequence, the appropriate tonal hierarchy would be activated. As well, according to Lundin's description of melody, it was decided to begin and end each melody on the tonic note, and utilise relatively small successive intervals (no more than five scale-steps), to ensure that the melodies conformed to stylistic regularities found in composed music (Jeffries, 1974). Finally, based on results obtained in pilot research, the targets were designed such that an alteration would not occur at a contour inflection, since these points in melodies appear to be highly salient, and overly easy to detect.

Methods

Subjects. Fourteen McMaster University Undergraduates participated in the study, as partial credit for an introductory psychology course. Ten subjects were female, while the remaining subjects were male. The subjects' ages ranged from 18 to 22 years of age, with the median being 19 years of age. All subjects reported at least six years of formal musical training on an instrument, and thus were classified as musically trained.

Apparatus. All stimuli consisted of pure tones generated by a Yamaha TX-802 FM tone generator slaved to a Comptech 386-33 computer, running a QuickBasic program designed to present stimuli and record responses. All stimuli were presented to subjects through AKG-340 earphones. Subjects were seated in an IAC sound attenuation booth, and indicated their responses using a Logitech mouse. The computer controlled the tone generator, and recorded all responses.

Stimuli. Pure tones were employed for all stimuli, in order to ensure that listeners were not using timbre differences in detecting alterations between standard and comparison melodies. The intensities of the pure tones were scaled according to equal-loudness contours (Robinson & Dadson, 1956), to ensure that detection was not being made on the basis of loudness differences. Overall, the average level of any given melody was approximately 75 dB SPL.

Each melody consisted of three distinct parts: a four-note context sequence, a four-note target sequence, and a one-note terminal sequence, presented at a rate of three notes per second. The context sequences consisted of the notes C-E-C-B to promote a C-major tonal hierarchy, and D-F-D-B to promote a D-minor tonal hierarchy (commencing on middle-C). The terminal sequences consisted of the tonic note from the tonal hierarchy activated by the particular context sequence. The target sequences consisted of twenty ascending four-note passages, depicted in Figure 3.1. Within the target sequence, a particular note was specified as the target, which would be either altered or left unchanged in the comparison sequence. Half of the target notes appeared in the second position of the target sequence, while the other half appeared in the third position (denoted the POS factor). An alteration consisted of replacing the target note with a note immediately adjacent to it in the diatonic scale, such that the contour of the melody would not be violated with the alteration. Ten such alterations were used: C-D, D-E, E-F, F-G, G-A, D-C, E-D, F-E, G-F, and A-G. As seen in Figure 3.1, two sequences were created for each of the first five CHANGE conditions. The remaining five conditions consisted of the altered version of these sequences. Based on the C-major tonal hierarchy, these alterations were subsequently collapsed into two change types (the CHANGE factor), akin to Bartlett and Dowling (1988), such that either a stable note (C, E, and G) was followed by an unstable note (D, F, and A) or the reverse. Half of the trials consisted of no alteration occurring, while the other half consisted of a single note differing between the two melodies (TYPE factor). In summary, then, there were ten possible alterations, two possible context keys, two positions in which an alteration could appear, and two possible trial types (same or different). A factorial design was used, meaning there were 10 X 2 X 2 X 2 possible stimulus types. Each stimulus type was repeated once, for a total of 160 trials in the experiment. The experiment tended to run for approximately 35-45 minutes, depending on a given subject's speed of responding.

Procedure. Following a brief interview to determine musical background, subjects were seated in the IAC chamber, and were told they would hear two melodies, with the task being to detect any alteration which appeared between the two presentations. The use of a six-point confidence scale was explained to them (1=very sure same, 2=sure same, 3=guessing same, 4=guessing different, 5=sure different, 6=very sure different). Subjects were told to reserve the end points of the scale for when their confidence was maximal, and to use the ratings closer to the middle when confidence was lower.

Order of trial presentation was randomly determined for each subject, using the subject number as the seed for the random number generator in the program. Subjects pushed a button on the mouse to start the experiment, and following the presentation of the two melodies on a trial, the computer displayed the confidence scale and waited for subjects to select and input a rating. Once a rating was indicated, a brief (three-second) graphic occurred, to indicate the selected rating, after which the screen was cleared, and the next pair of melodies was presented.

In order to ensure that subjects were paying attention and were interested in the task, following every ten trials the number of correct responses obtained on the previous ten trials was displayed. This score was based on whether the ratings were on the appropriate side (same or different) of the rating scale relative to whether the trial-type itself was the same or different.

Results and Discussion

As stated in the Stimuli section, the CHANGE condition consisted of either a stable note transformed to an unstable one (SU), or the other way around (US), with reference to the C-major hierarchy. Note that this nomenclature would reverse if it were made with reference to the D-minor tonal hierarchy. That is, an SU change in C-major is actually a US alteration in D-minor, due to the fact that the component notes of a particular change occupy different levels of the tonal hierarchy in C-major and D-minor. The reader is urged to keep this in mind when examining effects and interactions, since the crucial interaction between context key and change type will imply that such a reversal of tonal hierarchy positioning of the respective tones is the operating factor. The levels of the CHANGE factor, then, are similar to Bartlett & Dowling's SN and NS conditions respectively, except that the stimuli presented here were melodies consisting entirely of diatonic tones. For example, if the target note in the standard sequence was C, and this target was changed to a D in the comparison sequence, this condition would be part of the SU trials (once again, C-D alterations are actually unstable to stable changes in the key of D-minor). According to the hypothesis described above, SU alterations should be more noticeable in a C-major relative to a D-minor context, while the reverse should occur for US alterations. According to the perceived alternatives hypothesis, no such interaction should appear, since all melodies maintained the same scalar structure between standard and comparison.

Prior to analysis, alterations involving the tonics of the respective keys were removed from the data. The reason for this was that within a given melody, the tonic appears at least three times (twice at the beginning, and once at the end). This strikingly high density of tonics introduces an unavoidable confound into the ratings data involving those alterations. Specifically, the tonic, which is maximally stable in the tonal hierarchy, was also a frequently presented note in the stimuli employed here. It is therefore possible that alterations moving away from the tonic would be more noticeable due to a frequently heard note being replaced by a less frequently heard note. The opposite result would occur if the order were reversed. Thus, inclusion of those CHANGE levels involving tonics would only serve to artificially inflate the key interaction sought between CHANGE and KEY, and thus were excluded from the data. The remaining ratings data for all subjects was submitted to a repeated measures ANOVA, using the ANOVA/MANOVA module in CSS-Statistica. The within-subjects factors involved were the position of the target (POS), the context key (CONTEXT), the type of alteration (CHANGE), and whether the standard and comparison melodies on a trial were the same or different from one another (TYPE).

A main effect of POS was obtained ($F(1,13) = 13.49$, $p < .01$), such that the mean rating for the second position was 3.02, while the mean rating for the third position was 2.77. In other words, subjects had an overall bias to report 'same' when the target sequence consisted of the target in the third position, relative to when the target was in the second position. Due to the nature of the dependent measure, and the fact that POS did not enter into a significant interaction with the TYPE factor, nothing further can be justifiably concluded from this effect. In addition, it was not clear at this point why such an effect was obtained. While it might relate to a serial position effect of memory, there is no obvious explanation for how the serial position of a target, regardless of whether or not that target changes in the comparison melody, could influence confidence. This effect will be examined further in subsequent sections.

As expected, trial type (altered versus unaltered) also served to significantly influence ratings ($F(1,13) = 31.77$, $p < .0001$). Unaltered trials were associated with a mean rating of 2.63, while trials where an alteration appeared were associated with a mean rating of 3.15. In other words, subjects were differentially sensitive to same or different trial types, as reflected in the difference between mean confidence ratings. Were the TYPE factor not significantly influencing ratings, this would be a sign that listeners were not reliably discriminating altered from unaltered trials, and the experimental results as a whole would be questionable.

There was a significant CHANGE X TYPE interaction ($F(1,13) = 6.58$, $p < .05$). For

SU trials, mean ratings for same and different stimuli were 2.73 and 2.97, respectively, while for US trials, the ratings for same and different stimuli were 2.52 and 3.33, respectively. In other words, the SU alterations were associated with a smaller difference in same-different ratings than were the US alterations. This can be interpreted to mean that in general, confidence levels grew (as indexed by movement toward the appropriate end-point value of the rating scale, since the end-points were indicative of higher confidence) on US trials relative to SU trials, implying that US alterations were somewhat more noticeable relative to the SU alterations, provided noticeability influences confidence level.

A significant three-way interaction appeared between the CONTEXT, CHANGE and TYPE factors ($\underline{F}(1,13)=9.23$; $\underline{p} < .01$). This interaction is displayed in Figure 3.2, with the different shades of bars representing the TYPE factor. If one examines the differences between same and different TYPE means for each level of CONTEXT and CHANGE, an interesting pattern emerges. It can be seen that for SU trial-types, the difference in mean ratings in D-minor contexts between same and different trial types is small, relative to that obtained for C-major contexts. The reverse effect appears for US changes. This is tentative evidence in favour of the hypothesis specified at the outset of this thesis, if a difference in confidence ratings across same and different trial types is taken as an index of 'noticeability' of an alteration. The confidence ratings differential between TYPE conditions is greater in C-major than in D-minor for SU changes, but greater in D-minor than in C-major for US changes, exactly as would be expected, given an SU change in C-major is, in fact a US alteration in D-minor, and vice versa.

In order to further explore this effect more systematically (and avoid the 'noticeability' assumption), the ratings data were converted to proportions of hits and false alarms, using the method described in Chapter II. Eight area scores were calculated for each subject, comparing hits and false alarms on same and different TYPE conditions for each level of the CONTEXT, CHANGE, and POS factors. These scores were submitted to a repeated measures ANOVA.

A main effect of CHANGE was obtained ($F(1,13) = 7.10$; $p < .05$), such that overall sensitivity was greater for US changes (mean area = 66.5) relative to SU changes (mean area = 53.4). This factor was also involved in an interaction with the KEY factor ($F(1,13) = 14.01$; $p < .01$), which is depicted in Figure 3.3. One can see that for SU alterations, mean sensitivity levels were greater for C-major than for D-minor, while the opposite effect was obtained for US alterations. In other words, when the alteration resulted in movement away from a stable element relative to the context key, detection ability was enhanced, relative to when the alteration moved toward a more stable element. Recall earlier, that the reader was instructed to keep in mind that SU and US are terms provided with reference to the C-major tonal hierarchy. Therefore, in the key of D-minor, that nomenclature would be reversed, meaning that the SU alteration is actually a US alteration in the key of D-minor. When an alteration was classified as stable-to-unstable according to the context key (e.g. C-major, SU condition, or D-minor US condition), detectability was superior to when that same alteration was transformed into an unstable-to-stable one, via the instantiation of a different context key (D-minor, SU condition, or C-major US condition). Although the constituent elements and their temporal order in the alteration didn't change, the nature of the alteration did change, when the context key changed.

The end result was a reversal in sensitivity to the target as Krumhansl would predict, and contrary to a perceived alternatives explanation.

A Neuman-Keuls post hoc analysis was carried out, to confirm whether or not the differences between mean areas within each alteration type were, in fact, significantly different. The result confirmed the hypothesis that key was having a significant influence on detectability. That is, within the SU alteration type, the difference between D-minor and C-major areas was significant ($p < .05$), as was the comparison between keys for the US alteration type ($p < .05$). So, for a given alteration type, be it SU or US, changing the context key resulted in detectability being affected in the direction specified at the outset of this experiment.

The lack of a POS main effect in the area scores analysis implies that the significant effect of POS from the ratings data was likely due to an overall change in response tendencies, rather than reflecting any differential sensitivity for the two target positions. As mentioned in Chapter II, the Ag measure is largely insensitive to bias factors, so the finding that alteration location had a significant effect on ratings, but not detection sensitivity, points toward a bias explanation.

In summary, then, a particular alteration was associated with greater or lesser detection sensitivity in listeners, depending on whether that alteration moved towards or away from stability, as determined by the context key. This reversal is interpretable from within a framework which accounts for dynamic tone-quality changes with changing keys, such as the tonal hierarchy framework, but does not seem to be accounted for by a perceived alternatives explanation, since all melodies/alterations maintained diatonicity within their respective keys.

In fact, a perceived alternatives hypothesis would be hard-pressed to predict that any effect involving CONTEXT and CHANGE would emerge, considering all notes in both melody types were drawn from the C-major scale, and hence should have been associated with the same set of perceived alternatives.

Following this study, two issues concerning the context sequences came to light, and warrant attention. First, the context sequences were designed to promote a particular key based on repetition of the tonic, along with the mediant (third). There is no guarantee at all that subjects actually did hear the 'C-major' melodies in C-major; likewise for the D-minor melodies. In fact, a re-examination of Figure 3.3 reveals that the reason for the interaction was largely due to the D-minor sequences showing vastly different sensitivity levels for SU relative to US changes, while the C-major sequences show little movement at all between the two CHANGE types. One possible reason for this is that in the key of C-major, the distinction between SU and US is not as strong as for the D-minor contexts. A second possibility is that the C-major tonal hierarchy was simply not elicited as strongly as the D-minor tonal hierarchy, resulting in a less-strong asymmetry. The fact that US alterations were associated with greater sensitivity than SU alterations fits with this latter explanation, since one would expect this finding for the D-minor contexts, but not the C-major contexts. To further examine these possibilities, one would have to construct contexts where there is more certainty as to the strength of activation of the appropriate tonal hierarchy. Presumably, a more complete crossover effect would arise as a result of this.

