Cross, I. (1989). The cognitive organisation of musical pitch. (Unpublished Doctoral thesis, City University London)

Original citation: Cross, I. (1989). The cognitive organisation of musical pitch. (Unpublished Doctoral thesis, City University London)

Permanent City Research Online URL: http://openaccess.city.ac.uk/7663/

The
Cognitive
Organisation
of
Musical
Pitch

A dissertation submitted in fulfilment of the requirements
of the degree of Doctor of Philosophy
in The City University, London

by

Ian Cross

September 1989

ABSTRACT

This thesis takes as its initial premise the idea that the rationales for the forms of pitch organisation employed within tonal music which have been adopted by music theorists have strongly affected those theorists' conceptions of music, and that it is of critical importance to music theory to investigate the potential origination of such rationales within the human sciences. Recent studies of musical pitch perception and cognition are examined, and an attempt is made to assess their capacity to provide sustainable rationales for pitch organisation in tonal music. Theoretical and experimental studies that focus on sensory processes are critically reviewed, and it is suggested that these do not adequately characterise important aspects of musical pitch organisation. Studies that examine more central cognitive constraints are discussed, and a detailed critique is made of recent cognitive-structural approaches to the representation of musical pitch. It is proposed that a significant aspect of tonal pitch organisation, diatonic structure, is neither adequately investigated nor provided with any compelling rationale by these studies. Three series of experiments on the perception and representation of diatonic structure are presented; it is suggested that the sensitivity to properties of diatonic structure shown by listeners in these and in other experiments implies that a representation of diatonic interval structure constitutes an important component of the cognitive organisation of musical pitch. A possible basis for this sensitivity is further explored, and a group-theoretic rationale for the musical use of diatonicism is proposed. The nature of the cognitive representation of diatonic interval structure is discussed, and relationships between diatonic structure, other western scale forms, tonality and (briefly) atonality are outlined.

Ian Cross, September 1989

ACKNOWLEDGEMENTS

In submitting this thesis. I wish to acknowledge the debt which I owe to a number of individuals who have assisted me in many ways. First of all, I wish to thank Natasha Spender; she introduced me to this area of research, and provided valuable and stimulating comment in the early stages of my research. Secondly, most of the experiments described in Chapters six, seven and eight were undertaken with the collaboration of Peter Howell and Robert West. They acted as my de facto "supervisors" in the psychological domain, at times in a personal as well as an intellectual sense. To Rob I owe thanks for his ability to offer succinct interpretations of some of my more deeply "nested" ideas, while to Pete in particular I owe thanks for his tendency to instruct me to "go and do it" on those occasions when theory seemed more enticing than the corralling of experimental subjects.

The exact nature of our collaboration – in terms of specifically who did what – is now a little difficult to disentangle. At one level, Pete and Rob were concerned to ensure the application of appropriate psychological method, and to ensure that any broad psychological implications of the design and results of the experiments were not overlooked, while my concern was directed more towards the elucidation of specifically musical and musico-cognitive problems. This is not to imply a simple split in responsibility for psychological and musical issues, however, for at a more detailed level I was directly responsible for the design and conduct of many of the experiments (including several not reported here which, through a process of attrition, helped to determine the forms and aims of subsequent experiments). Similarly. while I would claim responsibility for most of the ideas expounded in the course of describing the experiments, many of these were tested, or had their seeds in, discussions with Rob and Pete. Having claimed this responsibility, I am also obliged to claim sole responsibility for any errors or omissions in methodology or logic that remain.

To many others I owe thanks for comments and encouragement. In particular. though, I would like to thank Malcolm Troup, for without the sheer and outrageous liveliness that he brought to Music at City University the act of musical discovery would not have been half so enjoyable. Lastly, most thanks to Jane, for understanding.

CONTENTS

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

CHAPTER NINE

CHAPTER ONE

Pitch has generally been regarded as the dominant structural dimension of western music: the way that pitch is organised within that music is complex and idiosyncratic. As musical styles have changed it has changed, with changes in pitch organisation often constituting the principle feature of style change; however, certain constancies appear to remain even through these style changes. The constancies of pitch organisation in western music over the last four hundred years are usually referred to as the principles of tonality, exemplified in the employment of successions of triadic chords within major and minor keys to articulate musical structure. In attempting to codify the principles of tonality (whether for ontological, aesthetic or pedagogical purposes) theorists have continually sought or advanced rationales for the particular forms of pitch organisation - triads, triadic successions and keys - employed within tonal music. These rationales have often come to play significant roles in shaping ideas about what constitutes the "stuff" of music.

The importance and influence of these rationales can be seen most clearly in musical analysis. The structure of a piece of music as revealed in analysis can often be related directly to the particular rationales for pitch organisation - or even to the nature or phenomenal locus of those rationales - that an analyst chooses to employ. That is, the criteria that are used in an analysis for expressing the actual pitches of a piece of music in terms of the representation peculiar to the analytic method applied can be directly referable to the rationales for the structure of triads or keys that have been incorporated into that analytic method. Hence the claims made by analysts as to what facets of music or musical experience an analysis represents - and the controversies that may arise around these claims - can be determined at least in part by the analyst's adoption of specific rationales for pitch organisation in tonal music. The point is perhaps best made by providing a concrete example.

Currently, the most influential and widely applied method of musical analysis of western tonal music is that developed by Heinrich Schenker. It is possible to see, in the structures that Schenkerian analyses reveal, the importance of the rationale for the structure of triads and keys that is used within that analytic system.

In a recent textbook on Schenkerian analysis, Forte and Gilbert (1982, p 327-328) provide a Schenkerian reduction of the second four bars of the theme of the last movement of Beethoven's Piano Sonata in E major, Op 109 (see **Figure 1.1**). In this analysis, the terminal sequence of melodic notes in the uppermost voice, which can be read as B-A-G# (see **Figure 1.2**), is taken to imply or "stand for" the melodic progression G#-F#-E (that is, a progression of mediant, supertonic and tonic). A first impression on seeing the score - or on hearing the piece - might be that the theme does in fact terminate with B-A-G#. This impression might be strengthened by the observation that each subsequent variation of the theme tends to finish with B-A-G#. Yet the structure of the end of the theme, according to Schenkerian theory, appears to consist of the notes G#-F#-E. These notes are indeed present as an inner voice but, it might be thought, are scarcely the most salient component of the final phrase.

Figure 1.1:Beethoven Piano Sonata in E major Op 109, III (analysis after Forte and Gilbert)

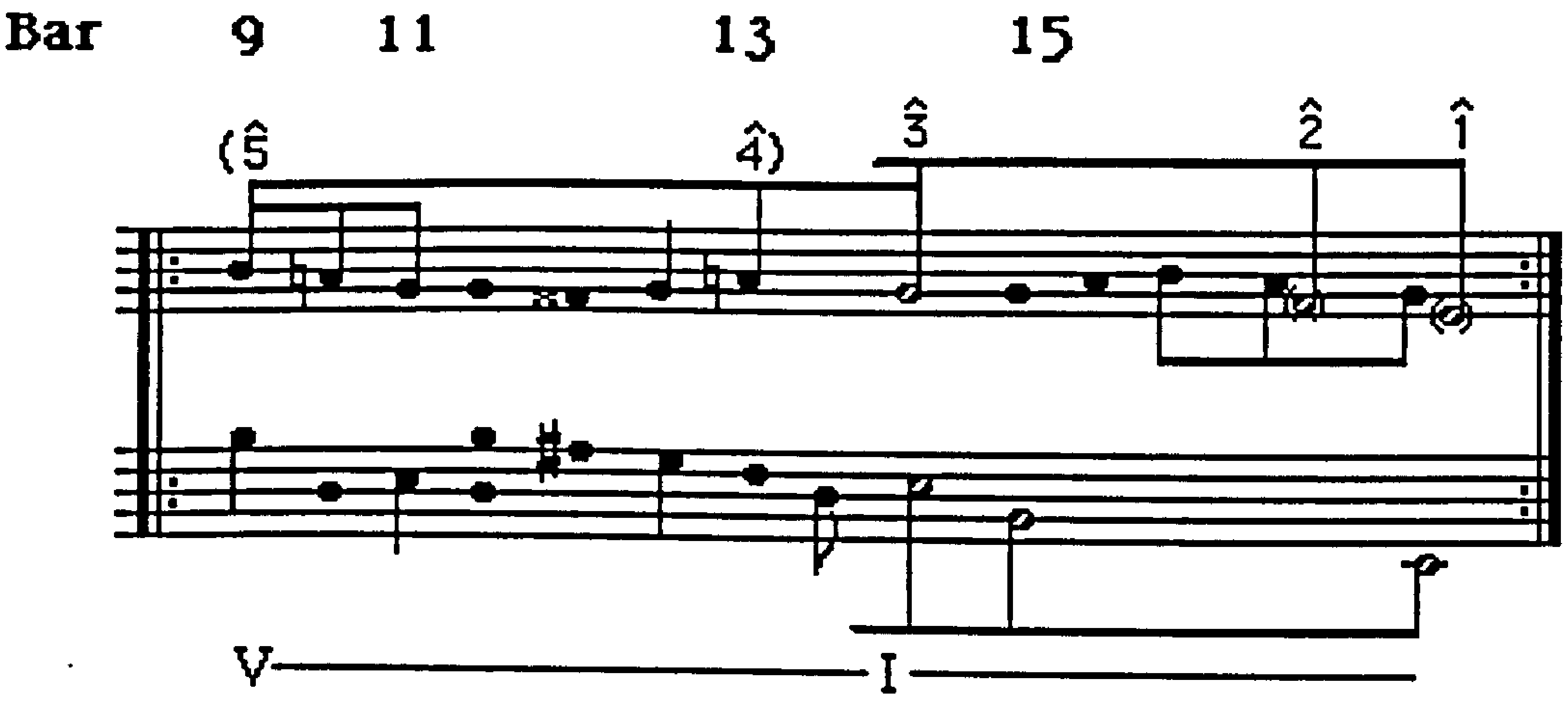
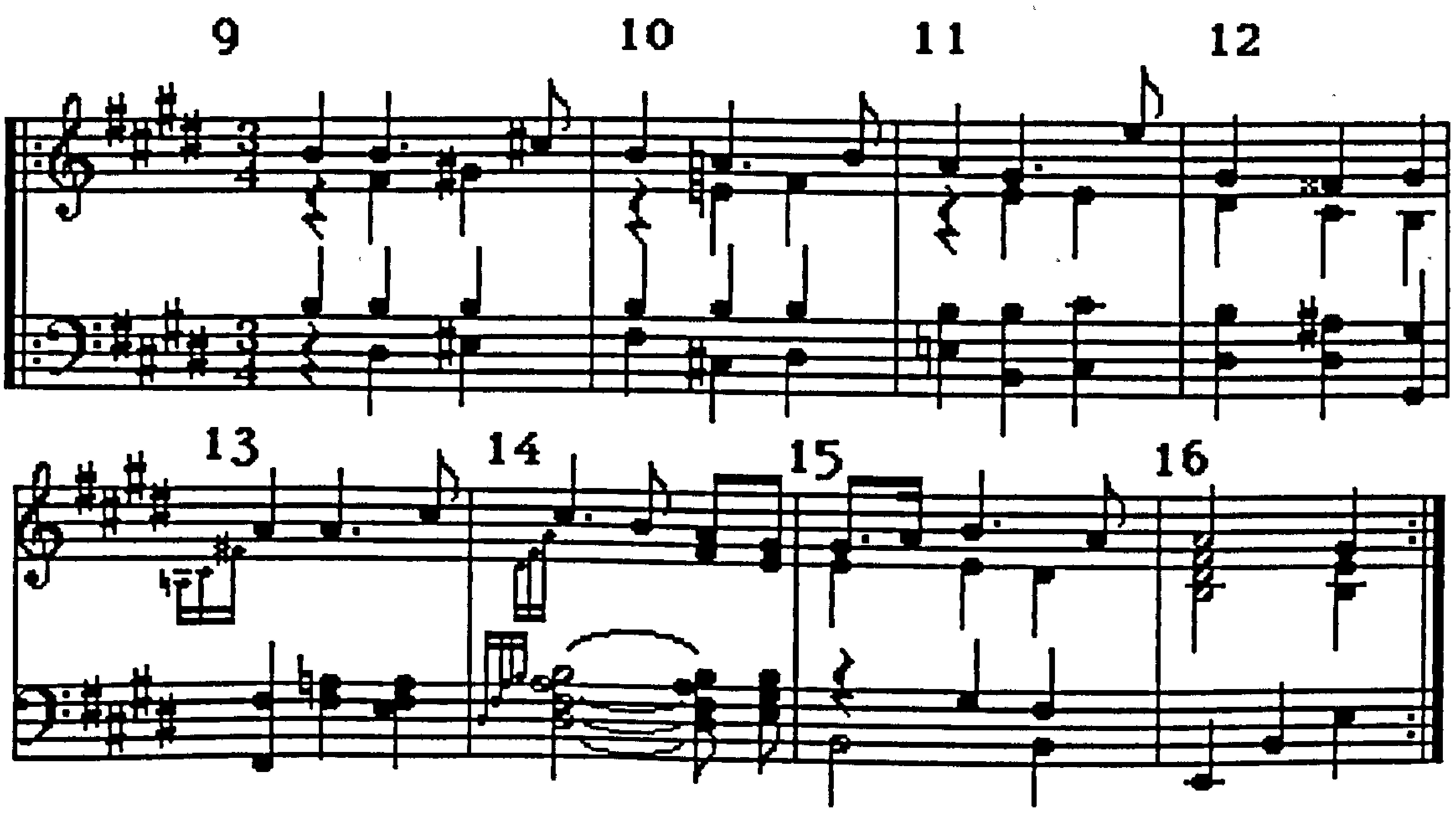