A second issue concerns the context sequences being presented in the same octave level as the target sequences, such that target notes were repeatedly sounded in the context sequences. This might have contributed to some confusion in the sensitivity data, since an alteration which resulted in a frequently presented note being changed to a less frequently presented note might have led to greater sensitivity than if the alteration was reversed. Despite the fact that alterations involving the tonics were omitted, it remains that the mediant for each key was consistently presented at the outset of each melody, thus guaranteeing a higher frequency of occurrence in each level of the CONTEXT set as a whole. To attempt to deal with these issues, a second study was carried out.

3.3. Experiment 2: Alteration detection with a triad context

Although the context sequences presented in the earlier experiment were designed with the intent of activating a specific tonal hierarchy, it is questionable whether that goal was achieved. As a result, new context sequences were constructed based on the criteria that they strongly activate a tonal hierarchy, as well as limit effects of note repetition.

Research involving the mapping of tone-profiles for various key-instantiating stimuli has revealed that the tones making up the root chord are quite effective in eliciting the tonal hierarchy in listeners, as indexed by tone profile studies (Krumhansl & Kessler, 1982). Hence, a reasonable means of hierarchy activation in listeners should be obtained through presentation of the root-chord notes. For this experiment, two opening triads were employed as context stimuli: C-major (C, E, G) and D-minor (D, F, A). In addition, the context sequences were presented an octave below the test sequences, to avoid the possible confounding effects of repeated pitches on detectability. If it is the case that listeners' detection was being influenced by movement toward or away from a frequently presented note, the critical interaction between type of change and context key should be attenuated by moving the contexts to a different octave. This prediction is justified by evidence from three different classes of research dealing with the perception of octave equivalent tones.

First, examination of the perception of 'octave-scrambled' melodies, in which the notes of highly familiar melodies are each transformed to different octave levels while maintaining their pitch chroma (same note name, different octave), reveals that listeners have a great deal of difficulty abstracting out the chroma information (Deutsch, 1972). It appears that this information is overridden by the unusually large interval jumps as well as the contour violations that are introduced by the scrambling transformation. When contour is preserved one finds that the ability to name an octave-scrambled melody does improve, although performance levels still fall well-short of perfection (Dowling & Hollombe, 1977; Idson & Massaro, 1978). It is therefore possible to conclude that octave equivalence, while strikingly apparent for simultaneously sounded tones, does not appear to be a salient feature of melody perception, and is not automatically apprehended in melody tones.

Second, Deutsch (1973) investigated the disruptive effects on memory produced by interpolating tone sequences between a standard and comparison tone. In general, she found that the greatest disruption occurred when the interpolated sequence was in the same octave level as the test tones, whereas the least disruption occurred when the interpolated sequence was an octave lower than the test tones. So, by presenting the context sequences an octave lower than the targets, the pitch chroma information should have a lesser interfering effect on listeners than if the context and targets were in the same octave, while retaining the ability to strongly promote a context key.

Third, psychophysical investigation has shown that isolated tones presented for similarity comparison result in little evidence of octave equivalence, provided pure tones are employed (Stevens, Volkman, and Newman 1937). Although one does find octave equivalence effects when complex tones are employed (Parncutt, 1989), this is easily avoided by using pure tone stimuli.

So, from the three general findings outlined above, it is possible to say that octave equivalence effects should be minimal, provided the tones occur in sequential fashion (a melody), the context appears in a different octave relative to the test tones, and pure tones are employed.

Methods

Subjects. Seventeen McMaster University undergraduates participated in the study, as partial credit for an introductory psychology course. The median age was nineteen years. Subjects were all classified as musically trained, such that they received a minimum of grade 8 Royal Conservatory of Music training on an instrument or voice. All of the subjects in this study received their training with piano.

Apparatus. All stimuli consisted of pure tones generated by a Yamaha TX-802 FM tone generator, which was slaved to a Comptech 386-33 computer, running a QuickBasic program designed to present stimuli and record responses. All stimuli were presented to subjects through AKG-340 earphones, at approximately 75 dB SPL. Subjects were seated in an IAC sound attenuation booth, and indicated their responses using a Logitech mouse. The computer controlled the tone generator, and recorded all responses.

Stimuli. Each stimulus sequence again consisted of a context sequence, a four-note test sequence, and a single terminating note. The context sequences consisted of the notes from either the C-major root chord (C-E-G), or the D-minor root chord (D-F-A), presented an octave below the target sequences. The terminal note was identical to the first note of each stream, and therefore was either C or D, depending on the context sequence of a particular trial.

Twenty test sequences were constructed, all of which consisted of an ascending stream of four notes, depicted in Figure 3.1. Half of the sequences contained a target note in the second position of the four-note sequence, while the other half contained a target note in the third position. The sequences in each subgroup consisted of the alteration C-D, D-E, E-F, F-G, G-A, D-C, E-D, F-E, G-F, and A-G. As seen in Figure 3.1, two sequences were created for each of the first five CHANGE conditions. The remaining five conditions consisted of the altered version of these sequences. In summary, then, there were two CONTEXT conditions (C-major, D-minor), two target positions (second/third position), and ten possible CHANGE conditions, with two melodies for each level. The only other condition included was trial TYPE, in which a change either occurred or did not occur. Overall, then this was a 2 X 2 X 10 X 2 design. Each trial type was presented twice to subjects, for a grand total of 160 trials. All other aspects of the stimuli were identical to that used in the experiments described earlier.

Procedure. Following a brief interview to determine musical background, subjects were seated in the IAC chamber, and were told they would hear two melodies, with the task being to detect if the melodies were identical or not. The use of the six-point confidence scale was explained to them (1=very sure same, 2=sure same, 3=guessing same, 4=guessing different, 5=sure different, 6=very sure different). Subjects were told to reserve the end points of the scale (1 and 6) for when their confidence was maximal, and to use the ratings closer to the middle when confidence was lower.

Order of presentation was randomly determined for each subject, using the subject number as the seed for the random number generator in the program. Subjects pushed a button on the mouse to start the experiment, and following the presentation of the two melodies on a trial, the computer displayed the confidence scale and waited for subjects to select and input a rating. Once a rating was indicated, a brief graphic indicating the selected rating was presented, taking approximately three seconds, after which the screen was cleared, and the next set of melodies was presented.

In order to ensure that subjects were paying attention and were interested in the task, following every ten trials the number of correct responses obtained on the previous ten trials was displayed. This score was based on whether the subjects' ratings were on the appropriate side of the rating scale relative to whether the trial-type itself was the same or different.

Results and Discussion

The ratings data for each subject for each trial type were submitted to a within-subjects ANOVA, using the ANOVA/MANOVA module of CSS-Statistica. The independent variables (KEY, POS, CHANGE, TYPE) were identical to the set described in the experiment above. Several effects of interest emerged from the analysis.

Once again, a significant effect of POS was obtained ($\underline{F}(1,16) = 6.49$; $\underline{p} < .05$). In particular, target melodies where an alteration appeared in the second position of the target sequence corresponded to a mean rating of 3.41, while the third position obtained a mean rating of 3.24. Listeners once again appeared to have an overall tendency to respond 'different' to melodies containing a target in the second position, relative to the third position.

The only other significant main effect was that of TYPE ($\underline{F}(1,16) = 31.4$; $\underline{p} < 0.0001$). As expected, trials on which no alteration occurred were given significantly lower ratings than were trials on which an alteration occurred. In general, then, subjects were reliably detecting when an alteration occurred, as indexed by significantly different ratings for unaltered and altered trials.

A significant three-way interaction occurred between TYPE, CONTEXT, and CHANGE ($F(1,16) = 26.75$; $p < 0.0001$), which is depicted in Figure 3.4. A similar interaction to that depicted in Figure 3.2 emerged here. One can see that for SU alterations, the difference in mean confidence ratings between same and different trial types for D-minor contexts is much smaller than for C-major, while for US alterations the difference in mean confidence ratings for D-minor contexts is greater than for C-major.

To this point, then, the pattern of ratings corresponds quite closely to that which was obtained in Experiment 1, above. As was done in the previous study, the ratings data were transformed to hits and false alarms, and areas under the MOC were computed for all combinations of the CONTEXT, POS and CHANGE factors, and submitted to a repeated measures ANOVA.

As predicted, a significant interaction was obtained involving the CONTEXT and CHANGE factors ($F(1,16) = 22.60$; $p < .001$). The interaction, pictured in Figure 3.5, resembles that obtained in the earlier experiment: Once again one sees that for SU alterations, C-major contexts were superior to D-minor, while the reverse occurred for US alterations. In fact, the crossover is clearer than in the earlier study, since the D-minor melodies show a rise in sensitivity when going from SU to US, while the C-major melodies show a drop in sensitivity. A Neuman-Keuls analysis confirmed that, as before, the difference between D-minor and C-major areas within each CHANGE condition were statistically significant ($p < .005$ for the SU condition, and $p < .05$ for the US condition).

Once again, the POS variable did not significantly influence sensitivity ($F(1,16) = 0.15$), implying that the POS effect found with the ratings data analysis was largely due to response bias, and not a shift in sensitivity.

So, using strong contextual stimuli similar to those employed by Krumhansl and her colleagues, one finds a strong influence of those stimuli on the asymmetry. Both contexts are associated with tonal hierarchies within which the two most stable levels have essentially the same notes, but on different levels of the hierarchy. If dynamic tone quality as determined by the tone's position in the tonal hierarchy is functional in the asymmetry, one would expect the result obtained. If, however, a perceived alternatives scheme is operational, no such reversal of the asymmetry should have occurred, since all melodies consisted of diatonic notes found in the key of C-major. Furthermore, the fact that the differential sensitivity was obtained when the melodies were presented with pure tones, and the contexts were in a different octave, weakens the argument that frequency of occurrence differences between targets was primarily responsible for the results obtained. Frequency of occurrence effects would only be an issue if octave equivalence were functioning in the study, a possibility with a low likelihood according to the available evidence (described at the outset of this chapter).

The issue of presentation frequencies, though, is not necessarily dead in the water. It might be that frequency of occurrences played a role in contributing to both crossover effects. This issue was addressed in the experiment presented below.

3.4. Experiment 3: Controlling for Frequency of Occurrence

When a strong tonal indicator was used (root chords) as the context stimulus, the asymmetry appeared to remain intact, even though the alterations were all scalar. As it was stated in Bartlett and Dowling (1988), the perceived alternatives hypothesis does not provide any compelling explanation for these results. According to that hypothesis, in order for the asymmetry to occur, one of the two melodies presented for comparison must deviate from scalarity, thus becoming part of a different set of possible alternative melodies. In the work described earlier, no such deviations occurred, yet asymmetrical perception was still observed, lending support to the notion that the tonal hierarchy was functional in the asymmetry.

There is a possible alternative explanation for those data, which does not resort to cognitive constructs, namely that there was a confound involving context key and frequencies of occurrences of constituent notes. As stated above, the reason for employing these stimuli was to provide a strong contextual cue that would lead to activation of the tonal hierarchy. Krumhansl (1979) stated that one possible method of activation of the hierarchy might be the relative frequency of occurrences of certain notes which would lead to a particular tonal hierarchy becoming active, a proposal with empirical support (Oram, 1989). Unfortunately, a high density of certain notes might also have been a main contributing factor in the obtained asymmetry, regardless of whether these tones served to activate a tonal hierarchy.

In the C-major context stimuli, there was a high density of root, third, and fifth notes, and a low density of seconds, fourths, and sixths. The opposite was true for the D-minor context stimuli. The root, third and fifth, though, are also the notes from the most stable level of the tonal hierarchy for C-major, while the second, fourth and sixth are the most stable notes for the D-minor tonal hierarchy. So, a stable note was also a very frequently-presented note, albeit presented in a different octave, using pure tones. While on a given trial, these stimulus characteristics would serve to minimize octave equivalence effects, it still remains that at the level of tonal hierarchies, two notes standing in an octave relation to one another are theoretically indistinguishable. As a result, the asymmetry described above might be a function of the differing frequencies of occurrences of the respective notes across the two context keys. Movement away from frequently presented notes, then, might be more salient than movement toward a frequently presented note. But according to Krumhansl, activation of a tonal hierarchy was made largely on the basis of frequency of occurrences of constituent note events. From that, it can be concluded that the method of activation would necessitate repetition of highly stable elements in the desired tonal hierarchy.

Obviously, this is a hopelessly paradoxical situation. In order to activate a tonal hierarchy, we apparently must present stimuli with a high density of stable elements (according to Krumhansl). But in doing so, a confound emerges between frequency of occurrence and position of a particular element in the tonal hierarchy. To avoid this, one must design context stimuli which have essentially the same frequencies of occurrences for all notes, but which have the added feature of activating different tonal hierarchies. In order to do so, one must assume that Krumhansl's method of 'counting note events' is not necessarily the only way key membership is determined. It is here that we must incorporate the theoretical issues of hierarchy activation described in the Introduction.

As pointed out by several groups of researchers (Cross, Howell, and West, 1983; Butler and Brown, 1984) it should be possible to rearrange the presentation order of the notes in the context to result in a different 'mode' or set of stability ratings being elicited in listeners. The study detailed below was nearly identical in design to that of the previous experiment, with an additional restriction of equal frequency of occurrences of constituent tonal elements.