Figure 1.2:Beethoven Piano Sonata in E major Op 109, III



The reason for this analytic assertion is that in the Schenkerian view a well-formed piece of tonal music is reducible, in the limit, to a fundamental structure. the <u>ursatz</u>. This fundamental structure is intended to express the fact that such a well-formed piece will exist within a particular key (i.e. will begin and end on a tonic harmony, and make use of the members of the tonic triad as important features in melodic movement), and is constituted of two concurrent elements. a fundamental melodic line and a bass arpeggiation. The components of the bass arpeggiation always outline the harmonic progression I-V-I, while those of the fundamental line always trace a descending step-wise diatonic sequence from a primary tone (which will be a member of the tonic triad) to the tonic. Accordingly, the fundamental line in the upper voice of the above example <u>must</u> employ the notes G#-F#-E, so as to terminate on the tonic. It appears that the application of Schenkerian theory to this tonal fragment imposes rather than reveals an underlying structure: other analytic methods might happily accept that the underlying structure of the melodic line involves its termination on G#.

The problematic nature of the Schenkerian reduction of this fragment can be traced to the rationales for pitch organisation that his method adopts. Schenker (1935/79) derives the fundamental structure of a piece from an arpeggiation of its tonic triad acting within a unified contrapuntal structure of upper voice and lower voice. The tonic triad in turn is regarded as being derived from the "chord of nature", which is to be found in the relations between the first five harmonics of a complex periodic sound. The leap of the bass arpeggiation derives from a composing-out of components of the "chord of nature", while the descending nature of the fundamental line is referred by Oster (in Schenker (1935/79, p 13) to the downward pull of the tonic acting as the "fundamental" of the chord from which the structure is derived. Thus for Schenkerian theory, the fundamental structure and the materials of a tonal piece are largely determined by "natural laws"; the triad and its issue (the bass arpeggiation and the fundamental line) represent the manifestation in music of a "natural phenomenon", which is indeed shaped by art but of which the limits and potencies are prescribed by nature (see Schenker, 1906/54)[1].

The fact that a periodic sound embodies several harmonic components yet <u>appears</u> to a listener as a unity (i.e. has the identity <u>in perception</u> of its fundamental component or <u>root</u>) is used by Schenker to account for the departure from and return to the tonic and for the centrality of the tonic triad with respect to a key that are generally evident in pieces of tonal music. So in the Schenkerian view the structure of the "chord of nature" as well as its unitary tendencies determine the structure and order of events within a well-formed piece of tonal music. Such a piece must start with and return to the tonic triad: the bass must outline a I-V-I progression, unfolding the principal elements of the "chord of nature": and the fundamental line must descend to the tonic to emphasise the primacy of that note as the root of the "chord of nature" of which the piece is a manifestation.

---

[1] The structure of scales and of keys is also referred directly to "nature", being predicated by Schenker on the occurrence in the overtone system of the "perfect fifth".

Of course this description of Schenkerian theory is simplistic. Many other factors such as rules of counterpoint, consideration of the nature of repetitions, "laws" of organic unity etc. play a role in Schenkerian theory. Nevertheless, the rationale adopted for pitch organisation within Schenkerian theory acts strongly to determine revealed analytic structure, and hence, the Schenkerian conception of what constitutes music.

In seeking to base the structure of triads and keys on the harmonics of a sounding note, Schenker implicitly appears to adopt a Lockean position of "direct realism" (see Gregory, 1981), wherein certain characteristics of objects are known directly[2], a position that is nowadays largely untenable (at least, in the auditory domain). It would seem that he holds that direct awareness of the physical characteristics of a complex tone acts to guide the perceptions of listeners and to constrain the sensibilities of composers. In a sense, the rationale that he adopts can be said to be predicated on the acoustical characteristics of a complex tone but to have a potency that derives from an assumed human sensitivity to such acoustical phenomena. In part, then, the way in which Schenkerian theory can appear to impose rather than to reveal structure can be regarded as determined by the immutability of the laws governing the acoustical phenomena that subserve its rationales for pitch organisation.

---

[2] Despite a cursory mention (see Schenker, 1906/54, page 25) of the possible mediating role of auditory physiology in the perceptual process, direct awareness of the laws that Schenker imputes to the "chord of nature" is simply assumed.

Schenkerian theory has been used here to demonstrate how the rationales for pitch organisation in tonal music that are employed within analytic methods can affect conceptions of music. Other examples could be given; the ideas of musical structure found in the theories of Hindemith (1945), for instance, can be regarded as almost wholly determined by the rationales for tonal pitch organisation that he adopts.

The selection of appropriate rationales for tonal pitch organisation is evidently critical for western music theory. The rationales used must enable theory to apply to a wide range of musical phenomena and musical intuitions, or at least enable the the theory to possess some coherent relation to actual music and to a broader theoretical context. It would seem unlikely that such rationales can be based on the operation of laws in the purely physical and acoustical domain; there is, after all, no necessity for a direct correlation between physical laws and our sensitivities to, or perceptions of, the operation of these principles. It would be more appropriate to seek rationales for musical pitch organisation within the principles which shape human sensitivities to the natural world, that is, which determine our perceptions and understandings. In other words, if rationales for pitch organisation in music theory are to have any "scientific" basis, then they must be rooted in human rather than in physical science.

The body of this thesis will examine recent studies of musical pitch perception and cognition, and will attempt to assess their capacity to provide sustainable rationales for musical pitch organisation. Theoretical and experimental studies that address psychoacoustics and the cognition of music will be discussed, and it will be proposed that a significant aspect of tonal pitch organisation, diatonic structure, is neither adequately investigated nor provided with any compelling rationale by these studies. Three series of experiments[3] on aspects of perception of diatonic structure will be presented, and it will be suggested that the sensitivity to properties of diatonic structure shown by listeners in these experiments implies a group-theoretic rationale for the musical use of diatonicism.

---

[3] The experiments described in Chapter 6 were published in Cross, West and Howell in Journal of Experimental Psychology: Human Perception and Performance,((3),444-460, 1983; those in Chapter 7 were published in Howell, West and Cross, Journal of the Acoustical Society of America, 76(6), 1682-1689, 1984; those in Chapter 8 are under revision for publication in Music Perception for 1990, while certain of the ideas presented in Chapter 9 appeared in print in the first chapter (Cross, "Music and Change...") of Musical Structure and Cognition, (Eds) Howell, Cross and West, Academic Press 1985.

CHAPTER TWO

The perception of pitch

Given that acoustical factors do not offer a sustainable rationale for musical pitch organisation and usage, it might reasonably be expected that a basis for musical relations in the pitch domain could arise from constraints imposed by the processes whereby acoustic signals are transduced into neural signals subserving the sensation or perception of pitch. Before examining three theories of musical pitch organisation that are based on this premise it is necessary to present a brief sketch of relevant aspects of current theories of pitch perception.

In the simplest case, the pitch of a single sinusoidal tone is more-or-less equatable with its frequency. The pitch of a complex tone containing a set of harmonically-related frequencies can be taken as corresponding to the pitch evoked by a sinusoid at the lowest or fundamental frequency. This led, in the nineteenth century, to the view that pitch is mediated primarily by place of excitation on the basilar membrane. However, in the case of a complex harmonic tone the amplitude of the fundamental may be reduced to very much less than that of one or more of the higher harmonics, or the fundamental may even be removed completely, yet the pitch percept will remain more-or-less constant. As Moore (1982, p 141) states, "the perception of a particular pitch does not depend on a high level of activity at a particular place on the basilar membrane or in a particular group of peripheral neurons...the pitch of a complex tone is mediated by harmonics higher than the fundamental, so similar pitches may arise from quite different distributions of neural activity".

Pitch perception is undoubtedly mediated by the mechanisms whereby frequencies are discriminated in the auditory periphery, but various different mechanisms may be proposed for its extraction from the complex pattern of events within the cochlea. As is implicit in the previous quote, Moore (1982) and a number of other psychoacousticians such as Patterson (1986) emphasise the role of neural periodicities in pitch perception. Others such as Terhardt (1977) have advanced the view that pitch perception is mediated primarily by place-type mechanisms, with particular patterns of excitation on the basilar membrane coming to be associated with particular pitch percepts. There is, however, a general consensus that the mechanisms for pitch perception at high frequencies (above ca. 2.5 Khz) are likely to be at least partly mediated by place of excitation on the basilar membrane, given that the period of such signals exceeds the latency of auditory nerves. Discontinuities in judgments of pitch relations which are attributable to a possible cross-over from a periodicity to a place-based region of pitch perception have been reported by, e.g., Attneave and Olson (1971). Notwithstanding this controversy, current models of pitch perception can be thought of as largely two-component; the first component represents the transduction of acoustical information by the mechano-neural "hard-wiring" of the auditory periphery, while the second component represents the operation of some more central, higher-level cognitive processes on the transduced neural information.

Patterson's spiral detection of periodicity

Patterson (1986) uses the properties of his model of pitch perception to explain the structure of musical scales and triads. The principal concern of his model is to address the fact that the auditory nerve seems to convey information about both spectral and temporal aspects of sound waves; accordingly, his theory outlines a means of transforming temporal information in the acoustic waveform into some time-independent neural code. He characterises the peripheral auditory system as two-part, comprising analog and digital subsystems. The analog, mechanical, subsystem transduces acoustic information into neural impulses while performing a spectral analysis; the digital subsystem organises these into neural pulse streams, more-or-less phase-locked to the movement of the basilar membrane. Patterson proposes a spiral model for the detection and representation of periodicity - or pitch - information from these pulse streams.