As in earlier studies, subjects here were presented two melodies for similarity comparison, and responded using the six-point confidence scale. The identical four-note target sequences were employed. In this study, however, the context and terminal sequences were designed to ensure that any given note occurred equally often with a C-major or D-minor context. This was accomplished by rearranging the presentation order of the notes, but maintaining the relative frequencies of occurrences across stimulus types, a practice which has a measurable effect on listeners' key perception (Butler & Brown, 1984). Obtaining the asymmetry with these stimuli would rule out explanations relying solely on subject sensitivity to differing frequency of occurrence of the target notes.

Methods

Subjects. Forty-two McMaster University undergraduates participated in the study, as partial credit for an introductory psychology course. Subjects were all classified as musically trained, having all received a minimum of grade 8 Royal Conservatory of Music training on an instrument or voice. The median age was nineteen years, and no subject reported having of absolute pitch (the ability to name a musical note in isolation, or without a reference). Thirty subjects were females, and the remaining 12 were males.

Apparatus. All stimuli consisted of pure tones generated by a Yamaha TX-802 FM tone generator, which was slaved to a Comptech 386-33 computer, running a QuickBasic program designed to present stimuli and record responses. All stimuli were presented to subjects through AKG-340 earphones. Subjects were seated in an IAC sound attenuation booth, and indicated their responses using a Logitech mouse. The computer controlled the tone generator, and recorded all responses.

Stimuli. Pure tones scaled according to equal loudness contours were used once again. The melody sequences consisted of a four-note context sequence, a four note target sequence, and a three-note terminal sequence. The same target sequences from Experiments 1 and 2 were presented in this study.

The context sequences, in combination with the terminal sequences, ensured that a given note was presented equally often regardless of the context, though not necessarily relative to other notes in a given context. C-major context sequences consisted of the notes C-E-F-G, with a terminal sequence of D-D-C. In summary, there were two C's, two D's, and one each of E, F, and G. For D-minor, the context sequences consisted of the notes D-E-F-G, with the terminal sequence being C-C-D. Once again, summing the various elements results in two C's, two D's, and one each of E, F, and G. Using these stimuli, any differences in the asymmetry that appear between the two contexts would not be attributable to differing frequencies of occurrences of the constituent notes, since those differences do not occur with these stimuli.

All the other aspects of the design were identical to that of Experiment 2.

Procedure. Following a brief interview to determine musical background, subjects were seated in the IAC chamber, and were told they would hear two melodies, with the task being to detect if the melodies were identical or not. The use of the six-point confidence scale was explained to them (1=very sure same, 2=sure same, 3=guessing same, 4=guessing different, 5=sure different, 6=very sure different). Subjects were told to reserve the end points of the scale (1 and 6) for when their confidence was maximal, and to use the ratings closer to the middle when confidence was lower.

Order of presentation was randomly determined for each subject, using the subject number as the seed for the random number generator in the program. Subjects pushed a button on the mouse to start the experiment. Following the presentation of the two melodies on a trial, the confidence scale was displayed and subjects were allowed to select and input a rating. Once a rating was indicated, a brief graphic indicating the selected rating was presented, taking approximately three seconds, after which the screen was cleared, and the next set of melodies was presented.

In order to ensure that subjects were paying attention and were interested in the task, the number of correct responses obtained was displayed after each ten-trial block. This score was based on whether the subjects' ratings were on the appropriate side (same or different) of the rating scale relative to whether the trial-type itself was the same or different.

Results and Discussion

As in the prior experiments, the CHANGE condition was reclassified into SU and US alterations. In addition, alterations involving the tonics of the two keys were removed from the analysis. The reason for this is that within a given context, the tonics appeared at least twice as often as the remaining notes. Within a given trial, then, any alteration involving those notes would be subject to the frequency of occurrence confound, thus artificially inflating any results that might be obtained with respect to the crucial interaction.

The ratings data for each subject were submitted to a within-subjects ANOVA, using the ANOVA/MANOVA module of CSS-Statistica. The independent variables (KEY, POS, CHANGE, TYPE) were identical to the set described in the experiment above.

A main effect of TYPE was obtained, with mean ratings of 2.95 and 3.35 for unaltered and altered trials, respectively ($\underline{F}(1,41) = 35.31$; $\underline{p} < .00001$), indicating that subjects were able to differentiate altered from unaltered trials. As in the prior two studies, the curious yet consistent main effect of POS was also obtained, with mean ratings being 3.24 and 3.06 for second- and third-position stimuli respectively ($\underline{F}(1,41) = 8.05$; $\underline{p} < .01$).

Finally, a significant three-way interaction emerged, involving the KEY, CHANGE and TYPE factors, depicted in Figure 3.6. One can see from the figure that listeners' confidence levels for unaltered trials tended to remain relatively static as context changed, while confidence for altered trials was higher when the alteration consisted of a stable to unstable note in the respective key, relative to when that ordering was reversed.

The ratings data were transformed into proportions of hits and false alarms, and subsequent areas under the MOC were calculated. The resulting area scores were then submitted to a within-subjects ANOVA.

The sole significant result to emerge from the analysis was an interaction between CONTEXT and CHANGE ($F(1,41) = 5.76$; $p < .05$). This interaction, depicted in Figure 3.7, should be familiar to the reader by this time. Once again, on SU trials, C-major contexts were superior to D-minor with respect to detection sensitivity, while the reverse occurred for US alterations. Although the effect appears to be weaker than that obtained when triadic contexts were used, the crossover still persisted in the face of equal frequency of occurrences between contexts.

The most obvious explanation for the weakening of the effect in this experiment is that the frequency of occurrence differences in earlier work did contribute to the overall differences across CONTEXT levels. That the effect was still obtained here, however, indicates that frequency of occurrence differences could not account for the entire effect. The obvious candidate, of course, is the manipulation of the tonal hierarchies, which influenced the relative positions of the two target tones on those hierarchies.

One issue left largely unaddressed to this point was the consistent POS main effect for the ratings, which did not arise with the area scores. In general, subjects had a tendency to rate sequences with the target note in the second position with more confidence in the direction of 'different' than sequences where the target appeared in the third position. The fact that this effect was not obtained with the sensitivity scores rules out explanations relying on ease of detectability or memorability.

The difficulty in explaining this effect is in part due to the multiple ways confidence ratings data can be interpreted. On the one hand, one might say that extreme values are indicative of maximal confidence. However, the mean values obtained here were quite close to the middle of the rating scale (in the 'unsure' area), giving rise to a second interpretation. Specifically, it might be that the larger value is indicative of subjects having a tendency to respond "different", rather than reflect any actual confidence differences. That is, this effect might simply be the result of a response bias toward the 'different' half of the ratings scale. This notion is supported by the fact that the effect dropped out when a sensitivity measure designed to remove the influence of response bias was calculated. If this second interpretation is correct, then we are left with the problem of determining the cause of this response tendency.

The sole factor all three studies had in common was the usage of the same target sequences. It is possible that the targets themselves might contain the explanation for this consistent influence of position on confidence ratings. Upon reflection, it was speculated that a biasing influence might have been due to the successive intervals of which the target sequences were constructed. That is, on any given trial, if the comparison melody contained several relatively larger interval jumps, this might have led to a tendency to respond ‘different’ regardless of the standard melody, possibly due to its sounding less natural (since composed melodies tend to consist primarily of small interval jumps). Although the target sequences were constructed such that successive intervals were roughly uniform between target sequences, with no drastic (i.e. a perfect fifth or greater) interval jumps, equivalence of cumulative interval sizes across POS stimuli was not stringently controlled in the design of the target sequences for pragmatic reasons. Specifically, the number of constraints already imposed on the targets (uniformly increasing, with particular tones appearing in one of the two positions, and outer boundaries delimited by the context sequences) meant that the flexibility of sequence construction was already being taxed.

To further investigate this possibility, a tabulation of the successive intervals was carried out on the target sequences. Briefly, the numerical value of each interval which appeared in each target sequence was summed for second and third position stimuli. It was found that overall, target sequences in which the alteration appeared in the second position had a summed interval value of 156, while the summed value for third-position alteration sequences was 142. Although this is a small difference, it might have been enough to produce the small but consistent ratings differences observed in the POS effects obtained in the prior studies. One should keep in mind that the response tendency in question was actually quite small in terms of rating-scale units. Rigorous testing of this notion is left to others who might be interested in such a line of research, and will not be pursued further here, considering it did not enter into any interaction on the more crucial detectability measure. Instead, the focus will turn toward more pressing matters of alternative explanations for the crucial interactions involving the context key and the identity of the altered tones.

Figures

Figure 3.1: Four-note target sequences employed. The left column, labeled (A), corresponds to target sequences where the second tone of the sequence was altered, while the right column, labeled (B), corresponds to target sequences where the third tone of the sequence was altered. Accents were drawn for visual aid, and were not actually sounded in the experiments.

Figure 3.2: Three-way interaction between CONTEXT, CHANGE, and TYPE factors for ratings data in Experiment 1. Confidence ratings are greater for C-major trials over D-minor trials with S-U alterations, while the opposite result appears for U-S alterations.

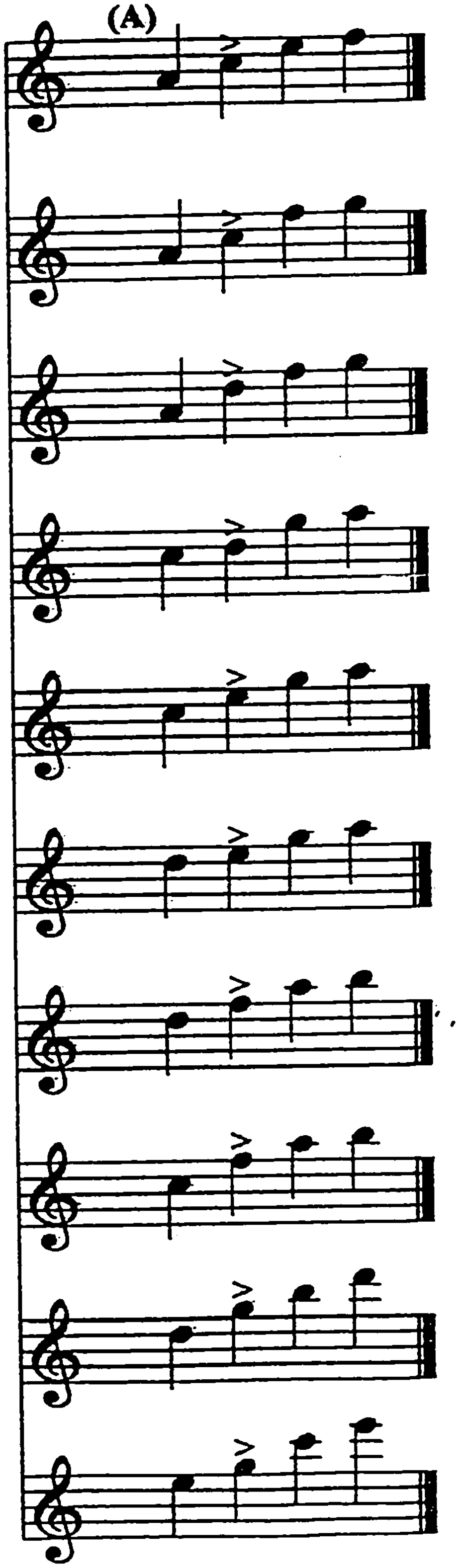
Figure 3.3: CONTEXT by CHANGE interaction for area scores in Experiment 1. For S-U alterations, C-major contexts resulted in superior detection levels relative to D-minor contexts, while the opposite occurred for U-S alterations.

Figure 3.4: Three-way interaction between CONTEXT, CHANGE, and TYPE factors for ratings data in Experiment 2. A similar pattern to that seen in Figure 3.2 is apparent here.

Figure 3.5: CONEXT by CHANGE interaction for area scores in Experiment 2. A more complete crossover is apparent.

Figure 3.6: Three-way interaction between CONTEXT, CHANGE, and TYPE factors for ratings data in Experiment 3. Again, a similar pattern to that obtained in Figures 3.2 and 3.4 is seen here.

Figure 3.7: CONTEXT by CHANGE interaction for area scores in Experiment 3.