Patterson's spiral model is intended to be a functional psychological pitch model rather than an accurate physiological model. It provides a graphical and computational account of the way in which periodicity information can be extracted from a constantly changing flow of neural inter-pulse time intervals. Its operation is explained by taking a plot of neural pulses (arising from the neural response to periodic acoustical input) against time, and wrapping the pulse-train "time-line" round in a logarithmic spiral with the base 2 (of which the length will double with each successive circuit). The most recent instant in time occurs at the centre of the spiral, and the pulses flow outwards along the path of the spiral as time progresses. Pulses of which the intervening durations are related as powers of two will thus "line-up" as "spokes" on the spiral at some particular point in time: given that the time interval between adjacent pulses is constant for a periodic input, at some particular time (once per cycle of the sound) the first, second, fourth, eighth etc. pulses line up on one spoke of the spiral while the third, sixth, twelfth etc. line up on a different spoke. Waveforms with different time-periods (frequencies) produce the same spoke patterns but with different orientations. Spoke activity is monitored by secondary sets of neurons, which detect and identify the presence of pulse-trains of differing frequency.

The spiral processor detects time-doublings rather than evenly-spaced events. It functions as a "common-mantissa" detector (to the base 2); coalescence of a set of pulses onto a spoke indicates that the logarithms of these pulse times have a common mantissa (fractional part of a logarithm) at a particular point in time. It thus provides a fast and computationally efficient way of automatically transforming a temporal flow of inter-pulse information into a time-independent form. That is, it provides a means of automatically representing the temporal detail of a sound (i.e. its period) within a time-independent domain - frequency or pitch - enabling the depiction of complex relations.

The spiral processor and the organisation of musical pitch

According to Patterson a (periodic) musical sound would generate a specific pattern of spokes on the spiral. The orientation of the spoke pattern would determine the frequency of the sound (here equated with its pitch), which would be detected and identified by the monitor units lying around the spiral.

Moreover, notes of which the frequencies are related by low-integer ratios will have spoke patterns that are orientated such that one or more spokes coincide. In fact, frequencies related by simple ratios between integers from one to five will always have at least one prominent coincident spoke.

Patterson notes that intervals which have been considered as "consonant" in traditional music-theoretic usage - i.e. of which the fundamental frequencies form low-integer ratios - will have such coincident spoke patterns, commenting (p 58) that "in the spiral model the combined spoke pattern produced by a pair of notes appears on the processor at the earliest stage [i.e. before any process that indicates the components present in an input spectrum] and overlapping spoke patterns that indicate consonance are immediately apparent". That is, the spiral model provides a low-level, automatic and immediate means of identifying musical consonances.

Patterson refers to Hindemith in arguing that "the rule of consonance provides a logical and elegant description of the basic notes of the two primary scales of Western music", further amplifying his argument by pointing out (p 58) that such "consonant" intervals were mentioned by Rameau as underpinning the Just Diatonic scale. He describes the major Just diatonic scale as consisting of the notes and ratios C(1/1), E(5/4), F(4/3), G(3/2), A(5/3) while the minor comprises C(1/1), Eb(3/5), F(2/3), G(3/4), Ab(4/5). The scales are "filled out" by adding the notes D(9/8) and B(15/8) to the major, and the notes D(9/8) and Bb(8/9) to the minor. Thus (p 57) "the primary intervals of the major and minor scales are those that produce spoke coincidences". In addition, spoke patterns produced by the major and minor triads will appear as "mirror-images" of one another. The diatonic scale - in the Just form presented here - and the triads are preferred (p 60) because of the relatively simple patterns of spoke activity that they produce.

Patterson's rationale for the organisation of musical pitch is open to objections on several levels: it is controversial both as a theory of pitch perception and as a justification for the forms of musical scales. Among objections to it as a theory of pitch perception is the fact that although the common-mantissa detector is a fast and computationally efficient way of transforming temporal information into a time-independent domain, it achieves this at the cost of assuming a simplicity of acoustical input signal that is incommensurate with many real-world "periodic" waveforms. That is, it relies on a consistency of neural inter-pulse interval that may simply not be derivable from waveforms which can be adjudged to have a pitch (see, e.g., de Boer, 1976). This could be overcome by building-in some tolerance in the secondary monitor units that detect spoke activity; however, Patterson does not explicitly allow for this in his model. Moreover, the model is entirely bottom-up; that is, it appears to allow no role for adjustment of its operating parameters on the basis of prior information or experience. In this, it conflicts with other theories of pitch perception (such as that of Terhardt, Stoll and Seewann, 1982, or of Scrulowicz and Goldstein, 1983) which employ both peripheral _and_ central mechanisms in the determination of pitch. That is, the spiral detector doesn't seem to allow any role for the _contextual_ identification of pitch (which is explicitly allowed for in other models such as that of Terhardt et. al., as will be seen); pitch, in Patterson's theory, is determined automatically, by more-or-less hard-wired processes acting only on the input signal.

The theory is also problematic as a justification for the forms taken by musical scales. Although Patterson's reliance on low-integer ratios in the formation of the Just diatonic scales seems to have support in that Rameau also appealed to this quarter in his theories, it is ahistorical in that scales tuned in Just temperament are almost entirely abstract constructs, with no evidence of historical usage. As Lloyd (Lloyd and Boyle, 1963) points out (p 147) "the seven-note scale of just temperament forms perfectly-tuned diatonic intervals with the tonic and gives perfectly 'true' intervals for the major triads on the tonic, subdominant and dominant. But there its practical usefulness ends. So limited is its musical objective that it cannot be used for real music, as the Tudor musicians well knew". According to Lindley (1984), the gamut (the practical range of available pitches for the late Medieval modal system) and its later Renaissance extensions were most likely to have existed in real music in the forms of Pythagorean tunings (vocal music), quarter- and third-comma mean-tone tunings (keyboards) and even - from the mid-sixteenth century - twelve-note equal temperament (fretted/stringed instruments). Moreover, the scale segments presented by Patterson as Just diatonic major and minor do not accord with practical or theoretical schemes for classifying pitch relationships prevalent during the period (ca. 1400-1600) when many of the "rules" of tonality could be said to have been being formed.

Even if the scale forms that Patterson relates to his system had had some historical use, the fact that relations between their primary elements strongly correlate with simple patterns on the spiral detector does not explain their relevance to musical syntactic structure. Although the model does provide a basis for different chord-types to be conceived of as having different degrees of stability (by virtue of their relative consonance, or relative complexity of their spoke patterns), it does not provide a compelling rationale for melodic pitch relations or for the apparent centrality of cadence forms in defining tonal structures.

Patterson's model relies on extrapolation of information about exact adherence to low-integer frequency ratios; this implies that many of the same objections as were advanced in the case of purely acoustical theories of pitch organisation can be advanced against this model. The fact that a variety of tuning systems which do not adhere to low integer ratios have existed in western tonal music, as well as the fact of the eventual dominance of the equal-tempered system - in which the only exact low integer ratio is to be found at the octave - implies that whatever the basis for harmonic and scalar structure in western music it is not to be found in such this type of model of pitch perception.

There are, however, other recent attempts to account for musical pitch organisation in terms of pitch-perception processes which appear to overcome many of the above objections. In particular, that developed by Terhardt, Stoll and Seewann (1982) seems to have a wider applicability and a greater explanatory power than does Patterson's model. It is claimed by its authors to provide a coherent rationale for musical concepts such as consonance and dissonance and the "roots" of chords (referred by Terhardt to Rameau's concepts rather than to those of Tartini or Hindemith) within the framework of a multi-stage model of the perception of pitch.

Terhardt's model of pitch perception

Terhardt's theory is based on the fact that it is possible to hear the same complex waveform as having different pitches depending on how it is listened to. For Terhardt, the pitch assigned to a complex waveform is called virtual pitch. The virtual pitch may correspond directly to the pitch evoked by one of the elements of the input spectrum or it may not. That is, a candidate for virtual pitch may derive directly from one of the components of the acoustic waveform - corresponding to a spectral pitch - or it may derive from some subharmonic of the spectral components which is not itself physically present in the acoustic waveform. In his model, pitch is thought of as existing at two different levels, spectral pitch and virtual pitch. The spectral pitches of a signal are those pitch percepts directly associable with input spectral components, while the virtual pitch is derivable from the spectral pitch pattern. In the earlier version of this theory (1977, 1978), Terhardt seems to treat the formation of virtual pitch as relatively deterministic - that is, for a given complex waveform, one virtual pitch is expected to be much more prominent in perception than others, and hence will serve as the pitch of the whole complex.

A later, more formal version of the theory is presented in

Terhardt, Stoll and Seewann's (1982) algorithmic approach to the perception of pitch. They intend this to be a formal, computable model of the abstract processes involved in the perception of pitch, stages of which may or may not be directly associable with discrete neurophysiological structures. The algorithm has five stages. Stage 1, frequency analysis, takes the form of an FFT with constant bandwidth of ca. 12.5 Hz, giving a power spectrum of the input waveform. Stage 2, extraction of harmonic components, proceeds by extracting all FFT components which are +7 dB greater than any adjacent six components. 7 dB seems to be taken as an empirically derived salience figure, representing the minimum difference for clear discriminability within the 150 Hz (six component) bandwidth. Stage 3, evaluation of masking effect, operates on the results of stage 2 - now regarded as more-or-less discrete frequencies - to evaluate the degree to which neighbouring harmonic components (and the "noise" component of the original FFT) act to suppress or swamp each other, as well as determining the degree of "pitch shift" of the harmonic components (deriving from contrast-enhancing mechanisms and from component sound pressure level, and, according to Terhardt (1978) the phenomenon likely to underlie the seemingly universal (Burns, 1974: Dowling, 1978) preference for "stretched" octaves).

These first three stages can be regarded as constrained by general, cross-modal neural functioning, as they are directly analogous to the processes proposed by Marr (1982) for deriving the grey-level image and primal sketch in the construction of visual images. The result of stages 1 to 3 is a set of spectral pitches (each of which might be expected to be "heard out" by an experienced listener).

The algorithm continues to Stage 4, weighting of components, which applies a weighting function - based on the absolute frequency range within which spectral dominance occurs (up to about 2 KHz, see Ritsma, 1967) - to the spectral pitches resulting from stages 1-3. This results in the spectral pitch pattern representing the relative perceptual salience of competing simultaneous spectral pitches. At Stage 5, extraction of virtual pitch, "subharmonics" (analogous to integer divisors of frequencies corresponding to the spectral pitches) are calculated, giving potential virtual pitches. These are assigned weights which depend on i) the number of spectral pitches which provide the same subharmonic pitch, ii) the spectral pitch weight assigned in stage 4 and iii) the "subharmonic number", analogous to the frequency corresponding to the subharmonic pitch (virtual pitch weight decreases with increasing subharmonic number).

This extraction of the range and weights of potential virtual pitches leads to their ranking in terms of probability of assignation to the stimulus. That is, the result of the algorithm is a set of pitches ranked in order of the likelihood of their being perceived as the pitch of the complex input signal: in general, the virtual pitch with the highest probability weighting (which may or may not correspond to a spectral pitch) is taken as the pitch of the perceived complex tone. However, this model allows for two possible modes of pitch perception. In the holistic or synthetic mode, virtual pitch is most likely to be perceived. In the analytic mode, spectral pitch - perhaps even spectral pitches, i.e. a chord - will be perceived. The operation of either mode is largely determined by context, or by factors not explicitly addressed in the algorithm. These include (among others) shape of overall amplitude envelope, coherence of frequency shifts among groups of harmonic components and the general "set" of the listener. Note, however, that the assignation of virtual pitch is not now particularly deterministic, but rather probabilistic.

Terhardt's model and music

Terhardt (1978) notes that "careful analysis of psychoacoustical transformations [such as those represented by his model, where an acoustical signal is transformed into an auditory sensation] may provide valuable theoretical insights into such basic musical phenomena as e.g. consonance and harmony". He points out that the perceived "roughness" of a complex tone - or of two or more superimposed complex tones - derives from envelope fluctuations caused by beating of adjacent partials within critical bandwidths. The lack of such roughness - arising from the components of two or more tones having small integer ratios between their frequencies - could thus be termed "sensory consonance". He qualifies this by pointing out that "other important features of musical sounds such as tonal affinity and harmony are not explained by this relation...thus musical consonance (in a general sense) may be composed of two main components, sensory consonance and musical consonance". He then attempts to account for aspects of musical consonance and the "harmonic identity" of chords in terms of his theory of pitch perception.

Terhardt uses his theory to propose that chords are experienced as having roots. He suggests (1978) that "virtual pitches...will be produced when musical chords are presented [to a listener]...every part-tone (of the chord) which may be considered as being resolved by the peripheral auditory system (i.e., not masked by adjacent ones) is considered as a cue to virtual pitch" resulting in the obtaining of several virtual pitch labels; some of these will correspond to the fundamental frequencies of tones present in the chord, while some will be lower, "subharmonic" virtual pitches. In the case of major chords, the virtual pitches obtained will correspond to "those assigned by conventional music theory [i.e. following the roots assigned by Rameau] to that major triad". In general, he proposes that the auditory system inspects the dominant spectral components of the perceived chord with respect to the question "which harmonics of what fundamental might these be ?" the chord being perceived as harmonic and as possessing a typical root if there are sufficiently unambiguous answers while if the answers are too ambiguous the chord is "disharmonic", having no root. He provides experimental evidence to show that subjects could identify short melodies outlined by the "fundamental-notes" (virtual pitch roots) of short sequences of chords (Terhardt, 1977).

Terhardt, Stoll and Seewann (1982) provide a rather less deterministic account of the assignation of roots to chords. In this later version of the theory, the degree to which one virtual pitch candidate will be more salient in perception than all the others varies according to the nature of the input waveform. Thus, e.g., although both major and minor chords will be perceived as having "roots", the unambiguous assignation of virtual pitch roots to major chords might be easier than for minor chords. Terhardt et al (1982) present an analysis of six triadic chords (major and minor in all inversions), illustrating the weights that the algorithm assigns to various virtual pitch "roots"; so, e.g., the chord (A4:C#5:E5) is shown to be most likely to have the root A4 with weighting (w) equal to 1.41, the second most likely would be A3 (w=1.09) and then A2 (w=0.59).

In making an equation between the pitches of chord-notes and the harmonics of a fundamental, Terhardt's theory seems capable of providing a rationale for musical pitch organisation that could supplement or supplant that used within Schenkerian and similar theories (although no explicit attempt to do this is made in any of the papers). To extrapolate, the nature of chordal relationships might be referred to the degree to which one chord exhibits different possible "root-candidates", or to the degree to which different chords share the same "root-candidates" with different weights; for example, the fourth most likely root of the chord (A4:C#5:E5) is given as D3, while the fifth most likely is E6, from which one might conceive of a tonic-subdominant-dominant hierarchy as being "latent" in the perception of a "root-position" major triad. Similarly, the "stability" of triadic chords might be referred to their relatively unambiguous virtual-pitch "roots".

If used within the framework of a fully-developed theory of tonal music, Terhardt's theory might permit chordal function to alter in context through its provision of multiple potential roots. Similarly, the idea of spectral pitches might provide a basis for the perception of note-to-note relations in a series of chords, and the theory might be able to characterise polyphonic musical textures. In addition, Terhardt's theory avoids the problem of rigid dependence on low-integer ratios between chordal constituents for its explanatory power by the incorporation of mechanisms for the detection of "pitch shift" in Stage 3 of the model.

However, none of these music-theoretic extrapolations are advanced in any of the papers. Terhardt et al make no attempt to provide a systematic account of the implications of the model for music theory. No attempt is made to relate the predictions of the model either to principles of musical syntax or to the forms of music-theoretic pitch structures such as keys or scales. The theory is presented as though it simply enabled the labelling of different chord-types (rather after the manner of Hindemith) within a static framework, consonance and dissonance - despite Terhardt's (1978) explicit acknowledgement of their contextual nature - being treated as intrinsic properties of particular chord-types rather than as properties that may derive from the contexts within which those chord-types are used. Questions such as "why should a particular chord or set of chords imply or define a particular key ?" are not addressed. Chords are assumed to be homophonic entities - the possibility that, in certain musical usages, they might be better characterised as the results of voice-leading processes is not considered.

It is possible that these extrapolations remain unexplored in part because of their unfalsifiable nature and in part because of their ethnocentric bias. As no <u>explicit</u> perceptual mechanism is proposed in the model that could account for the syntactic logic of cadence forms, the possibility that these forms may be dependent on low-level processes of auditory perception remains untestable. In addition, the statement that triadic homophony is intimately related to general listening processes is of little value in accounting for pitch organisation in musical cultures (such as that of Java, India, Japan etc.) which make little or no use of triadic homophony but are relatable to western music in their use of octave-equivalence, discrete pitch steps and scale structures.

Moreover, the experimental evidence provided by Terhardt (1977) for the perception of chord "roots" is less than compelling. Only well-known melodies were used, and their rhythmic profiles were maintained. The melodies themselves were present as voices of the chord sequences presented (though admittedly, the melody-notes did not consistently appear in the same voice in consecutive chords). Given these conditions, it is possible to advance plausible alternative explanations for the identification performance of his subjects; as the melodies' rhythmic profiles were maintained, subjects may have used this as a cue to melodic identity, given the relative strength of this parameter as a cue in other melody-identification tasks (as, for example, in Deutsch, 1972 or in Dowling, 1973). As the melody was actually present in the pitches of the chords, subjects may have been able to infer the identity of the melody by picking out individual voices of the chords. Acting together, both sets of cues would seem quite sufficient to explain Terhardt's results without the necessity of adducing a "virtual-pitch extraction-and-tracking" strategy on the part of the subjects.

Terhardt's theory differs from that of Rameau (to which he compares it) in providing an explicit account in terms of general processes of auditory perception of how "overtonal" relations might be significant for the cognitive organisation of musical pitch; as Terhardt, Stoll and Seewann (1982) demonstrate, the predictions of their algorithm are in close accord with current empirical data on pitch perception. However, their theory ultimately undermines the notion of the inversional equivalence of chords - one of the central tenets of Rameau's musical theory - as the results that they quote make clear that the assignation of "roots" to different chord inversion or to major and minor triads on the same tonic is likely to be highly indeterminate. For example, the second-inversion major chord (E5:A5:C#6) would have only a 0.60 weighting for its most prominent root A3, while the next most prominent root would be E5 with w=0.54, rather too close to the weight for A3 for completely unambiguous assignation of chord root. Worse yet, the two pitches E4 and A3, representing the highest weighted roots for the chord (E5:A5:C6) have the same weighting (w=0.54). Thus two inversionally-equivalent chords such as (C5:E5:A5) and (E5:A5:C6) might be assigned quite different roots (respectively, A4 and either E4 or A4) - a possibility that accords rather better with C. P. E. Bach's (1787/1961) and also Schenker's (1935/79) view of harmonic function and equivalence than with that of Rameau.

Parncutt and "pitch commonality"

Many of the objections to Terhardt's theory seem to be overcome in the revision and extension of his algorithmic model proposed by Parncutt (1986, 1988). Parncutt (1986) suggests that in dealing with musical chords, stages 1 to 3 of the Terhardt model should be simplified slightly; this seems to involve (i) adopting a higher frequency resolution (narrower bandwidth), (ii) calculating the "auditory" level of each component by subtracting its "ideal" threshold SPL from its observed SPL and (iii) calculating each component's audible level by subtracting the overall "masking level" at the position of the component from its auditory level. The output of these processes provides a set of what Parncutt refers to as "pitch-classes" (more properly, pitches labelled categorically according to note-name and pitch height). These pitch-classes serve as input to three further stages: (i) each pitch-class of a chord is assigned a weight (which refers to its calculated perceptual salience); (ii) the pitch weights are used to estimate the "root ambiguity" of the chord and (iii) the calculated root ambiguity value is used to convert pitch weights into absolute estimates of salience for each significant pitch-class component (all of which are octave-generalised over central musical pitch-range).

Parncutt uses an extension of Terhardt's theory to estimate the weights of pitch-classes within a chord so that, for example, octave-related spectral pitch components falling within the spectral dominance region will interact at the fourth stage of Parncutt's model to produce greater weights for lower components than would have been the case if no octave relation were present. These weights are then used to calculate the root ambiguity of the chord, giving a single value that expresses the number of possible different chord-roots. These two sets of values are then used to calculate the salience - or probability of being noticed - of each pitch class associable with the chord.

Parncutt (1988) claims that his theory provides more appropriate estimates of likely roots of, for instance, minor chords than does that of Terhardt. Whereas, according to Parncutt (1988, p72) the latter's algorithm would be likely to calculate the note F to be the most probable root of a chord containing the notes (C:Eb:G), with three times the weight of the next two most probable roots, Eb and C, Parncutt's model calculates the most likely root as C (with weight 1.60), followed by Eb (weight 1.33) and G (with weight 1.00). Parncutt's result certainly does seem more plausible than Terhardt's, according to most music-theoretic criteria.

Parncutt (1986) proceeds to develop his model by explicitly proposing that it can be used to provide a rationale not only for the holistic perception of single chords but also for the nature of harmonic syntactic structure, by means of the concepts of pitch proximity and pitch commonality. He suggests that musical sounds such as the chords in a chord progression are perceived to be more or less related to each other to the degree that they have pitches in common or have roots which are close to each other in pitch, or both. He defines pitch commonality as the degree to which two simultaneities evoke tone sensations in the same pitch-class and pitch proximity as the overall perceived closeness in pitch of the sounds (closeness here being defined as either semitonal proximity or proximity within the circle-of-fifths), and produces formal versions of these which can be applied to the output of the earlier algorithm.

He suggests that these concepts - commonality and proximity - can account for the ways in which different chords are heard as similar or different, and describes an experiment which tested his predictions in this respect. He used four groups of subjects (western musicians, non-western musicians, western non-musicians and children), who were required to judge the similarity of pairs of close-position major triad chords. The similarity of each chord-pair was also calculated by means of the algorithm. Responses for all groups appeared to indicate that (a) chords of which the voices "led well" according to music theory, or (b) which had a note in common, or (c) of which the roots were close on the cycle of fifths were adjudged most similar, with the influence of pitch commonality being greatest for the western musician group. By-and-large, these results appeared to concur with those predicted by his algorithm, allowing different weightings for the effects of commonality (result (b)) and of proximity (result (c)), although result (a) cannot be accounted for in this way. Nevertheless, Parncutt's theories seem to capture some highly significant features of musical perception, and provide a rationale for some important music-theoretic relations.

However, several important types of harmonic relationship - perhaps the most music-theoretically significant types - appear to remain outwith the scope of Parncutt's theory. For instance, its characterisation of chordal syntactic relations in terms of pitch commonality and proximity does <u>not</u> appear to provide any way of capturing the relations between the chords of a cadence, nor does it enable relations between chords which are not temporally adjacent to be described coherently. As mentioned, it also provides no rationale for the effects of voice-leadings on adjudged chordal similarity. Moreover, Parncutt does not attempt to derive rationales for particular music-theoretic pitch schemes - such as interval types, or scale-forms - from aspects of his theory. Indeed he invokes the idea of listeners' acquiring "pitch and interval categories" through "cultural exposure" without in any way indicating how such learning processes and their products might relate to the "sensory" processes that his model seeks to describe. His theory would appear to be severely limited in its capacity to account for anything more than aspects of more-or-less isolated pairs of pitch simultaneities.