(A)

(B)

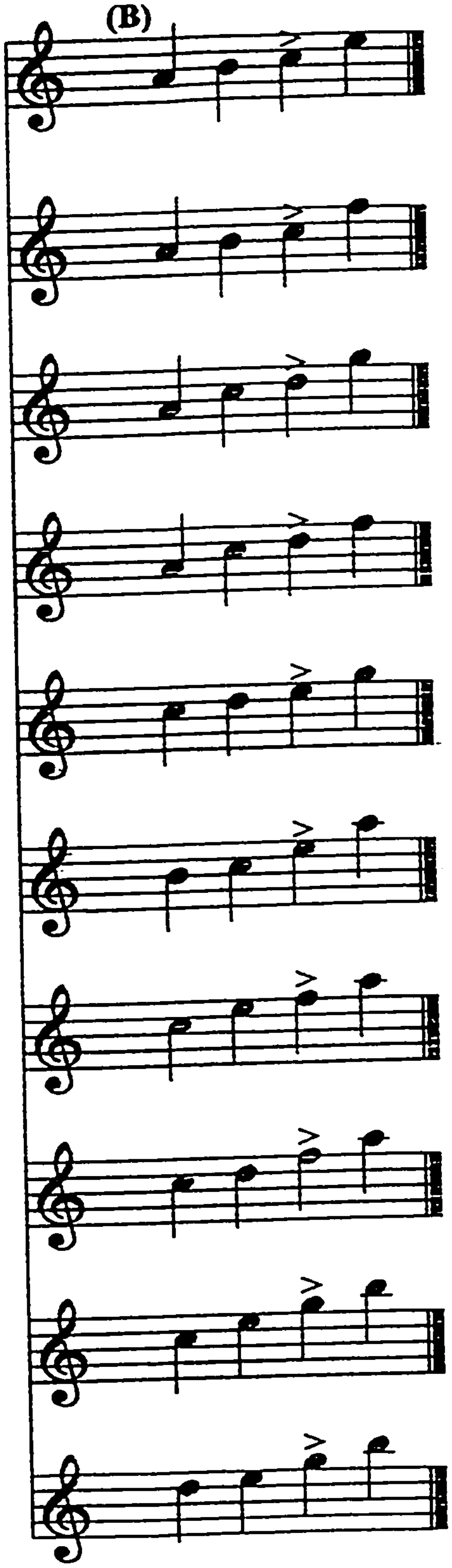



Figure 3.2

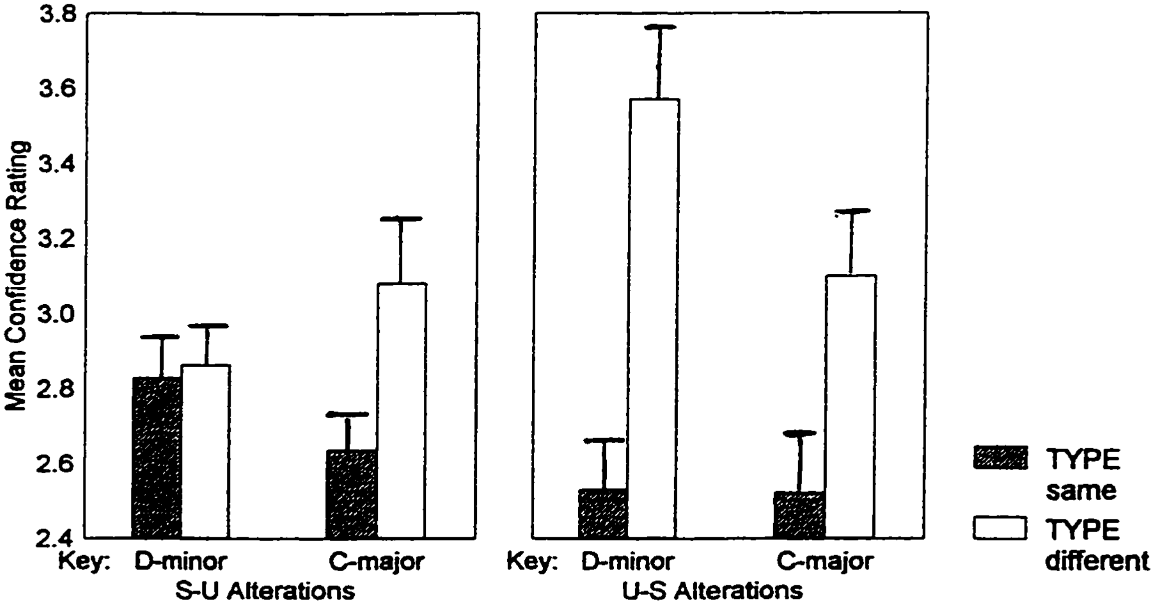

Mean Confidence Rating
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
Key: D-minor
C-major
S-U Alterations
Key: D-minor
C-major
U-S Alterations
TYPE same
TYPE different

Figure 3.3

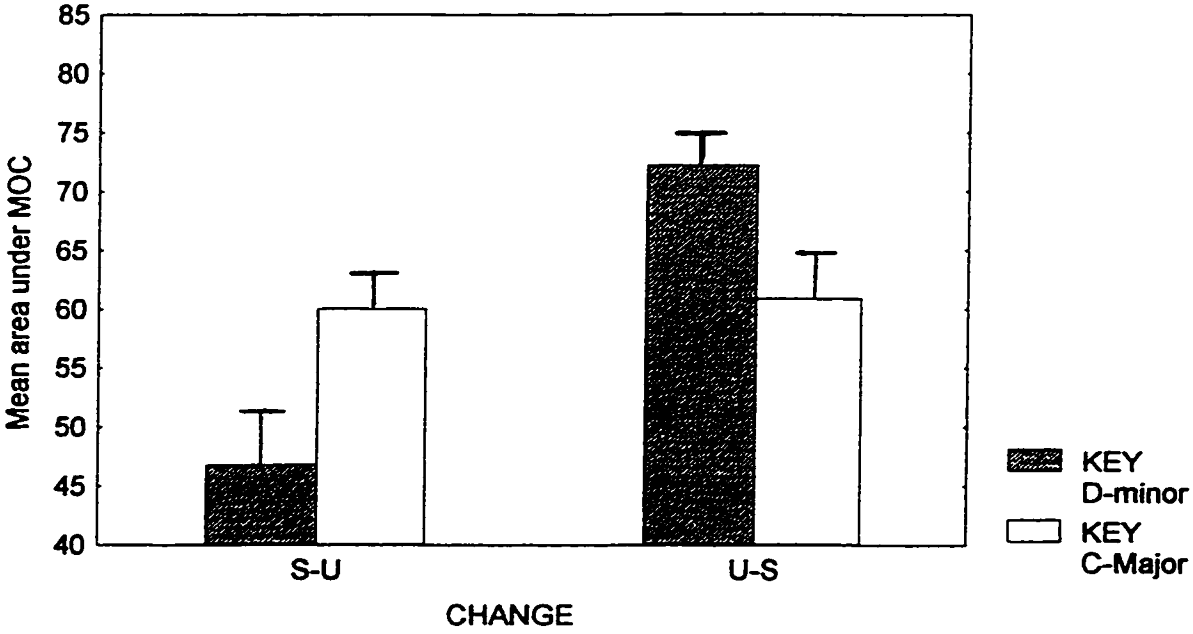
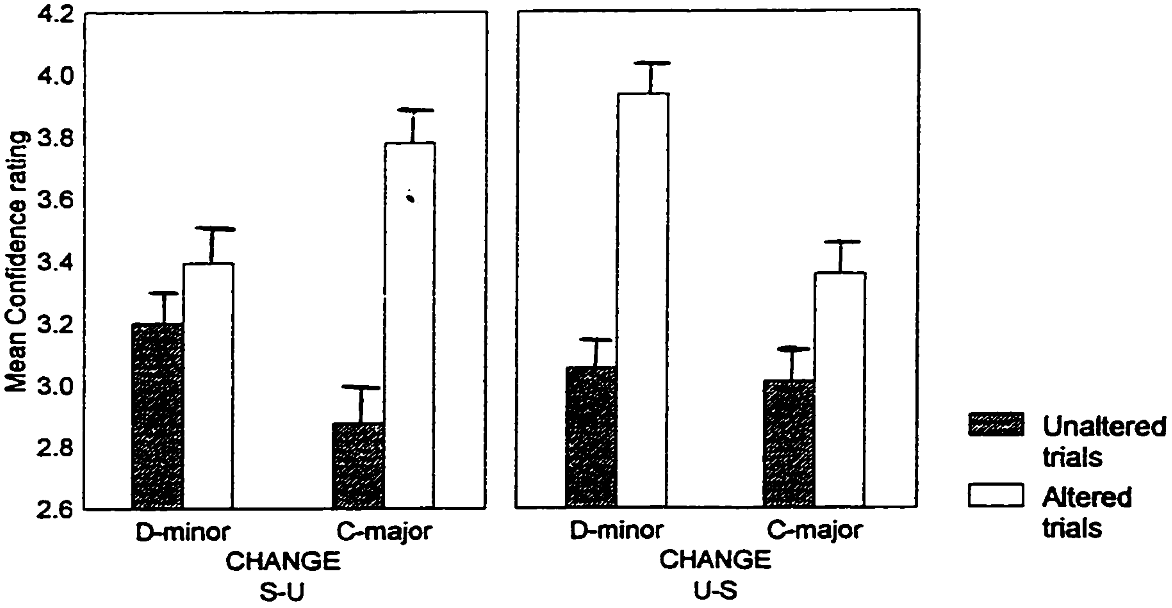

85
80
75
70
65
60
55
50
45
40
Mean area under MOC
S-U
U-S
CHANGE
KEY D-minor
KEY C-Major

Figure 3.4



Mean Confidence rating
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
D-minor
C-major
CHANGE
S-U
D-minor
C-major
CHANGE
U-S
Unaltered trials
Altered trials

Figure 3.5

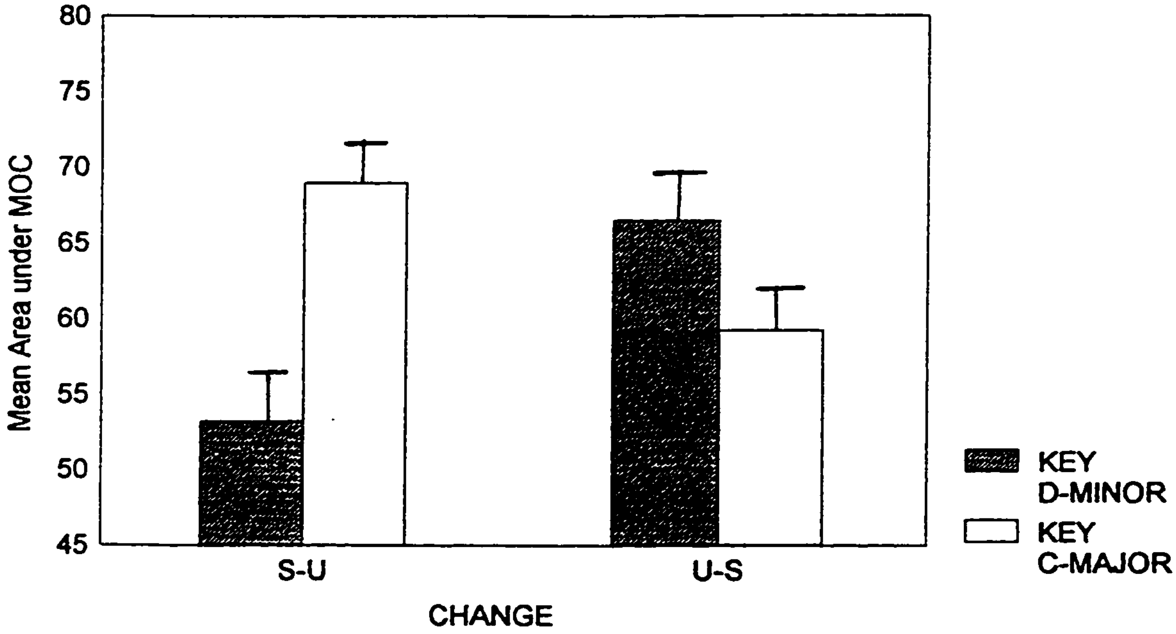
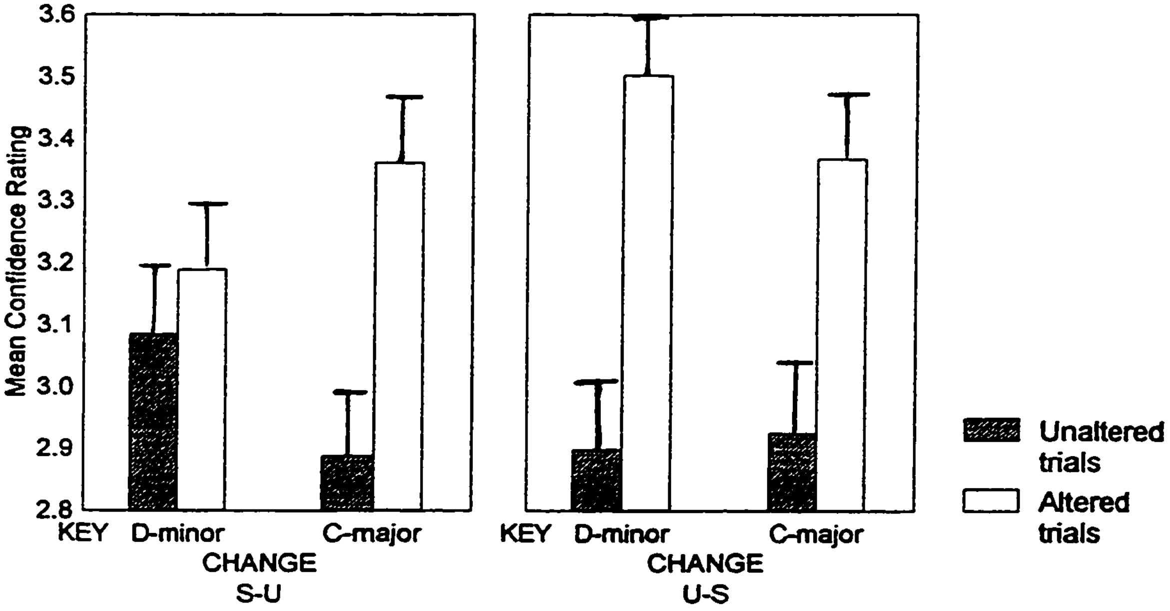
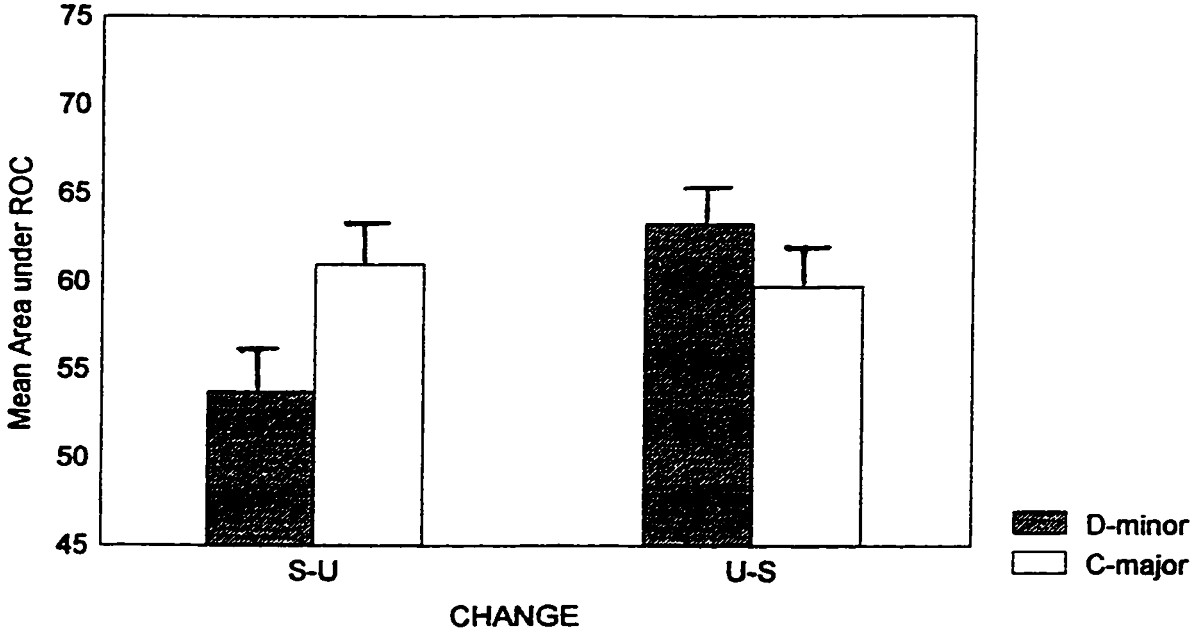

80
75
70
65
60
55
50
45
Mean Area under MOC
S-U
U-S
CHANGE
KEY D-MINOR
KEY C-MAJOR

Figure 3.6


Mean Confidence Rating
3.6
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
KEY D-minor C-major
CHANGE
S-U
KEY D-minor C-major
CHANGE
U-S
Unaltered trials
Altered trials

Figre 3.7


75
70
65
60
55
50
45
Mean Area under ROC
S-U
U-S
CHANGE
D-minor
C-major

Chapter IV. Control Studies

4.1 Introduction

Although the predicted differential sensitivity to a particular alteration between context keys was obtained under a variety of different context-setting conditions, several issues merit attention prior to drawing any convincing conclusions.

First, it is possible that there was something idiosyncratic about the target sequences themselves which was producing the obtained effects involving KEY and CHANGE. That is, it might be the case that the obtained interaction between CONTEXT and CHANGE came about as a result of the pre-composed stimuli. Perhaps it would not have mattered that the context was explicitly presented to listeners.

Second, there is still the issue of whether or not subjects truly were hearing the melodies in the proposed keys as determined by the experimenter. This issue is particularly acute for Experiment 3, where the context sequences consisted of a rearrangement of the component tones, in order to promote one or the other keys. Although these context sequences were, indeed, the first four notes of the diatonic scales of the respective keys, this by no means ensures that listeners perceived them in those keys. To address these issues, two additional studies described below were carried out.

4.2. Experiment 4: No context condition

If the results obtained were truly due to the effect of the opening context, then removal of that context should result in the observed asymmetry disappearing. If, however, the ordering of the trials (as indexed by the CONTEXT variable, perhaps) were somehow idiosyncratically producing the results, one should still see the asymmetry, or at least suggestions of it, in the data. Furthermore, there were several unusual effects obtained which were likely to have come about due to the target stimuli configurations, and not from the contexts. In particular, a main effect of POS was repeatedly obtained from the ratings data ANOVA. If it is the case that the target sequences were themselves producing this result, then such an effect ought to emerge when those sequences are presented in isolation.

Methods

Subjects. Twenty-one McMaster University undergraduates participated in the study, as partial credit for an introductory psychology course. The median age of participants was 19 years. Subjects were all musically trained, such that they received a minimum of grade 8 Royal Conservatory of Music training on an instrument or voice. All of the subjects in this study received their training with piano.

Apparatus. All stimuli consisted of pure tones generated by a Yamaha TX-802 FM tone generator, which was slaved to a Comptech 386-33 computer, running a QuickBasic program designed to present stimuli and record responses. All stimuli were presented to subjects through AKG-340 earphones. Subjects were seated in an IAC sound attenuation booth, and indicated their responses using a Logitech mouse. The computer controlled the tone generator, and recorded all responses.

Stimuli. The melodies in this study consisted solely of the four-note target sequences which were used in the previously described experiments. To ensure that temporal differences would not contribute to the results, the four-note context sequences and the three-note terminal sequences were replaced by silent periods of equal durations.

Procedure. Following a brief interview to determine musical background, subjects were seated in the IAC chamber, and were told they would hear two melodies, with the task being to detect if the melodies were identical or not. The use of the six-point confidence scale was explained to them (1=very sure same, 2=sure same, 3=guessing same, 4=guessing different, 5=sure different, 6=very sure different). Subjects were told to reserve the end points of the scale (1 and 6) for when their confidence was maximal, and to use the ratings closer to the middle when confidence was lower.

Order of presentation was randomly determined for each subject, using the subject number as the seed for the random number generator in the program. Subjects pushed a button on the mouse to start the experiment, and following the presentation of the two melodies on a trial, displayed the confidence scale and waited for subjects to select and input a rating. Once a rating was indicated, a brief graphic indicating the selected rating was presented, taking approximately three seconds, after which the screen was cleared, and the next set of melodies was presented.

In order to ensure that subjects were paying attention and were interested in the task, the number of correct responses obtained in each ten-trial block was displayed. This score was based on whether a listener's ratings were on the appropriate side (same or different) of the rating scale relative to whether the trial-type itself was the same or different.

Several predictions concerning the results may be advanced. First, one would expect this task to be much simpler for subjects to carry out, since only 4 notes are heard instead of 11. It is therefore expected that subject confidence should be much higher in this study relative to confidence obtained in the previous work. In other words, the distribution of ratings are expected to cluster around the end-points of the rating scale. Likewise, one would expect that subjects' accuracies at detecting an alteration ought to be higher than in previous work, again, considering the ease of the task. Finally, the interaction between KEY and CHANGE factors should disappear when these stimuli are employed. In fact, any effect involving the KEY factor ought to disappear with the removal of this factor, since an overt key context is no longer present.

Results and Discussion

All subject data was submitted to a MANOVA, using the ANOVA/MANOVA module of CSS-Statistica. There were two significant main effects. The first was the expected TYPE main effect ($\underline{F}(1,20) = 466.68$; $\underline{p} < 0.0001$). Unaltered trials were associated with significantly lower ratings (mean = 1.90) than were altered trials (mean = 5.26). As predicted, this difference between mean ratings is much greater than that observed in prior studies, and is reflective of the greater ease of this task for subjects. The only other significant effect, surprisingly, involved the key-context variable, which was not really part of the design, in the sense that the context stimuli were not presented to subjects. The mean rating for a 'D-minor' context was 3.62, while the mean rating for the 'C-major' context was 3.54 ($F(1,20) = 7.77$; $p < 0.05$). The only possible explanation for this small result, considering the variable in question did not really vary in the experiment, is that something about the order of presentation resulted in the slightly lower ratings. Since the order of presentation was randomised for each subject, no systematic explanation readily comes to mind, other than that the randomized ordering collectively contributed to an overall weak order effect. As well, this differential rating for key was not obtained in other experiments where context was explicitly presented to subjects, and was undoubtedly overshadowed by the more salient key information that was overtly provided in previous experiments.

The only interaction to emerge from the study involved the target position (POS) and trial-type (TYPE) factors. Specifically, targets in the second position of the sequence were associated with a greater differential between "same" and "different" trial-types than were targets in the third position of the sequence. This interaction appears to reflect a slight increase in subjects' confidence for trials involving alterations in the second position relative to the third position.

No other effect or interaction was significant in this experiment. As a result, it appears that any effects involving the asymmetry arose for the most part as a direct consequence of the context stimuli being present, and not on some order influence or other idiosyncrasies of the stimuli.

The ratings data were converted to proportions of hits and false alarms, and area scores were computed for each subject in each condition. The resulting scores were submitted to a repeated measures ANOVA.

The position of the target (POS) was the only influencing variable on subjects' detection abilities (F(1,20) = 6.93; $p < .05$). Targets appearing in the second position of the sequence were associated with slightly higher area scores (mean Ag = 93.96) than were targets appearing in the third position (mean Ag = 92.65). Target position appears to be a more salient feature here, relative to earlier work, where a surrounding context essentially buried the targets into the middle of the melody. In this study, however, only four notes made up the test sequences, so subjects may have focused more attention on the second position's identity relative to the third, resulting in a slight drop in detectability of the latter.

More importantly, there was no evidence of asymmetric similarity arising, which is unsurprising, given there was no specific tonal context provided, beyond the four notes which made up each stimulus. Those notes were drawn from the C-major scale, so one might expect there to have been, if anything, an asymmetry based on the C-major tonal hierarchy. However, possibly due to the impoverished nature of this impromptu "context", activation of the hierarchy did not take place with enough strength to observe it in the response measures.

While it is possible that a ceiling effect might be responsible for the lack of asymmetric relations emerging, several aspects of the study cast doubt on this conclusion. First, one might at least hope to find some semblance of an asymmetric pattern emerging from the data, albeit a nonsignificant one. That is, in spite of the responses being accurate, they were not perfect, such that the ceiling influence would serve to obscure any asymmetric relation, but not necessarily completely. However, the graph for the (nonsignificant) interaction reveals a pattern nearly opposite to the asymmetry observed in prior studies. In particular, the U-S alterations were seemingly more accurately detected for the C-major context, relative to the D-minor context. The opposite pattern was obtained in the studies presented earlier, when a more explicit context was introduced. Obviously, the lack of a significant effect prohibits drawing any definitive conclusions concerning this interaction. However, even when examining the response patterns of those listeners who clearly were not performing at perfect levels (e.g. an average A' of less than .80), no response tendency resembling the pattern observed in the other three studies emerged. That is, even for the listeners performing below what could be called 'ceiling' performance, asymmetric perception did not arise.

A second aspect of the study that calls into question the ceiling effect argument concerns the nature of the stimulus presentation itself. Specifically, the "C-major" and "D-minor" levels of the CHANGE variable did not, in this study, exist. Both levels were identical to one another in all respects, except for the order of presentation in which they appeared across trials for each subject. Since that order consisted of an intermingling of both types of trials, one is hard-pressed to find any rational explanation for subjects systematically responding differently between the two CHANGE levels.

In summary, then, it appears that the removal of the explicit tonal contexts from the stimuli employed in the previously-reported studies resulted in a disappearance of asymmetric similarity perception, lending support for the notion that those contexts were the determining factors in the asymmetric pattern of ratings.

4.3. Experiment 5: Testing for key identification

While the asymmetries observed in Experiments 1 and 2 generally conformed to the hypothesis that the differing tonal hierarchies being activated were responsible for the observed sensitivity differences between key contexts for a given alteration, there has thus far been no other corroborating evidence to support the conclusion that subjects actually heard the "C-major" and "D-minor" melodies in the keys of C-major and D-minor, respectively. This experiment was conducted in order to confirm that subjects were hearing these melodies in the keys specified by the experimental design. The melodies used in Experiment 3 were presented individually to subjects, and were followed by a chord probe. Subjects were asked to rate how well the chord represented the key of the previous melody. Probes consisted of C-major, D-minor, and their relative complement keys (A-minor and F-major). Although it would have been better for subjects to rate all melodies against all possible major and minor chords, it would have entailed exposure to a minimum of 480 trials (20 melodies followed by each of 24 chords), which was simply too much to ask of even the most patient listener. A smaller set of possible comparison chords was necessary.

Methods

Subjects. Fifteen McMaster University undergraduates participated in the study, as partial credit for an introductory psychology course. The median age was 19 years. All subjects had reported receiving a minimum of grade 8 Royal Conservatory of Music training on an instrument or voice, and were therefore classified as moderately trained musicians.

Apparatus. All stimuli consisted of pure tones generated by a Yamaha TX-802 FM tone generator, which was slaved to a Comptech 386-33 computer, running a QuickBasic program designed to present stimuli and record responses. All stimuli were presented to subjects through AKG-340 earphones. Subjects were seated in an IAC sound attenuation booth, and indicated their responses using a Logitech mouse. The computer controlled the tone generator, and recorded all responses.

Stimuli. The test melodies (context and target sequences) from Experiment 3 were presented here. The chord probes consisted of the root chords of C-major, D-minor, F-major, and A-minor. All chords were presented in the same octave as the melodies (middle C to C-above middle-C), using pure tones. The additional 'lure' chords (F-major and A-minor) were selected to test for two possible confounding factors in the stimuli.

First, these two lure probes were chosen to investigate the possibility that the melodies were being heard as merely major or minor, without reference to a particular key. If that was the case, such that listeners in the earlier experiments heard the streams as major or minor without necessarily assigning a specific key, then the F-major probe should have higher ratings in the C-major context than in the D-minor context, while the opposite outcome should occur for the A-minor probe.