Objections to psychoacoustical accounts

In general, it seems that a major problem with attempts to found theories of musical pitch on models of auditory perception that focus on low-level transductional processes is that they provide accounts of the organisation of musical pitch in terms of features likely to be applicable only to a limited domain of musical pitch organisation. They may provide accurate accounts of how harmonic structure and chordal identity arise in musical perception, but they limit the musical percept to that dimension; they provide one-dimensional images of musical pitch. As Parncutt (1986) explicitly admits, his algorithm can only account for certain aspects of completely homophonic music. Moreover, the applicability of Parncutt's theories of pitch commonality and proximity to the perception of linear relations appears to be limited to relations between temporally adjacent chords. There appears to be no way in which these concepts can account for harmonic relations within a hierarchical framework.

Moreover, these approaches do not provide compelling accounts of how the linear, melodic and syntactic domain might be perceived. It is difficult - if not downright inappropriate - to account for the organisation of monophony and polyphony in purely harmonic terms. As Dahlhaus (1980a) comments "[the] theory that the harmony is always a resultant rather than a starting point can still remain reasonable as a requirement to be met by musical perception...It will mean that the listener is expected to give his attention primarily to the 'movement features' of the individual parts".

The largely data-driven nature of the psychoacoustical theories discussed here means that although they may account for the "forms" that musical entities take in perception, these "forms" tend to be considered in isolation and provide of themselves little or no basis for perceived musical function. The theories are algorithmic in that they specify sets of procedures which enable the transformation of physical signals into psychological representations that seem to be appropriate models of isolated musical elements. However, the theories are not complete "computational theories" of musical perception (to adopt the terms used by Marr, 1982) because although they show how the physical signal may be transformed into some mental representations commensurate with the elements of music, they do not show how important types of relations between those elements might be represented; they provide no generalisable account of the potential musical <u>functions</u> - and hence the larger-scale musical forms, relations and structural entities - that their inputs might subserve in perception as a musical work unfolds in time. The psychological representations arising from the bottom-up processes that they specify may serve as clues to some of the constituent elements of music-as-perceived, but these psychological representations do not of themselves provide sufficient clues as to how these elements might interact or function with respect to one another in a musical context.

Pitch categorisation

A further problem arises in that despite claims to the contrary (as in Parncutt, 1986), these psychoacoustical models tend to rely on low-integer ratios, or at least on "best-fits" to the harmonic series of the discriminable components of complex tones or chords in their explanations of the cognitive organisation of musical pitch. The fact that such low-integer ratios do not seem especially significant in pitch discrimination tasks would seem to indicate that the psychoacoustical theories offer at best partial solutions to the question of how pitch might be organised in cognition. Several studies have demonstrated that listeners tend to discriminate between musical intervals in both melodic (Burns and Ward, 1978) and harmonic (Halpern and Zatorre, 1979) contexts on the basis of musical interval category rather than on the basis of degree of approximation to "ideal" frequency ratios. This is in spite of the fact that listeners can exhibit much finer pitch discrimination ability; Houtsma's (1968) estimates of just-noticeable differences between frequency ratios are in the order of 13 to 26 cents (between about one-eighth and one-quarter of a semitone).

As Burns and Ward (1982) point out (p 245-254), the fact that subjects show evidence of categorical perception in an experimental task implies a) that the subjects are able to discriminate two stimuli only to the extent which they can identify them, and b) that subjects have access to some verbal or symbolic code in terms of which the stimuli can be represented. The subjects in their 1978 experiments were both musicians and non-musicians; only the musicians - who might be expected to have access to some verbal or symbolic code whereby musical intervals can be identified and represented - consistently showed "sharp and reliable" category boundaries between musical intervals. Non-musicians showed no evidence of categorical perception, either based on low-integer ratios or on musical interval-types. However, Siegel and Siegel (1977) do report some evidence for the categorical perception of musical intervals by non-musicians, although this appears to derive from a range effect rather than reflecting over-learned interval categories. The conclusions that Burns and Ward (1982) draw - from their own and others' studies of perception <u>and</u> performance - are that although natural (low-integer ratio-based) intervals have probably influenced the development of Western musical scales, intervals are perceived categorically in all other than minimal-uncertainty conditions, and the "the standards of intonation for a given [musical] culture are the learned interval categories of the scales of that culture". They conclude that even octave-equivalence - often held to be not only innate, but to have cross-species generality (Blackwell and Schlosberg, 1943) - is probably learned, a conclusion supported by Sergeant's (1983) study.

In the light of the previous general criticisms and of much evidence for categorical perception of musical intervals, it would seem that psychoacoustical theories are unlikely to be able to provide complete rationales for the ways in which pitch may be organised in music. None of these theories offers any coherent and substantiated account of the perception of monophony or polyphony, nor for the particular interval classes, or the particular scale formations, that appear in western music. This appears to arise in part from their data-driven nature; these theories derive what is perceived almost exclusively from momentary sensory input, and appear largely unable to account for musical relations as they may unfold in time. Their focus on sensory-transductive processes seems to provide little or no basis for elucidating how particular pitch configurations might come to subserve specific musical functions. If the nature of music-theoretic relations in the pitch domain is indeed to be referred to the nature of our experience of music rather than to some more abstract and extrinsic constraints, it would appear necessary to examine broader principles and processes which underlie perception of - and memory for - more general classes of events and their inter-relations in time.

As Sanford (1985, p 69) points out, "it makes a real difference to models of word recognition if we remember that we normally need to recognise words in the <u>continuous reading of text</u>" (his emphasis). The analogy is plain; musical events such as chords are normally encountered as constituents of musical passages rather than in isolation. In order to understand the ways that music might be identified and remembered, it is thus necessary to examine those studies which have been directed towards exploring the principles underlying general classes of pitch relations between heard events. These principles can be thought of as falling into two domains: general principles of auditory pattern perception, and principles that are apparently specific to the perception of musical pitch relations (although the latter might also be construed as a specialised subset of the former). It seems that these general principles are bound up with the need to make sense of what we hear, in terms of locating, identifying, remembering and recognising the characteristic sounds and noises of objects and events in the environment. These types of activity are obviously dependent on sensory processes and their results; however, they cannot be entirely "data-driven", implying as they do a reliance on prior experience and knowledge. It is likely that consideration of the ways in which such knowledge is acquired, represented and applied in perception will provide means of accounting for musical pitch organisation which have more general applicability and greater explanatory power than do those deriving from purely psychoacoustical approaches.

CHAPTER THREE

Sound sources and auditory streams

Moore (1982, p 185) states that "the auditory world is analysed into discrete sound sources, each of which may have its own pitch, timbre, location and loudness". While the sensory transduction of individual and multiple attributes (frequency, intensity, modulation rate etc.) of sound sources can be shown to influence such analyses, it would appear that general principles founded in likelihood of applicability to a wide range of real-world sources provide a more coherent basis for analysis. Sound sources must not only be located but also identified, and their location and identity must be tracked over time. Cues associated with these variables are likely to be multiple; sources might be differentiated and identified on the basis of their loudness, pitch, timbre or location. Although perception of any or all of these attributes can be construed as the result of sensory (bottom-up) mechanisms, their perception as a source is tracked over time seems to be mediated by more central and constructive cognitive processes[1].

[1] Some negative evidence that can be held to support this is reported by Bregman (1978), in an account of an experiment following-up Bregman and Campbell's (1971) finding that listeners can misidentify one sound source which consists of a sequence of pitches alternating between two discrete ranges as being two separate sources. This might be explained as arising from the operation of some bottom-up, peripheral tracking mechanism of the sort proposed by Luce and Green (1978), whereby sequential proximate frequencies would sensitise (lower the firing threshold of) proximate or overlapping groups of cochlear neurons; if this were the case, then some discontinuity in the source-misattribution effect would be expected (arising from the possible cross-over from a periodicity to a place-based method of pitch perception). However, Bregman & Bernstein (in Bregman, 1978) tested the generality of Bregman and Campbell's effect over a very wide frequency range and reported that no such discontinuity existed; accordingly, it may be inferred that some higher-level processes were involved in the source misattribution.

Bregman describes the percepts that arise when sources are tracked over time as auditory streams. He describes the principles underlying their formation as arising from rules that relate to "ideal" forms and relations, stating (1981, p 109) that "[the application of] these ideal relationships [in perception] are 'bets' about the world". In a sense, this is to give these principles a similar status to that which Marr (1982) accords to uniqueness and continuity in stereopsis; they are principles of ecological likelihood. Bregman likens the auditory stream (source tracked over time) to an "auditory object", and states that an "ideal" stream is one that has homogeneous features, such as continuation over time at a relatively constant pitch, any changes being smooth rather than sudden, with simultaneously-occurring components changing in parallel. The description of an auditory environment is thus built up from descriptions of separate streams and their interactions in terms of "ideal" forms and relations; listeners use features of "ideal" streams to parse their sound world.

As principles underlying the perception of "ideal forms and relations", Bregman employs the Gestalt "laws" of proximity, similarity, good continuation, closure and common fate. He refers the findings of Bregman and Campbell's (1971) experiment, in which a sequence of alternating high and low tones was perceived as constituted of two separate streams (high and low), to the operation of the similarity principle; as alternate notes were identical, the material was grouped in perception on the basis of pitch similarity rather than on the basis of temporal order. The findings of Divenyi and Hirsh (1978) expand the possible operation of this principle, bringing in the Gestalt concepts of "figure" and "ground"; they showed that the saliency of a three-note pattern which was embedded in a longer sequence was dependent on the degree to which the three-note pattern was construed as "figural" as opposed to the "background" formed by the sequence within which it was embedded. This "figure-ground" distinction was easiest for listeners to make when the respective pitch ranges of the "figure" and "ground" note-sequences did not overlap, and when the figure occurred at a "temporal edge" of the longer sequence (i.e. at its beginning or end).

The good continuation principle is taken to underlie Bregman and Dannenbring's (1973) results, in which the tendency of such a sequence of alternating high and low tones to split into two streams was reduced by connecting sequential tones by frequency glides or glissandi; the fact that movement between temporally contiguous pitches was now smooth and regular assisted listeners to hear the sequence as one rather than as two streams. In a similar way, Bregman holds that the principle of closure determines the illusion that a tone "interrupted" by a noise burst actually continues through the noise (Bregman and Dannenbring, 1977), although in fact the tone has an offset and onset that are synchronous with the onset and offset of the noise. As the tone and noise "share" perceptual edges the noise is perceived as "occluding" or masking the tone; the tone is perceived as continuing under the noise, and the (real) gap in the tone is perceptually "closed". The principle of common fate is shown to be in operation in Bregman and Pinker (1978), in which the degree of onset synchrony of two potential components of a complex tone was shown to be directly related to their likelihood of fusing as a single percept; that is, the fact that two elements of the experimental sequences underwent the same changes at the same times and in the same way made these elements more likely to be perceived as part of the same event.

In all of these experiments - but perhaps most particularly in the last - these principles are to some extent in competition with one another. The pitch, timbre, loudness and location of the sound material are capable of being arranged so that the organisation of each parameter provides some possibility of streaming the material on the basis of that parameter. That is, relations between possible "ideal" streams are being experimentally manipulated in attempts to determine not only what principles might be operative but also the relative strengths or weights of the different principles in the different perceptual parameters to which they might apply. Deutsch's (1975a) results can be taken as indicating that the operation of these principles in the pitch domain can tend to override other streaming cues. She found (1975a) that presenting two-note chords drawn from simultaneously ascending and descending major scale formations arranged so that high and low chord components were fed dichotically to alternate ears resulted in the perception of streams organised by pitch proximity rather than by location. Butler (1979) extended these findings, showing that the effect of streaming by pitch proximity was highly robust by replicating the effect via loudspeakers rather than headphones, introducing differential amplitude cues and also timbral cues; both of these potential cues to stream the sounds by spatial location were overridden by the tendency to stream by pitch proximity.

However, this "dominance" of pitch as a means of stream organisation is not absolute; Deutsch found (1981) that presenting simultaneous high and low notes (octave-related) dichotically through headphones to alternate ears resulted in streams being formed on the basis of apparent spatial position, though alternate notes at each ear were perceptually "suppressed" in the formation of these streams. Moreover, Erickson (1982) reports that in the performance of repeating cyclic melodies by five different instruments each playing one melody note in turn, listeners reported either streaming based on the melody itself, or streaming resulting from proximate (but temporally non-adjacent) pitches or streaming on the basis of individual instrumental timbres: the basis for streaming could be switched to some extent at will. It would appear that the "preferred" parameters within which auditory streams are adjudged to occur is somewhat dependent on the experimental method used to elicit reports of these streams, and is to some extent under the conscious control of the listener. Notwithstanding these caveats, pitch appears to provide a powerful means of identifying and tracking sound sources as auditory streams. Many further experiments (e.g. Divenyi and Hirsh, 1974; Bregman and Rudnicky, 1975; Dannenbring and Bregman, 1976) all indicate that pitch organisation provides very strong cues to the streaming or grouping of auditory events.

Given the strong relations that hold between pitch organisation and stream formation, it would appear that the principles underlying auditory stream formation can be thought of as underlying the perception of global aspects of melodic structure. Most western tonal melodies conform to the conditions for stream perception - they occupy limited pitch ranges and tend to make use of conjunct rather than disjunct melodic motion (i.e. tend to change pitch gradually rather than abruptly). Moreover, prescriptions for the composition of western polyphony (musical texture consisting of multiple simultaneous melodies) seem to be directed towards the production of multiple simultaneous streams that will be more-or-less separate in perception. For instance, among other prescriptions, the rules of Fuxian counterpoint state that the preferred movement of each polyphonic voice should be by step rather than by leap (thus assisting the perception of each voice as a stream or melody); they also imply that two or more parts changing in pitch simultaneously should not do so in the same direction and by the same pitch interval, thus favouring the conditions that Bregman (1981, p 113) indicates will maintain their perceptual separateness or segregation (Wright and Bregman, 1987).

The principles that underlie streaming explain to some extent why we hear melodies as melodies rather than as series of consecutive events, and also explain the perception of some relatively local connections between temporally discontiguous notes. However, our representations of melody are framed not only in terms of these general streaming principles but also in terms of further general global aspects of melodic structure. If melodies are "octave-scrambled" by maintaining the <u>chromas</u> or note-names of their pitches (i.e. maintaining that attribute shared by two notes an octave apart), but randomly switching the <u>pitch height</u> (the dimension of pitch directly correlatable with frequency) at which the notes occur, this should reduce the likelihood of a listener being able to perceive them as melodies by the application of streaming principles. It might, in fact, lead to the perception of several streams, each constituted of temporally discontiguous notes which occupy discrete and narrow pitch ranges. However, Dowling and Hollombe (1977) found that listeners can still recognise a melody under these conditions when the contour or pattern of ups-and-downs in the pitch domain is preserved.

Contour

Perception of aspects of musical structure which are explicable in terms of melodic contour seems to have been first experimentally investigated by Ortmann (1926) and later by Werner (1940), in his study of the perception of melodies based on "micro-scales". More recent investigations, however, have indicated that listeners appear able to identify and recognise melodies on the basis of contoural information, or patterns of ups-and-downs. The principles whereby melodic contour may be perceived can be thought of as similar to those underlying the perception of auditory streams; however, in the light of Dowling and Hollombe's (1977) findings, contour might be better thought of as a "figural" property of a melody or phrase. Divenyi and Hirsh (1974) found that ascending or descending three-note sequences were easier to identify than were sequences which changed melodic direction; they related this in part to the simplicity of the melodic "figure" or contour of the fragments, as well as to the operation of the (stream-forming) principle of good continuation.

The perception of melodic contour does not require assimilation in cognition of relations between consecutive pitches in any other than a nominal sense (see Jones, 1978); that is, a contoural representation takes into account the direction but not the size of melodic steps between notes. Dowling and Fujitani (1971) found that listeners were largely unable to distinguish between transpositions of five-note fragments that retained contour although altering the exact intervals of the original sequence and transpositions which were exact; however, listeners could easily distinguish between transpositions that altered the contour and those that did not. Dowling and Fujitani also showed that listeners could could recognise distorted versions of familiar tunes in which pitch intervals between notes were changed but contours were preserved.

Edworthy (1985) tested the degree to which memory for contoural information differed from memory of information about the actual pitch intervals used in melodies. She found that contoural information seemed to be important only in the short-term; when melodies were increased in length (or, as was shown by Dowling and Bartlett (1981), if subjects were required to make recognition judgments over relatively long time-spans of ca. 5 minutes) contoural information was of less use than was information about the actual pitches and pitch relations in a melody in making accurate judgments. This seems to indicate that listeners are more likely to represent melodic phrase similarities in contoural terms when they fall within the span of immediate apprehension (see Lindsay and Norman, 1977) or within the boundaries of working memory (see Hitch, 1980) rather than when they occur widely separated in time, and when they are unable to coherently represent the pitch relationships within a melody in any other form, points made by Dowling (1982a) in his review of the area.

However, the perception of melodic relations is not mediated solely by the global perceptual principles governing stream formation or the identification and recognition of contoural identity. Relations between and within groups of notes can be represented in finer detail and in more structured ways in cognition than are available within a stream- or in contour-based representation. For example, Kallman and Massaro (1979) showed that listeners were unable to recognise octave-scrambled melodies on the basis of contour preservation alone; when they altered the relative pitches or chromas of the notes of the melodies in the octave-scrambled versions, listeners were unable to recognise the melodies although they were able to recognise the melodies so long the as the octave-scrambled versions used the same pitches as in the normal melody. That is, when _pitch chroma_ was altered in the octave-scrambled versions, recognition performance diminished even though the contours of the melodies (in terms of _nominal_ pattern of ups and downs) remained constant. In fact, contour cannot really be manipulated without alteration of these finer details of pitch relation, as Davies (1979) points out.

Interval, scale and key

It is obvious that judgments can be made about melodies on the basis of the particular notes that they contain, and the specific relations between these notes. In the light of Dowling and Hollombe's (1977) findings and those of Kallman and Massaro

(1979), these relations may be perceived in terms of pitch height or of pitch chroma or both. Although listeners can exhibit very fine discrimination in judging pitches (as Houtsma (1968) shows), the previously-cited studies by Burns and Ward (1982) would appear to indicate that these fine discriminatory abilities may not be directly relevant to musical judgmental tasks; it would appear that the perception of musical pitch may be fairly coarse-grained. Further evidence from Deutsch (e.g. 1970, 1975b: summarised in Deutsch, 1982) seems to indicate that memory for musical pitch is in general very poor.

Deutsch (1970) showed that although the pitch of an isolated tone can be recognised after a (fairly long) silent time-span of 15 seconds, the interpolation of other tones between the tone to be remembered (standard) and the comparison tone drastically reduced listeners' ability to retain the first tone in memory. Further studies revealed that the inclusion in the interpolated sequence of a tone that was a semitone removed from the standard tone disrupted memory even further, and that an even larger disruption occurred if standard and comparison tones differed and a tone that was the same as the comparison tone was included in the interpolated sequence. Deutsch (1975b) generalised this finding to show that maximum memory disruption occurred when one of the interpolated tones differed from but was close in pitch to one of the test tones, and that this maximal disruption occurred when tones differed by two-thirds of a tone; she accounts for these results (Deutsch, 1982) by referring to the operation of some lateral inhibitory network in pitch memory.

These findings seem to indicate a rather poor memory for pitch in the short-term. However, Deutsch's studies are primarily orientated towards uncovering factors affecting memory for isolated pitches; the degree to which listeners could place the material to be remembered in some sort of context that would enable more accurate retention is not explored. Studies of this type appear to measure ability to discriminate between, or to remember, stimulus material which is represented solely in terms of "sensory traces"; it has long been recognised that stimulus material may be represented in short-term memory in some transformed or encoded form. It could be that Deutsch's studies have produced results that do not correlate with listeners' abilities to perceive and remember either isolated tones or relations between tones when some sort of contextual coding of the material can be carried out.

Deutsch herself in an earlier paper (1969) suggests that a possible basis for the contextual encoding of musical pitch might take the form of a network representing musical intervals (although she is not specific about how such a representation might arise), a proposal that seems to be supported by experiments demonstrating categorical perception of musical intervals. As Burns and Ward (1982, p 264) point out, the use of the relatively small number of available musical intervals to encode pitches and their relations could be determined by "inherent limitations on the processing of high information-load stimuli by human sensory systems" leading to the categorical perception and representation of intervals in all other than minimal-uncertainty conditions. The distinction between pitch height and pitch chroma already referred to can be considered as reflecting these limitations. The number of differentiable pitches and pitch relations (intervals) can be drastically reduced from that available within the pitch height domain alone if fundamental frequencies are taken as differentiable not only in the pitch height domain but also in terms of pitch chroma or note-name - that is, if pitches which are octave-related are considered as perceptually equivalent in some way.

However, many studies indicate that the cognitive representation of pitch relations is not simply expressible in terms of musical intervals, but is rather better accounted for in terms of the forms in which notes and intervals manifest themselves when constituting elements of scales or keys; that is, a major consideration in the ability to discriminate between, identify and remember relations between pitches and note patterns seems to be the degree to which they can be interpreted as notes within musical scales or keys. This sensitivity to scale or key may arise through repeated exposure to tonal music, involving a process of assimilation of the structural and functional regularities that such music embodies. The "knowledge" encapsulated in such an "overlearned representation" would function as a form of mental model (see Johnson-Laird, 1983), incorporating in some rule-structure the general principles of pitch organisation in the corpus of music to which the listener had been exposed. In the process of perception, scale or key may thus be thought of as functioning as a form of schema. To adopt Neisser's (1976, p 54) definition, a schema can be thought of as "that portion of the entire perceptual cycle which is internal to the perceiver, modifiable by experience and somehow specific to what is being perceived...the schema accepts information as it becomes available at perceptual surfaces and is changed by that information: it directs exploratory activities that make more information available, by which it is further modified".

Evidence for the importance of key or scale organisation in perception is shown in the study by Dewar, Cuddy and Mewhort (1977). They asked listeners to recognise short melodic sequences in exact, contour-preserving and different-contour transpositions. They found that their subjects could distinguish correct repetitions of seven-note sequences from incorrect (one note altered) repetitions significantly better with tonal sequences (i.e. fitting into a single scale or key) than with atonal sequences, and also that listeners could recognise single notes drawn from a preceding sequence more easily if that sequence was tonal than if it was atonal. Cuddy, Cohen and Miller (1979) went on to show that the ability of listeners to detect the alteration of one note in a three-note fragment was affected positively by the addition of two-note suffixes and prefixes which "set" the fragment in a strongly tonal context, while the provision of atonal suffixes and prefixes actually depressed performance.

Further evidence of the importance of tonal relations in perception is found in the results reported by Dowling (1978), as well as some indication of the nature of listeners' representations of those relations. In a task replicating that of Dowling and Fujitani (1971) he asked two groups of listeners - musically experienced and musically inexperienced - to discriminate between five-note sequences which were either exact transpositions, "tonal" transpositions (which simply shifted the original so that it occurred at a different point in the same scale as the original, changing the exact interval sizes but retaining the same number of scale-steps between consecutive notes), contour-preserving atonal versions of the original and different-contour tonal five-note melodies. Dowling found that while listeners could distinguish between transpositions that preserved contour and different-contour melodies, they were largely unable to distinguish between exact and "tonal" transpositions although they could distinguish between transpositions that were either exact or tonal and transpositions that resulted in atonal melodies. Both of the groups that he tested exhibited this pattern of responses, although the more experienced listeners showed a higher level of performance. It would appear that his subjects were unable to discriminate between intervals that, although comprised of different numbers of semitones, constituted the same number of functional melodic-intervallic steps when melodic patterns could fit into an underlying scale or key.