Second, the two lure chords were chosen on the basis of their relatedness to the hypothesized 'true' context chords (C-major and D-minor). One can specify two keys as related if their component diatonic tones are identical. The tones making up the F-major diatonic scale are identical to those of D-minor; the same may be said for the relation between C-major and A-minor. Musicians, then, would refer to A-minor as being the 'relative minor' of C-major. The only difference between the scales of two related keys is their order of presentation: a C-major diatonic scale begins and ends on the note C, while the A-minor diatonic scale begins and ends on the note A. In the case of a melody written in one or the other key, this information might be conveyed by a combination of temporal ordering, stylistic regularities, and frequencies of occurrences of specific tones, as discussed earlier in the Introduction.

The choice of lure chords indicative of relative keys of the hypothesized true keys was intended as a finer test of the extent to which listeners hear each stimulus in a particular key. If there is insufficient information concerning key being conveyed by these stimuli, then one might expect listeners to report the streams as being in the lure key in addition to the context key specified by the experimenter; The ratings for lure chords should resemble those of their relative chords. In our case, this means that the pattern of ratings for C-major and A-minor should be similar to one another, as should those for D-minor and F-major. If, however, a strong sense of key is conveyed, then similar patterns for related key chords should not appear, such that the lures would be associated with significantly lower ratings relative to the true context chords.

Procedure. A typical trial consisted of a single melody being followed by a one-second silent interval, and then one of the four chord probes. Subjects were asked to rate how well the chord represented the key of the melody, using a six-point rating scale (1=very bad, 2=moderately bad, 3=mildly bad, 4=mildly good, 5=moderately good, 6=very good). All melodies were presented with each of the four chord probes.

Subjects were instructed to base their response on key representativeness, and to avoid responding on the basis of how well the chord 'finished' the melody, or how pretty/interesting a cadence was formed. These instructions seemed necessary, due to pre-test listening by the researcher of the various trial-types. Specifically, while listening to the trial types to ensure the program was performing properly, it was noted that several of the melody-chord pairs formed what can best be described as 'catchy' chord progressions. For this reason, subjects were strongly instructed to avoid using this as a criterion for fitness ratings.

The 20 melody types from Experiment 3 were the test melodies. Each was followed by each of the chord probes, for a total of 80 trial types, and presentation of each combination was repeated once, for a grand total of 160 trials. The independent variables melody CONTEXT (C-major/D-minor), probe CHORD (C-major, D-minor, F-major, A-minor), CHANGE, and POS. These last two variables were included as coding variables to ensure similarity between this and the previous experiments in the analysis. Of interest is whether any of the same interactions obtained with the same-different task would be obtained when it came to key membership judgments.

Results and Discussion

The ratings data for all subjects and trial-types was submitted to a within-subjects ANOVA, using the ANOVA/MANOVA module of CSS-Statistica.

A main effect of CHORD was obtained ($\underline{F}(3,42) = 12.93$; $\underline{p} < .00001$). The mean ratings for each chord type are depicted in Figure 4.1, where it can be seen that overall, A-minor chords received the lowest ratings, while C-major received the highest. This factor produced an interaction with CONTEXT, and is shown in Figure 4.2. One can see that for D-minor melodies, the D-minor chord received the highest rating, while for C-major melodies, the C-major chord was rated as most representative. The fact that this effect is stronger for the C-major than D-minor melodies is undoubtedly a result of competing cues for key membership in the D-minor melodies. In other words, although the context sequences were promoting a D-minor key, having begun and ended on the note D, the notes were actually entirely drawn from the C-major scale. No such competition occurred for C-major melodies, resulting in a stronger set of ratings. Finally, one can clearly see that a similar pattern did not emerge for the F-major or A-minor chords, indicating that listeners were not hearing the stimuli in their related keys.

Since the main chord ratings of interest were those of C-major and D-minor probes, a second ANOVA was performed on those CHORD types alone. Upon doing this, it was found that the CHORD main effect disappeared. Apparently, then, the CHORD main effect obtained in the first ANOVA was undoubtedly due to the very low ratings assigned to the A-minor chords.

Continuing with this second ANOVA, the interaction between CHORD and CONTEXT persisted ($\underline{F}(1,14) = 11.59$; $\underline{p} < .005$). Examining the left half of Figure 4.2, one can see that for D-minor contexts, the D-minor chords were rated as more representative than were the C-major chords, while the reverse occurred for the C-major contexts. So, the assumption of key membership of the respective context types was supported here, such that subjects were hearing the melodies in the keys as defined by the experimenter.

Figures

Figure 4.1: Main effect of chord-probe type on fitness ratings in Experiment 5. This main effect appears to have been due to the low ratings of the A-minor probe relative to the other chords.

Figure 4.2: Chord-probe by context-key interaction in Experiment 5. Listeners show a clear bias to choose the D-minor probe as fitting the D-minor context, and the C-major chord as fitting the C-major context. When only the D-minor and C-major chord-probes were analyzed, the effect was still highly significant.

Figure 4.1

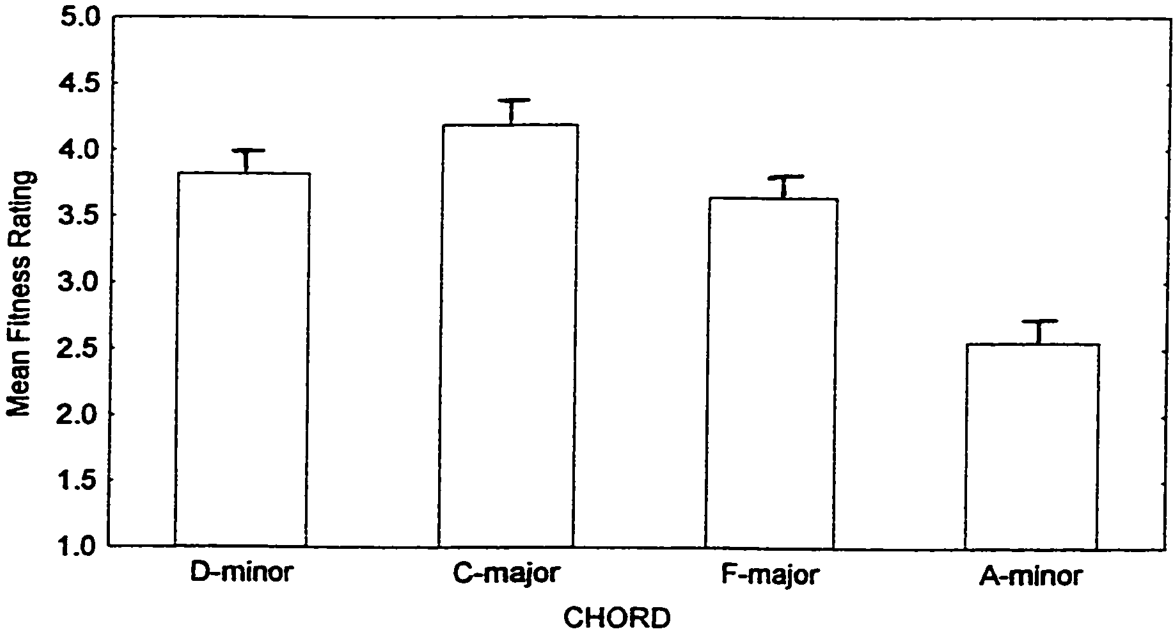
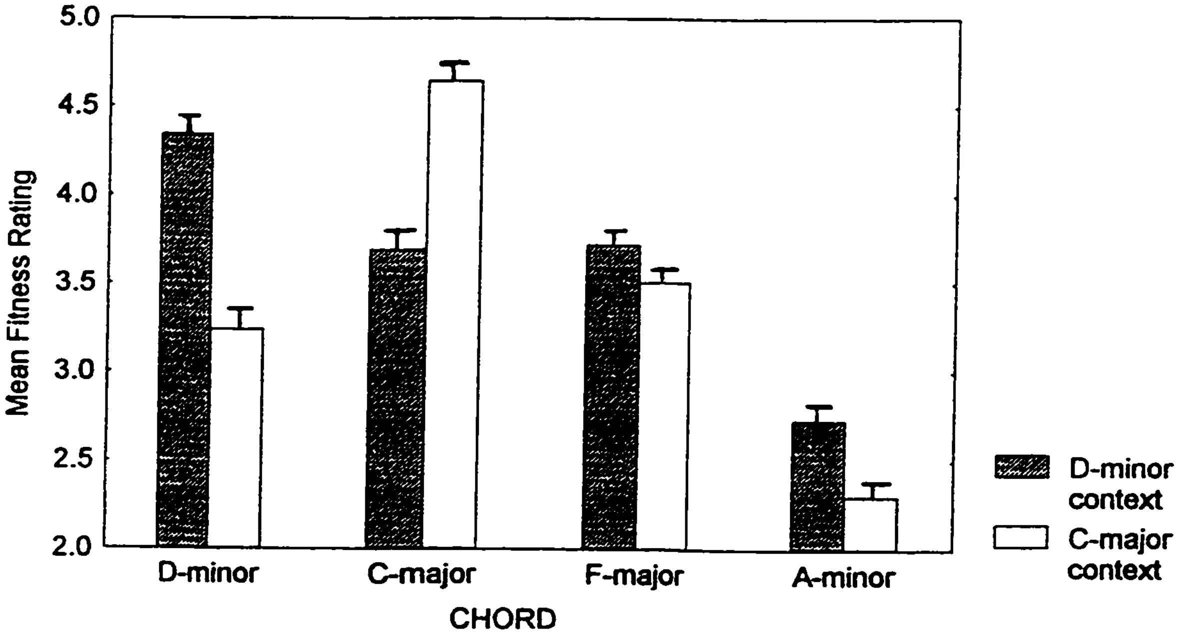

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Mean Fitness Rating
D-minor
C-major
F-major
A-minor
CHORD

Figure 4.2


5.0
4.5
4.0
3.5
3.0
2.5
2.0
Mean Fitness Rating
D-minor
C-major
F-major
A-minor
CHORD
D-minor context
C-major context

Chapter V. General Discussion

In general, two main conclusions may be drawn from the work that has been presented. First, it is clear that the perceived alternatives hypothesis, as it was stated in Bartlett and Dowling (1988) is not able to adequately account for the effects obtained here. Alterations which did not violate the tonality of the piece were associated with differential detection sensitivities depending on the tonal hierarchy being activated. Second, this work demonstrates that the position a tone occupies in the tonal hierarchy can indeed influence detection ability in a pattern-matching task, counter to Browne's proposal. This can best be seen in Experiment 3, where the frequency of occurrences of notes was controlled between context types, yet context still produced differential sensitivity to particular tone alterations. In other words, when the position of a tone in the tonal hierarchy was systematically manipulated, systematic changes in detection sensitivity emerged. Considering the perceived alternatives hypothesis cannot explain these data, another model becomes necessary.

There are two potential explanations for this effect. First, it might have been the case that the target tone in the first melody was encoded more accurately when that tone represented a highly stable as opposed to a less stable member of the instantiated tonal hierarchy. Consequently, listeners would have access to a better memory representation of a more stable note relative to that for a less stable note, resulting in better detection accuracy. As evidence in favor of this notion, Krumhansl (1979) presented listeners with a standard and comparison tone, and interpolated a diatonic or nondiatonic sequence between them. She found that diatonic standards were associated with much higher alteration detection accuracies than were nondiatonic standards, when the interpolated sequence was itself diatonic. She concluded that this was evidence that the memory representation for highly stable elements was stronger than for less stable elements. As well, she hypothesized that the less stable elements might, over time, have a tendency to be transformed in memory toward more stable elements. Although her experiment had several flaws, which were highlighted in the Introduction, it does not rule out the possibility that this mechanism was operating in listeners to bring about the asymmetric relations obtained in both her work and that presented here. Bartlett and Dowling (1988), however, disputed this notion by pointing out that if it were the case that scalar elements are more accurately encoded, then a scalar to scalar alteration ought to have a low associated similarity rating, which is the opposite to what they obtained.

A second possible explanation does not rely on the memory representation per se, but rather examines the relative saliences of the two target elements in the tonal context. In other words, the ability to detect an alteration will depend on listeners' sensitivity to the relative stability weightings assigned the target tones. As pointed out in the Introduction, Tversky's (1977) model appears to be amenable to explaining the asymmetries described here. However, several ambiguities exist with respect to the details of its application. First, while the stimuli used in the work presented here were clearly directional in nature, it is unclear as to which stimulus was seen as the subject, as opposed to the referent. Usually, posing a question such as 'compare A to B' clearly implies that A is to be compared to B, and not the other way around. In this case, stimulus A would be the subject, and B the referent of the comparison. It might appear obvious, then, that when presented with melody A and melody B for comparison, listeners should see melody A as the subject, and B as the referent. However, it is equally possible that the situation is seen by subjects in the opposite direction. That is, when two melodies are presented sequentially, it is likely that listeners will hear and store the melody as best they can, then compare the second melody to the first. Construing the situation in this fashion means that the first melody is the stimulus 'already on the blackboard', while the second melody is the stimulus 'handed to the subject' after seeing the first one. In other words, the first melody would act as the referent, while the second melody acts as the subject, resulting in predictions opposite to those implied by the reverse situation. While Tversky (1977) cites some evidence from auditory tasks which appear to indicate that listeners were taking the first stimulus as the subject, and comparing it to the second stimulus (or referent), the studies he reported (Rothkopf, 1957; Wish, 1967) dealt with auditory signals of varying length. The importance of this lies in his assumption that the length of the signal determined the signal's salience, in particular, that short signals are less salient than long signals. Only under this assumption does it then becomes possible, based on the direction of the asymmetry observed in the data, to conclude that observers saw the first stimulus as the subject, and the second as the referent. This does not imply that employing auditory stimuli will always result in observers taking the ordering as subject-referent. Recall that in her original study, Krumhansl found that the second tone's identity had a greater influence on the observers' pattern of ratings than did the first tone. This is arguably evidence that, with her task, the second tone was serving as the referent. As well, if Tversky's assumption is wrong, then it implies that observers in Rothkopf's and Wish's studies must have seen the first stimulus as the referent, and the second as the subject. In fact, it appears that determining which stimulus serves as subject versus referent seems to largely depend on what one means by 'salience'.

According to Tversky, salience depends on the number of distinctive features a stimulus contains, relative to the other stimuli in a given set. The greater the number of distinctive features a stimulus has, the more salient it is said to be. So, for example, when comparing a square to a rectangle, the associated similarity will be greater than, say, comparing a square and a circle, since in the latter case, the two stimuli share far fewer common features than the former. While this notion is quite apparent when visual stimuli are employed, the issue is somewhat less cut-and-dry in the auditory domain. The psychological correspondent of frequency, referred to as 'pitch', is multidimensional in nature, as depicted in the models described in the Introduction. According to several of those models, as well as music theory in general, when a key is instantiated by a set of tones, this gives rise to dynamic tone quality, wherein particular tones are endowed with different degrees of functionality in the melody, in terms of their structural importance to the key in question. Krumhansl's conical structure is a useful thumbnail sketch of these different sets of tones, and from this structure it can be seen that the root-chord tones are highly structurally stable, the remaining diatonic tones less so, and those tones not forming part of the diatonic scale least of all. It is here where one might choose to define 'salience' in one of several ways.

For example, one might make the claim that the most structurally important tones must also be the most salient. However, this description does not seem to capture what was meant by Tversky's definition, since dynamic tone quality differences persist when features such as intensity, timbre, and duration are constant across tones. In other words, one need only manipulate frequencies, to observe dynamic tone quality. Frequency, however, is unidimensional. Although 'pitch', or the perception of frequency, is said to be at least bidimensional (Shepard, 1982), the fact remains that a strict usage of Tversky's definition of salience should give rise to equivalent levels of perceived similarity regardless of the direction of the comparison. To illustrate, consider two tones, 1000 Hz and 1010 Hz, presented for similarity comparison. What features do these stimuli have in common? What 'stimulus set' serves as the context? It quickly becomes apparent that it is impossible to say what the common or distinctive features are, if Tversky is strictly applied. To state that one or another of these tones is more salient on the basis of its being more structurally important in the context of a melody's key is inappropriate, since there is no evidence to back up that claim.

In fact, it is obvious to any music listener that just the opposite claim seems more appropriate. That is, the least stable elements appear to be highly salient in a piece of music, such that when an out-of-key tone is sounded in an otherwise diatonic melody, it stands out to the listener. Even in the absence of an appropriate explanation as to why this phenomenon occurs (although several come to mind, including expectancy violation), this appears to be the more appropriate definition of salience.

What we are left with, then, is a model which resembles Tversky's original idea, but with several features vastly different from the original conceptualization. First and foremost, the idea of salience is not blindly described as 'most structurally important', since it is apparent that this is not the case when listening to music. Rather, salience is an inverse function of structural stability, such that the more stable a tonal element, the less salient it is. Second, the issue of which stimulus is considered the subject and which the referent appears resolved by considering what listeners must do in order to make the comparison: Upon presentation of the first melody, subjects attempt to store and remember it in anticipation of comparing the second melody to that representation. In this fashion, it can be seen that the first melody is the referent, and the second melody the subject. Given this, the salience of the target tone in the second melody will carry more weight than that of the first in the comparison, due to the focal attention paid it by the listener, according to Tversky's model.

In this form, the model quite nicely captures the data presented in this thesis. For the SU comparisons, the first melody contained a more stable element, and the second melody a less stable one. The less stable element's salience would be intensified by the focal attention being given that stimulus by the listeners, since it is considered the subject of the comparison, resulting in a low similarity rating. For the US comparisons, however, the first melody contains the less-stable element, while the second melody contains the more-stable one. This greater stability, though, is associated with low salience, and as such, the focal attention would give rise to a lower outcome of total salience as a result of intensification, resulting in higher levels of perceived similarity.

This model also nicely captures Bartlett and Dowling's (1988) set of findings. They found that SN comparisons resulted in lower similarity ratings that NS, and also that NN comparisons resulted in lower similarity ratings than SS (a result which they were hard-pressed to explain with the perceived alternatives hypothesis). Examining the SN stimulus set, the second stimulus contains the highly salient nonscalar element, and due to focal attention, this salience will be further enhanced, resulting in a low similarity rating. For the NS pair, however, the second melody contains only scalar elements, resulting in lower salience. Consequently, the influence of focal attention on the second melody will be attenuated relative to the SN situation, causing the overall similarity ratings to be higher. Turning to the SS/NN data, the model predicts that SS pairs should receive higher similarity ratings than the NN case. For the SS melodies, focal attention will result in the second melody having an associated intensified salience. However, the second melody contains only scalar elements, and as such, the influence of focal attention will be small. For the NN comparisons, the second melody contains the highly salient nonscalar element. This already-salient feature will be further intensified due to focal attention, resulting in a greater absolute differential in relative saliences than was obtained in the SS case. The end result, then, is that the SS pairs should be perceived as highly similar, while the NN pairs should be perceived as somewhat less similar.

To clarify what has been said thus far, a simplified example might help. Assume that S melodies have an associated salience of 1, while N melodies have an associated salience of 2. Also, assume that the effect of focal attention is to cause the salience of the feature receiving the benefit of focal attention to be enhanced by a factor of 2. Using these parameters, in the NS condition, the N melody has a salience of 2, while the S melody has a salience of 1, which will be enhanced by focal attention, resulting in a salience of 2. The two stimuli should therefore be perceived as highly similar. For the SN pairs, however, the N melody would have an overall salience of 4, since focal attention will enhance the already-salient N melody. As a result, the difference between saliences for these melodies is greater (4-1=3) than in the NS case (2-2=0). For the SS/NN situation, the SS pairs would have an associated salience differential of 1 (2-1), whereas for the NN situation, the associated salience differential would be 2 (4-2), resulting in the NN pairs being heard as more different than the SS pairs. It must be pointed out that even without attempting to derive more accurate values for the various salience levels, the model adequately describes the data obtained in Bartlett and Dowling (1988) as well as that presented here.

There are several possible reasons for the asymmetric effect being stronger for tonic triad as opposed to when frequency-controlling contexts. First, it might be that the triad more strongly elicited the appropriate tonal hierarchies in listeners, resulting in a stronger crossover effect. Several researchers have found that the tonic triad does give rise to strong tonal hierarchies, as indexed by tone-profiles from probe-tone studies with a triad context (Krumhansl & Kessler, 1982). Therefore, the broken triad context stimuli may have simply been more effective 'elicitors' of the tonal hierarchy than were the other context types employed. A similar explanation is that the differing contexts were associated with differing salience levels as described above. That is, the actual values for saliences were possibly different for triad contexts as opposed to the frequency-controlling contexts, as a result of their differing context-setting capabilities.

A second possible explanation involves the frequency of occurrence confound repeatedly mentioned in the earlier text. It is possible that movement toward or away from a previously sounded note had an influence on detection ability. This argument is weakened by the fact that the context sequences were in a different octave from the targets in Experiment 2, but the asymmetry still emerged. However, on the basis of the frequency-controlling contexts of Experiment 3, one can abandon the possibility that frequency of occurrences produced the entire effect. At best, the differential densities of root-chord and non root-chord notes served to enhance the effect, but cannot completely account for it.

Finally, it was demonstrated in the last two experiments that the removal of the context sequences resulted in a disappearance of the asymmetry, and that the melodies employed in Experiment 3 were in fact heard in the keys defined by the experimenter. This further reinforced the notion that the tonal hierarchy was playing an active role in determining listeners' responses in the earlier experiments.

Based on these data, it becomes necessary to re-think the role of dynamic tone quality in the processing of melodic information. Some have relegated the existence of the tonal hierarchy to that of an epiphenomenon of note-repetition, allowing it no role in either key identification (Brown, 1988; Butler, 1989) or short-term melody recognition (Browne, 1981). Although the author is willing to grant that the tonal hierarchy's activation may not play an active role in key identification, it appears that key identification does give rise to a set of dynamic tone quality functions, which in turn have measurable effects on detectability in particular, and melody perception in general.

One can find similar evidence in the literature, in which subjects are asked to perform a same-different comparison between two melodies, varying along the dimension of tonality (Dowling, 1990). When tonality is degraded, corresponding degradations occur in a listener's ability to detect an alteration, even when that alteration is based on a contour violation (which is considered to be a very salient feature of melodies). In addition, melodies that do not conform to diatonic structure are difficult to remember/recognize (Dowling, 1990; Dowling and Fujitani, 1970). Part of the reason for this difficulty might be because subjects were not able to access a tonal hierarchy with these stimuli, and were basing their responses on less information than would have been available with melodies adhering to diatonic structure. This information comes in the form of relative importance of each incoming note event, with respect to the key of the melody. Use of this information in organizing the melody in terms of structurally important/nonimportant notes would not take place in melodies with ambiguous tonalities, thus contributing to the confusion experienced when performing a paired-comparison task. In order to clarify what is meant here, the reader is asked to consider memory in the language domain. If a sentence is constructed using standard rules of grammar and syntax, it will be much more readily memorized than if the sentence violates those rules. The more violations, or less grammatical, the less well-remembered/recognised the sentence will be. The reader will be confused as to the subject of the sentence, what action was taken, and so on. In the same way, when diatonic scale structure is violated (drawing constituent elements of a melody from the chromatic scale) it is tantamount to presenting a sentence consisting of unrelated words. The representation of that melody will be fragmented at best, and listeners will not be able to sort out the 'meaning' or direction the melody is taking, and therefore be less able to perceive the melody as a coherent whole.

In fact, one can call into question, on the basis of prior research and rational thought, studies which attempt to investigate the processing of melodies with nondiatonic stimuli. It is known that deviations from scalar structure result in associated melodiousness ratings dropping (Cuddy, Cohen, and Mewhort, 1981), as well as a sharp decline in retention of these streams (Dowling & Fujitani, 1971). In other words, subjects appear to cease perceiving these stimuli as music at all, regardless of whether the construction of those 'melodies' was rule-governed. To extend one's conclusions from this work into the domain of day-to-day music processing is not justified, since the stimuli employed are clearly not what subjects consider to be music. In a similar vein, experiments employing 'musical stimuli' consisting of two- or three-tone sets would rule out the possibility of discovering the interplay between the many factors contained in a more complete, ecologically valid musical stimulus. That is, one cannot be sure that a factor which, in isolation, might produce measurable effects would have the same influence when more cue-rich features typical of written music are also incorporated into the stimuli.

The thrust of this thesis was not to promote the role of dynamic tone quality to the top of the 'influential factors' heap in the processing of melodies. Such a single-factor view would be narrow minded to say the least, given the growing body of evidence indicating the multidimensional and interactive nature of melodies and the features of which they are constituted. Rather, the intent was to demonstrate that the hierarchical nature of representations of musical notes can influence recognition memory, and should not be neglected or ignored when discussing these types of tasks. Bartlett and Dowling (1988) did just that, however, by neglecting to acknowledge that their stimuli, while maintaining or violating scalar structure, also consisted of alterations which came from vastly different levels of the tonal hierarchy, and with vastly different levels of salience. One might attempt to salvage the perceived alternatives hypothesis by including further subdivisions of the scalar set, such that the scalar set could be broken out into root-chord and non root-chord elements. However, such an account still could not explain the results obtained in the work presented here, considering the stimuli consisted of melodies which were made up of a mixture of root-chord and non root-chord notes (hence always conforming to the same set of perceived alternatives). That is, altering a single note from root-chord to non-root-chord status should still have no influence on detectability, since the rest of the two melodies consisted of a mixture of these two note types, thus being part of the same set of perceived alternatives.

The research presented here is not subject to the same criticisms levelled at Krumhansl and her colleagues for using the probe-tone technique. The task mandated to listeners here was unambiguously circumscribed, so that no alternative interpretations of 'belongingness' or 'representativeness' could occur, as was possible to do in probe-tone tasks. Of interest was not key identification or belongingness of an element, but rather whether the processing of an element in a melody depended at all on its position in the tonal hierarchy, or in other words, its dynamic tone quality. The issue of position of a note in the tonal hierarchy being associated with differential presentation frequencies was controlled for, yet systematic effects based on those positions endured. Finally, one would be hard-pressed to come up with a coherent argument against the conclusions which is based on short-term memory or primacy-recency effects, considering the nature of the tasks and the stimuli employed in these studies. Specifically, the alterations occurred between tones appearing approximately in the middle of each melody, where primacy and recency ought to be at their weakest. As well, the target position did not change with changing contexts, but detection sensitivity was significantly influenced. Aside from Krumhansl's original probe-tone study, this is the first time such asymmetric perception has been systematically investigated and obtained at the level of melodies, using solely diatonic stimuli. Part of the reason for this might be due to methodological differences between this work and other research which might have attempted it but failed (another possibility might be that no one else was ever interested in it enough to bother). By keeping the alterations themselves static while manipulating the stability levels of constituent elements, reliable and predictable changes in sensitivity arose. It appears that this novel approach to the problem is a superior manipulation to those involving a reversal of the temporal ordering of the elements, since such factors as possible changes in anchoring of the target note, or differences in detectability of interval changes that arise when order is reversed, are negated with this technique. Those differences cannot be occurring, since the alterations were identical between key contexts. Detectability of an alteration, then, was seen to be a function of the particular dynamic qualities of the tones involved in the alteration, which came about from the context in which those tones were presented. Stability reductions resulted in a more noticeable alteration than stability increases. The only physical feature which varied was the particular context key; other features such as the temporal ordering, and serial position of the tones remained static. In this way, it was possible to change the dynamic qualities of the component tones unconfounded by other potential variables. This method, then, is a useful tool for further investigating the properties of the tonal hierarchy and its role in the processing of melodic information.