Bartlett and Dowling (1980) further investigated Dowling's (1978) findings, using a similar recognition task. Listeners had to distinguish between the same types of transpositions as in the 1978 study, but transposed into different scales or keys from the original. Bartlett and Dowling found that "tonal" transpositions into keys closely related on the circle of fifths to the key of the original were more difficult to distinguish from exact transpositions than were "tonal" transpositions into keys more distantly related. They related this result to the fact that nearly-related keys share more notes with the key of the original than did distantly-related keys, and hence the "tonal" transpositions into nearly-related keys were intrinsically more confusable with exact tranpositions than were transpositions into distant keys.

Dowling (Dowling, 1982b) supports the hypothesis that scales and keys are gradually assimilated as "mental models", fulfilling the role of schemata in the perception of pitch relations. He suggests that listeners with little exposure to music are likely to remember melodies in terms of interval sizes between consecutive notes, and only with more experience - i.e. more exposure to tonal music - will they come to represent melodies in memory in terms of chromas within a given scale or key, citing the results of a further experiment in support of this claim. In this experiment, listeners (with either little or moderate experience of music) were required to differentiate between exact and "tonal" transpositions of brief melodies preceded by four-chord harmonic contexts. Dowling found that inexperienced listeners were actually better than his "moderately experienced" subjects at distinguishing exact from "tonal" transpositions, and concludes that abilities of the more experienced listeners are characterisable by the degree to which they have abstracted some "tonal scale" representation from the music that they have heard; for these listeners, melodies are more likely to be integrated into some "tonal scale schema" than for the less experienced listeners, and hence the melodies' constituent intervals would become less easily recoverable.

Scale, key, mode: music theory

So far, evidence has been adduced to indicate that some cognitive representation of pitch relations in terms of scales or keys plays a major role in the ability to discriminate between, identify and remember relations between pitches and note patterns. The questions of what constitutes scale or key and of how these might be represented in cognition in some schematic form has not yet been addressed. Music-theoretic definitions of these (and related) terms will be presented before consideration of various attempts to characterise the forms that they might take in cognition.

In traditional music theory (see ABRSM, 1958), a scale is "an alphabetic succession of sounds ascending or descending from a starting note", characterised by a particular distribution of tones and semitones; the diatonic major scale will have a distribution of intervals (in semitones) between adjacent alphabetically-ordered notes of 2, 2, 1, 2, 2, 2, 1. A key is constituted of "the set of notes on which a piece [of music] is built, each note having a definite relation to a note known as the key-note or tonic". Thus the set of notes D, B, F, C, A, G, E constitute the set of notes available within the key of C major; re-ordered as C, D, E, F, G, A, B, (C) it can be seen that they conform to the interval structure of the diatonic major scale, with each note forming a particular interval with either the lower or the upper note C, the keynote or tonic.

Twelve enharmonically-different diatonic major keys can be formed within the total available chromatic set of notes;

differentiating between enharmonically-equivalent notes, nineteen diatonic major keys can typically be formed (although more are available, and are occasionally implied within late Romantic music). Scales and hence keys can be major or minor, the minor scale existing in two forms (harmonic and melodic), although the minor key is taken as one single entity having variable internal pitch relations. With the addition of the idea that triadically grouped notes (e.g. C-E-G) form some privileged or stable grouping, and the provision of some hierarchy of importance of notes other than the tonic within a key, these definitions more-or-less suffice in the elementary teaching of the typical forms of pitch relations within tonal music.

However, definitions at this level are of limited value; they only describe the materials that are likely to be used in tonal music, and these only partially. Moreover, they do not indicate how the concepts scale and key might be related to "real" music in that they do not provide a functional theory of these materials. If this level of definition is used in the interpretation of the experimental findings outlined above then these findings are likely to have little or no explanatory power for music theory. It is necessary to undertake a more detailed review of the role and significance of these concepts and their functions in music theory in order that their relation to the experimental results may be made more explicit.

As is evident from the description given, scale and key represent two different types of pitch organisation. The conventional description of scale appears to indicate that it is a piece of specific "passage-work", although implicit in that description is the idea that it is the intervallic structure between scale-notes that defines the scale. Key, on the other hand, is defined not only by adherence to a specific scale but by the maintenance of a set of specific relationships between the tonic of the key and all other notes; that is, not only adherence to the intervallic structure of the scale is necessary, but some differentiation in terms of importance must be made between the tonic and other notes of the key as well as between these other notes. It would appear, then, that scale is a simpler and more abstract level of description of pitch relations than is key, incorporating no real differentiation between scale-notes.

Key, on the other hand, is a more complex musicological phenomenon; different theories of key can be advanced. These vary as to whether relations between different scale-degrees or relations between different triads are regarded as the principal features of a key. That is, the tonic could be considered important as a scale-degree or as the root-note of the central triad within the key. Both viewpoints are equally tenable, and neither need exclude the other. In general, the principal elements of key are taken to be the relations between triads separated by the interval of a fifth as ordered within "typical" tonal musical phrases - that is, as the chords I-IV-V-I - and, to a lesser extent, the "substitutive" relations between triads on scale-degrees separated by a third, such as II and IV (i.e. in the sequence I-IV-V-I, II could substitute for IV to produce I-II-V-I). In music theory, triadic chords within a key can be regarded as both "fused" entities having roots which may be sequentially ordered to form "fundamental progressions" involving steps of a fifth (hence I-IV(II)-V-I, following Rameau), and as collections of notes comprising (between them) all the notes of the scale that underlies the key (following Riemann, see Dahlhaus, 1980b). Therefore each triadic chord within a key is related to - and is differentiable from - the tonic and the tonic triad of the key by means of its root or by virtue of the notes that it incorporates. Each triadic chord may be supplemented by additional notes (as in the chord of the dominant seventh) or it may employ altered notes (as, for instance, in the case of the minor chord on the subdominant); in this way, the various non-diatonic notes can be integrated (however loosely) into an "extended key".

The relations between notes or triads in a key take the form of orderings of the "degree of stability" of the various scale-degrees, or of the various triadic chords which may be formed within a given key. That is, the definition of a key embodies expectations about the degree to which its elements will act as stable events in a piece of music, the more stable events tending to occur at the beginnings or endings of phrases; the significance or function of other elements will largely be determined by their structural or temporal relations to the stable reference element(s).

However, these levels of definition of scale and key seem to be addressed towards a level of complexity of pitch organisation which goes considerably beyond that which is addressed by the experiments described in the preceding section. Moreover, it has been argued that the definition(s) of key which have been given are too specific, and may be applied only to a highly restricted sub-corpus of "tonal" music. For example, Palisca (1980) points out that a description of key that is expressed in terms of the theory of "fundamental progressions" *might be best construed* as a theory applicable only to Western art-music of the late 17th and early 18th centuries. It would seem desirable to attempt to express the concepts of scale and key within some simpler but more broadly applicable framework; examples of these have been suggested by musicologists who are concerned to develop and refine music-theoretic tools that might be equally appropriate to Western tonal music and the music of earlier periods as well as to non-Western music. A particularly cogent specimen of this type of definition is found in Powers' (1980) account of a "scale-mode-melody" continuum.

Powers (p 377) states that "if we think of scale and melody as representing the poles of a continuum of melodic predetermination, then most of the area between can be designated ... as being in the domain of mode. To attribute mode to a musical item implies some hierarchy of pitch relationships and some restriction on pitch successions; it is more than merely a scale". In this definition, mode is seen as a concept embracing aspects of pitch organisation such as key - at least, insofar as key is manifested in melody - and is differentiated clearly from scale. For Powers, scale is the alphabet of pitches and their inter-relations (intervals) which constitute a piece of music (or part of a piece). Scale - as distinct from mode - does not embody any "hierarchy of pitch relationships"; it is a neutral level of organisation insofar as it enables pitches to be differentiated one from another with no implication that any one pitch is more important than another. Within this definition, reference to "the diatonic major scale on C" would not imply that the note C has any particular primacy, but would simply indicate a particular collection of notes. Mode, on the other hand, is defined by Powers as either a "particularised scale" or a "generalised melody"; mode embodies some pitch hierarchy, and is likely to be identifiable both by the particular pitches that occur and by the order in which they occur in a piece of music; at the same time "a mode is always at least a melody type or melody model, never just a fixed melody". The medieval systems of differentiating between pieces in the chant repertoire on the basis of their underlying intervallic structure and the terminal note of the pieces and their subphrases may thus be subsumed under the term mode, as may the North Indian system of differentiating between rags on the basis of their intervallic structure and permissible melodic formulae.

In this way, Powers seeks to present a general means of classifying melodic pitch organisation in terms of which the materials of a wide range of different types of music can be ordered and discussed. This sort of systematisation would seem to provide a level of discourse more appropriate to consideration of the cognitive organisation of musical pitch - and to the elucidation of any rationales for musical pitch organisation that this might offer - than might the detailed, culture-specific framework provided by conventional music theory. Indeed, Dowling (1978) proposes a "conceptual scheme" for pitch organisation in cognition that has several features in common with Powers' approach.

Scale, key, mode: cognitive representations

Dowling's scheme has four levels, the psychophysical pitch function, the tonal material, the tuning system and mode; he claims that there is good evidence - based on experimental work, and on common applicability across various cultures - that these "levels of analysis" have some psychological reality. The psychophysical pitch function is "the general rule system by which pitch intervals are related to frequency intervals of tones", which takes an approximately logarithmic form (following Attneave and Olson, 1971). The next level, tonal material, comprises "the set of pitch intervals in use by a particular culture or within a particular genre", taking the form in Western culture of the intervals available within the equal-tempered chromatic scale. The third level, tuning system, consists of "a subset of the available pitch intervals from the tonal material that are used actual melodies, but not anchored to any specific frequencies" (in Western culture, major or minor scales). The last level, mode, is characterised by the "translation of what were the pitch intervals at more abstract levels [i.e. of the pitch intervals constituting an available tuning system] into the pitches of specific notes", together with the "establishment of a tonal hierarchy within the tuning system".

As can be seen, these last two levels are relatable to Powers' scale-mode distinction. However, it would appear that Dowling is using the concept "mode" in a rather more restricted way than is Powers, whose definition is both more particular and more broad-ranging. Powers' concept of mode is more dynamic than is Dowling's, in that for him, modal organisation is likely to be made manifest by the syntactic (temporal) as well as the alphabetic (structural) organisation of melodies, whereas Dowling is not specific in making this distinction between scale (or tuning system) and modal levels of analysis or representation. Indeed, Dowling is not specific about the role that these different levels might play in musical cognition, nor does he indicate how these representations might arise. He does (after Helmholtz) give an account of the necessity of discrete pitch steps together with some form of scale-type representation (to enable listeners to judge the degree of melodic and rhythmic movement within a melody). However, he does not proceed to demonstrate either the necessity of or the detailed functioning of his levels of analysis in the perception of music.

Cuddy (Cuddy, Cohen and Mewhort, 1981: Cuddy, 1982) presents a pragmatic representational scheme which seems to derive more directly from the concepts of scale and key as they appear in traditional music theory than does that of Dowling; however, in her scheme the different levels of analysis are also less fully explained. She states (1982) that "understanding the detection of pattern [in melody] requires us to explore ... the role of repetition, contour, intervals contained within the melodies and implied harmonic progression ... we must direct research towards the analysis of temporal structure". She adduces two dimensions (in addition to contour) within which melodies might be ordered in terms of the degree to which they exhibit characteristics of <u>tonality</u> (defined as being dependent on scale membership and harmonic syntax); these are harmonic structure and "excursion" (i.e. whether or not a melody returns to its starting note).