In particular, the locus of the alterations was restricted to the two most-stable levels of the tonal hierarchy (those levels incorporating only diatonic elements). But by using this technique, it is possible to examine the differential effects of movement between any two of the three levels of the tonal hierarchy, rather than simply examining the directional effect on sensitivity between two levels. For example, one could see if an alteration representing movement between the most and least stable levels results in greater sensitivity than when that alteration represents movement between the two more stable levels, or movement between the two least stable levels, in any direction one chooses, depending on the context key that is instantiated. Numerous comparisons are possible, with testable predicted effects on detectability, according to what we know about the tonal hierarchy from both music-theoretic descriptions, and the data obtained from probe tone tasks.

In summary, this thesis attempted to describe a more all-encompassing theoretical framework for the perception of melodies, and bring together seemingly disparate ideas concerning tonality perception, and the subsequent influence on perception that arises from gaining a sense of key. From the group-theoretic domain, there is the notion that diatonicity has been so strongly embraced in modern culture due to that set of tones possessing properties of simplicity, uniqueness, and coherence not found in any other set of tones. From the work of Krumhansl and her associates, it was seen that the tones in a diatonic sequence are conferred with the property of dynamic tone quality, wherein the role of a tone in the structural framework of the melody can vary from highly stable to highly unstable, depending on that tone's identity, and the act of key instantiation can come about as a result of frequency of presentation of certain elements from a particular diatonic scale. Butler and his associates, on the other hand, demonstrated that frequency of presentation was not necessarily the sole means by which key activation takes place, by providing evidence supporting the notion that the intervallic relations between elements in a tonal sequence can give rise to consistent and accurate key identification. The conclusion drawn here is that, depending on the task and the cues available in the stimulus, listeners may employ one, the other, or both methods in arriving at a sense of key. Following this, however, the incoming tonal information is imbued with dynamic tone quality, as depicted in representations of the tonal hierarchy, and generally described by music theorists long before these studies were carried out. These differences in structural stability, in turn, can be used to explain the asymmetric similarity perception observed in earlier work, first by regarding stability as inversely related to salience, and second by presuming that the directional nature of the paired-comparison task results in the listener's focal attention falling on the second melody of the pair. This model accounts for both the scale-structure violating stimuli of Bartlett and Dowling, as well as the scale-preserved alterations presented to listeners in the studies described in this thesis. It is this author's opinion that the influence of key and tonal structure in the perception of melodies has wrongfully entered a Dark Ages of sorts, wherein the establishment of key in the mind of the listener is assumed to have little to no influence whatsoever on said perception. The data presented here are an attempt to rectify that situation, and the model sketched out above is an attempt to provide a testable framework to further explore this seemingly neglected facet of music, wherein the dynamic features which emerge as a result of serial presentation of musical tones can and do have an influence on melodic processing.

Bibliography

Attneave, F. & Olsen, R.K. (1971). Pitch as medium: A new approach to psychophysical scaling. American Journal of Psychology, 84, 147-165.

Balzano, G.J., The pitch set as a level of description for studying musical pitch perception. In M. Clynes (Ed.), Music, Mind and Brain. Plenum: New York, 1982.

Bartlett, F.C. Remembering: A Study in Experimental and Social Psychology. New York: The MacMillan Co., 1932.

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, 501-515.

Bartlett, J.C., & Dowling, W.J. (1988). Scale structure and similarity of melodies. Music Perception, 3(3), 285-314.

Bharucha, J.J. (1984). Anchoring effects in music: The resolution of dissonance. Cognitive Psychology, 16, 485-518.

Bharucha, J.J. (1984). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. Journal of Experimental Psychology: General, 113(3), 421-425.

Bregman, A.S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology, 89(2), 244-249.

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. Music perception, 5, 219-250

Balzano, G.J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. Computer Music Journal, 4, 66-84

Browne, R. (1981). Tonal implications of the diatonic set. In Theory Only, 5, 3-21.

Carlsen, J.C. (1981) Some factors which influence melodic expectancy. Psychomusicology, 1, 12-29.

Cross, I., Howell, P., & West, R. (1983). Preferences for scale structure in melodic sequences. Journal of Experimental Psychology: Human Perception and Performance, 3, 444-460.

Brown, H. (1988). The interplay of set content and temporal context in a functional theory of tonality perception. Music Perception, 5(3), 219-250

Butler, D. & Brown, H. (1984). Tonal structure versus function: studies of the recognition of harmonic motion. Music Perception, 2(1), 6-24

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.

Castellano, M.A., Bharucha, J.J., & Krumhansl, C.L. (1984). Tonal hierarchies in the music of North India. Journal of Experimental Psychology: General, 113, 394-412.

Cook, N. (1987). The perception of large-scale tonal closure. Music Perception, 5, 197-206.

Cross, Howell, & West. (1983). Preferences for scale structure in melodic sequences. Journal of Experimental Psychology: Human Perception and Performance, 9, 444-460

Cross, I., West, R., & Howell, P. (1991). Cognitive correlates of tonality. In Howell, P, West, R., & Cross, I. (Ed.) Representing Musical Structure, AP, San Diego, CA.

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.

Deutsch, D. (1972). Octave generalization and tune recognition. Perception and Psychophysics, 11, 411-412.

Deutsch, D. The deterioration of pitch information in memory. Unpublished doctoral dissertation, University of California at San Diego, 1970.

Deutsch, D. (Ed.) The Psychology of Music. 1982, Academic Press: New York, NY.

Deutsch, D. & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. Psychological Review, 88, 503-522.

Dowling, W.J. (1978). Scale and contour: Two components of a theory of memory for melodies. Psychological Review, 85, 341-354.

Dowling, W.J. (1984). Assimilation and tonal structure: Comment on Castellano, Bharucha, and Krumhansl. Journal of Experimental Psychology: General, 113(3),417-420.

Dowling, W.J. (1991). Tonal strength and melody recognition after long and short delays. Perception and Psychophysics, 50(4), 305-313.

Dowling, W.J., & Bartlett, J.C. (1981). The importance of interval information in long-term memory for melodies. Psychomusicology, Spring.

Dowling , W.J., & Fujitani, D.S. (1971). Contour, interval, and pitch recognition in memory for melodies. Journal of the Acoustical Society of America, 49, 524-531.

Dowling W.J. & Harwood D.L. Music Cognition. New York. Academic Press. 1986.

Dowling, W.J., & Hollombe, A.W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. Perception and Psychophysics, 21, 60-64.

Edworthy, J. (1985) Interval and contour in melody processing. Music Perception, 2(3), 375-388.

Erlich, S.F., & Rayner, K. (have to dig this one up again...) Journal of Verbal Learning and Verbal Behavior, 1981, 20(6), 641-655.

Farnsworth, P.R. (1938). The pitch of a combination of tones. American Journal of Psychology, 51(3), 536-539.

Garner, W.R. (1970). Good patterns have few alternatives. American Scientist, 58, 34-42.

Gould, S.J. (1991). Bully for Brontosaurus: Reflections in Natural History, W.W. Norton & Company Inc., New York: NY.

Green, D.M., & Swets, J.A. (1966). Signal Detection Theory and Psychophysics. New York: Wiley.

Idson, W.L., & Massaro, D.W. (1978). A bidimensional model of pitch in the recognition of melodies. Perception and Psychophysics, 24, 551-565.

Janata, P. & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. Music Perception, 6(2), 161-172.

Jarvinen, T. (1995). Tonal hierarchies in jazz improvisation. Music Perception, 12(4).

Jeffries, T.B. (1974). Relationships of interval frequency count to ratings of melodic intervals. Journal of Experimental Psychology, 102, 903-905.

Jones, M.R. (1981). Music as a stimulus for psychological motion: Pt. I. Some determinants of expectancies. Psychomusicology, 1, 34-51.

Jones, M.R. (1982). Music as a stimulus for psychological motion: Pt. II. An expectancy model. Psychomusicology, 2, 1-13.

Knopoff, L, & Hutchinson, W. (1983). Entropy as a measure of style: The influence of sample length. Journal of Music Theory, 27, 75-97.

Krumhansl, C.L. (1979). The psychological representation of musical pitch in a tonal context. Cognitive Psychology, 11, 346-374.

Krumhansl, C.L. (1991). Music psychology: Tonal structures in perception and memory. Annual Review of Psychology, 42, 277-303.

Krumhansl, C.L. (1990). Tonal hierarchies and rare intervals in music cognition. Music Perception, 7(3), 309-324.

Krumhansl, C.L., & Keil, F.C. (1982) Acquisition of the hierarchy of tonal functions in music. Memory and Cognition, 10(3), 243-251.

Krumhansl, C.L. & Schmuckler, M.A. (1986). Key-finding in music: An algorithm based on pattern matching to tonal hierarchies. Paper presented at the Mathematical Psychology Meeting.

Krumhansl, C.L., & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychological Review, 89, 334-368.

Luce, R.D. (1963) Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), Handbook of Mathematical Psychology. New York: Wiley.

Macmillan, N.A. & Creelman, C.D. Detection Theory: A User's Guide. Cambridge University Press, 1991.

Meyer, L.B. (1973) Explaining music. Berkeley: University of California Press.

Miller, G.A. & Heise, G.A. (1950) The trill threshold. Journal of the Acoustical Society of America 22, 637-638.

Moore, B.C. (1982). An Introduction to the Psychology of Hearing. Academic Press: San Diego, CA.

Narmour, E. (1989) The "genetic code" of melody: Cognitive structures generated by the implication-realization model. Contemporary Music Review, 4, 45-63.

Norman, D.A. (1964). A comparison of data obtained with different false-alarm rates. Psychological Review), 71, 243-246.

Norman, D. A., and Wickelgren, W.A. (1965) Short-term recognition memory for single digits and pairs of digits. Journal of Experimental Psychology, 70, 479-89.

Oram, N (1989). The responsiveness of Western adult listeners to pitch distributional information in diatonic and nondiatonic melodic sequences. Unpublished Doctoral dissertation, Department of Psychology, Queens University, Kingston Ontario.

Parncutt, R.(1989). Harmony: A Psychoacoustical Approach. New York: Springer-Verlag.

Robinson, D.W., & Dadson, R.S. (1956). A redetermination of the equal-loudness relations for pure tones. British Journal of Applied Physics, 7, 166-181.

Rosch, E. (1975). Cognitive reference points. Cognitive Psychology, 7, 532-547.

Rosner, B.S., & Meyer, L.B. (1986) The perceptual roles of melodic process, contour, and form. Music Perception, 4(1), 1-39.

Rosner, B.S., & Narmour, E. (1992). Harmonic closure: Music theory and perception. Music Perception, 9, 383-412.

Rothkopf, E.Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. Journal of Experimental Psychology, 53, 94-101.

Schellenberg, E.G., & Trehub, S.E. (1994). Frequency ratios and the discrimination of pure tone sequences. Perception and Psychophysics, 56(4).

Schenker, H. (1979) Die Frie Satz (Free composition). (E. Oster, Trans.) Vienna: Universal Edition. (Original work published 1935).

Schmuckler, M.A. (1989) Expectation in music: Investigation of melodic and harmonic processes. Music Perception, 7(2), 109-150.

Serafine, M.L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music. Music Perception, 6, 397-430.

Shepard, R.N. (1964). Circularity in judgments of relative pitch. Journal of the Acoustical Society of America., 36, 2346-2353.

Shepard, R.N. (1982). Geometric approximations to the structure of musical pitch. Psychological Review, 89, 305-333.

Shepard, R.N. (1982b). Structural representation of musical pitch. In D. Deutsch's (Ed.) The Psychology of Music, Academic Press.

Stevens, S.S., Volkman, J., & Newman, E.B. (1937). A scale for the measurement of the psychological magnitude of pitch. Journal of the Acoustical Society of America., 8, 185-190.

Swets, J.A. (1973). The relative operating characteristic in psychology. Science, 182, 990-1000.

Swets, J.A. (1986) Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. Psychological Bulletin, 99(1), 100-117.

Swets, J.A. & Pickett, R.M. (1982). Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press: NY.

Tversky, A. (1977). Features of similarity. Psychological Review, 84(4), 327-352.

Unyk, A.M. & Carlsen, J.C. (1987) The influence of expectancy on melodic perception. Psychomusicology, 7, 3-23.

Werner, H. (1925) Uber Mikromelodik und Mikroharmonik. Zeitschrift fur Psychologie, 98, 74-89.

White, B. (1960). Recognition of distorted melodies. American Journal of Psychology, 73, 100-107.

Wish, M. (1967). A model for the perception of Morse Code-like signals. Human

Factors, 9, 529-540.

Youngblood, J.E. (1958). Style as information. Journal of Music Theory, 2, 24-35.

Zuckerkandl, V. (1956). Sound and Symbol: Music and the External World
Princeton University Press, Princeton, NJ.