Cuddy suggests that melodies embodying the highest level of structure (i.e. that in which the "rules of tonality" are most closely adhered to) can be derived from one single diatonic scale (Dowling's tuning system), imply an underlying tonic-dominant-tonic chord progression (as in Schenker's theory) and end by moving from the leading-note to the tonic. They are also likely to start on the tonic as well as ending on it (zero excursion). Lower levels of melodic structure are produced by gradual relaxation of these rules; so the next lower level will end on the tonic but the penultimate note will not be the leading-note, the next lower level will relax the rule of chord progression while still maintaining the key and the next lower level allows some violation of key membership, still lower levels allowing increasing violation of key membership.

Cuddy, Cohen and Mewhort (1981) tested the validity of this system of classification by asking listeners to rate the "perceived tonality" of short melodies drawn equally from the different levels of melodic structure. Their listeners were classed in three groups: advanced music students, less advanced music students and students with little or no musical training. The results from all three groups agreed very closely with the levels of the classification system, indicating some criteria for recognising "adherence to tonal rules" that was shared by musicians and non-musicians alike, and hence could be ascribed to a process of acculturation rather than to formal training. In a further series of experiments, listeners were asked to differentiate between correct and incorrect transpositions of the same short melodies; again, ease of error recognition correlated strongly with the degree to which melodies conformed to the "tonal rules".

It would seem that Cuddy's scheme for differentiating between levels of melodic structure does capture some essential features of how melodies are perceived - and by inference, encapsulates something of the cognitive representation underlying musical perception. However, as stated, she fails to make explicit the exact criteria for her different levels of analysis. Unlike Dowling, she conflates scalar and modal structure in presenting her scheme in a way which leaves the criteria for construction of her experimental melodies and the significance of her results open to a range of interpretations. For instance, her finding that "implied" harmonic structure is of primary importance is in fact not necessarily borne out by the melodies that she uses. Other explanations - for example, expressed in terms of pattern simplicity, or the degree of parsimony with which a melody could be represented in perception and memory - might be equally well advanced.

Deutsch (Deutsch and Feroe, 1981: Deutsch 1982) in fact adopts this somewhat different approach - of pattern parsimony - to the question of scale and key representation in cognition. She proposes a model of pitch relations which is based on the idea of different overlearned pitch alphabets. Thus the chromatic scale (Dowling's tonal material), the diatonic major scale and the minor scales (Dowling's tuning system) and the triads formable within the major and minor scales are all taken as constituting alphabets and as having some cognitive representation in that form. Melodic sequences can be generated or described in terms of elementary operations (such as "same", "next", "prior" etc.) carried out on reference elements within these alphabets. Thus the operation "next" carried out on the reference element C within the alphabet of the diatonic major scale of C would produce the note D. A further application of the same operation would produce the note E. In this way the sequence of notes C, D, E could be represented by the "next" operation carried out twice on the note C, or rather by the <u>recursive</u> application of the "next" operation to its product (D) when applied to the note C. Longer sequences may be described in terms of a further set of operations (such as "prime", "inversion" or "retrograde") carried out on shorter sequences in a similarly recursive way. This use of recursion to describe the relationship between single notes and sequences, and between sequences and larger sequences of notes, provides an explicitly hierarchical model, larger components encompassing subcomponents (for a more detailed account, see West, Howell and Cross, 1985).

By representing melodic sequences as deriving from recursive application of operators to reference elements drawn from alphabets, Deutsch and Feroe provide a means of assessing the "information load" of different melodies (i.e. the degree to which - and the manner in which - melodic sequences can be represented as multi-levelled hierarchical structures), the implication being that listeners will interpret melodies in perception and retain melodies in memory in ways which minimise their information load. For example, the (ascending) sequence B-C-Eb-E-F#-G-B-C can be encoded in hierarchical (recursive) terms fairly concisely: as a thrice repeated "next" operation on the reference element C in the alphabet "C major triad" (producing the notes C-E-G-C), operated on - by use of the prime operator - by a further sequence constituted of a "prior" operation on a reference element in the chromatic alphabet, the reference element being each note in turn of the first-described sequence. Although this sequence description may seem fairly verbose, it is certainly capable of being more parsimoniously described within Deutsch and Feroe's scheme than is, for instance, the sequence Eb-B-E-F#-C-B-G-C (which uses the same notes re-ordered).

Deutsch (1981) presents the results of two experiments designed to test the hypothesis that listeners would interpret melodies in perception and retain melodies in memory so as to minimise the melodies' information load in accordance with Deutsch and Feroe's scheme. She found that musically trained listeners were better able to recall sequences correctly when those sequences were capable of parsimonious description and were performed with temporal gaps occurring between subsequences which could have some simple pattern representation as part of the whole sequence. So, for example subjects could recall a sequence such as G-F#-G-D-C#-D-B-A#-B-G-F#-G fairly easily, particularly if it was segmented in time so as to occur in three-note chunks. Conversely, subjects found difficulty in recalling such sequences if they were segmented into four-note chunks, and could not easily recall sequences with complex non-recursive pattern descriptions such as F#-B-G-A#-D-B-A#-C# etc. Deutsch takes these results to indicate that listeners do encode melodic sequences in perception and memory in some way that is consonant with Deutsch and Feroe's pattern-description system.

Several objections can be raised to Deutsch's model, and to her interpretation of these experimental results. Although she demonstrates an apparent effect of pattern parsimony on recall performance this does not necessarily validate Deutsch and Feroe's scheme; the sequences which in her experiments are considered as "well-formed" in terms of pattern-description economy are equally "well-formed" in terms of other - different - pattern-description systems such as that of Simon and Sumner (1968). Moreover, she does not consider the strong possibility that to her musically-trained subjects these "well-formed" sequences might be overlearned melodic formulae (as the sequence G-F#G-D-C#-D-B-A#-B-G-F#-G undoubtedly is in Western common-practice period music); enhanced performance on such melodies might simply reflect their overlearned status. However, more fundamental objections to her theory can be raised.

The theory has been criticised by Jones (1981); as she states, there are no apparent criteria in the theory for identification in perception of the reference element within a sequence, which would seem to make sequence encoding within the system somewhat probabilistic. Other criticisms that can be raised relate to the implausibility of the parsing mechanism implicit in the scheme; the alphabets and operations in use in sequences can in some instances only be identified retrospectively, which would seem to work against the prospective, right-branching nature of the scheme. It is when the scheme is assessed as a theory of the cognitive organisation of musical pitch, however, that its lacunae are most apparent. It offers no means of relating alphabets; although the same note might occur in several different alphabets, and although certain alphabets contain others (e.g. the diatonic major alphabet on G contains the major triad alphabet on G), there is no attempt to resolve the ambiguities that these multiple possible attributions may give rise to in sequence description. There is no equivalent of Dowling's "modal level" of analysis, wherein certain notes within the tuning system may be accorded some referential or dominant status. Moreover, the model's strict differentiation between alphabetic and syntactic structure offers no possibility of encapsulating anything much of Powers' idea of the distinction between mode and scale.

It would appear that none of the models of musical pitch representation so far described offer possible rationales for the particular forms that pitch takes in western tonal music; the value for music theory of these schemes and their associated experimental findings seem to be more confirmatory than explanatory (although the intended generality of Deutsch's pattern representation model should have enabled it to be used in an explanatory way, were it not for its intrinsic defects). It must be said, however, that all these models of the cognitive representation of musical pitch are in some way incomplete; for instance Dowling, Deutsch and Cuddy present some implicit model in using the concept of key-relatedness to account for experimental findings, yet the structure underlying such relatedness is not spelled out. More fully articulated models have been developed by other psychologists; these structural approaches, together with experimental evidence adduced in their support, will now be considered in detail.

CHAPTER FOUR

Structural representations of musical pitch

Shepard (1964) laid the foundations for a structural approach to the cognitive representation of musical pitch in proposing that pitch is better thought of in terms of a "multidimensional cognitive-structural representation" - which could be construed as acting as a form of schema (in Neisser's terms, see above) - rather than in terms of a unidimensional psychoacoustic scale - such as the "mel" scale proposed by Stevens, or even a logarithmic scale as suggested by the results of Attneave and Olson (1971). That is, although pitch can be conceived of as the simple psychological analogue to the unidimensional physical parameter, frequency, enabling a one-to-one mapping between the frequency and the pitch domains (as in psychoacoustical accounts), Shepard argues that a more appropriate account of the "relations holding between pitches of tones that are interpreted musically" will be arrived at by "focusing on the precise form of the interpretative cognitive structures [schemas] themselves". Several of the results of the experiments on the perception of music just discussed, together with some musical intuitions about pitch and pitch relations, indicate that musical pitch cannot be characterised in "simple" psychoacoustical terms; some cognitive representation or model must be sought that is consistent with the psychoacoustical facts, but also encapsulates the types of relations apparent in musical perceptual judgments and rememberings.

Shepard's suggestion that pitch is best conceived of in terms of a multidimensional spatial schematic model - or mental image - is motivated by a concern to capture the essence of three conditions which he states underlie all psychoacoustically-based models of musical pitch, while also capturing relations that have been experimentally shown to hold between pitches in a musical context (Shepard, 1982a, 1982b). These three general psychoacoustic conditions are unique correspondence (each pitch corresponding to a particular point in a metric space), preservation of equivalence (pitches perceived in the same psychological relation corresponding to points separated by the same distance in the space) and monotonicity (pitches perceived to be more similar to a given pitch correspond to points that are closer in the space to the point corresponding to the given pitch). These three conditions can be thought of as informal statements of the underlying principles of representation within a metric space, the principles of (i) positivity, (ii) symmetry and (iii) triangle inequality. That is, (i) two different points (or representations of individual pitches) cannot occupy the same spatial position (otherwise they must be regarded as the same point or pitch), (ii) the distance between two points or pitch representations must be the same whether measured from point a to point b or vice versa and (iii) if points a, b and c do not lie in a straight line the distance from point a to point c must be less than the sum of the distances from point a to point b and from point b to point c.

Initially, Shepard (1964) attempts to account for the phenomenon of octave-equivalence; accordingly, his model makes use of pitch height and pitch chroma (see above), chroma being represented as a circle (two-dimensional) on which pitches are conventionally ordered by note-name (i.e. C, C#, D, D# etc.)

while pitch height forms a linear (unidimensional) continuum. The integration of these two components into a single spatial model produces a helix, which Shepard construes as completely regular; thus pitches separated by an octave will fall on the same straight line on the surface of the cylinder "enclosed" by the helix. Shepard (1982a, 1982b) acknowledges that this model had been proposed previously (by Drobisch in 1855, Revesz in 1954 etc.); however, Shepard (1964) adduces his own experimental evidence in support of the claim that this model provides an accurate account of perceived musical pitch. In his experiments, Shepard produced the illusion of "eternally-rising" (or falling) chromatic scales; these were generated on computer, using complex tones having only spectral components related by powers of two (i.e. using octave-related partials); these were played to listeners through a broad-band filter. According to Shepard, this should saturate the height dimension so that pitch can only be differentiated in terms of chroma; in this way, although listeners are able to judge movement from note to adjacent note in the chroma circle as taking place in the pitch height and the chroma domain, when the octave is reached listeners cannot differentiate between the pitch of this note and that of the note an octave below, giving rise to the "continued rise or fall" illusion.

However, the claim that this illusion arises because of the nature of the cognitive model whereby pitch relations are represented is challenged by the fact that Burns (1982) was able to produce the same illusion using complex tones having only spectral components related by powers of inharmonic numbers (e.g.

2.119/1). Burns points out that not all of his subjects experienced the experimental material as "continuously-rising", and suggests that the illusion arises because each tone will possess a large number of possible fundamental-candidates (comparable to Terhardt's spectral pitches); as Burns points out, "the task is then essentially a multi-channel listening task rather than the tracking of a single pitch, with the majority of channels showing an increase in pitch for each successive tone". Burns' results favour a psychoacoustical rather than a "cognitive-structural" interpretation of the processes underlying the illusion, thus contradicting Shepard's appeal to the illusion to substantiate his model.

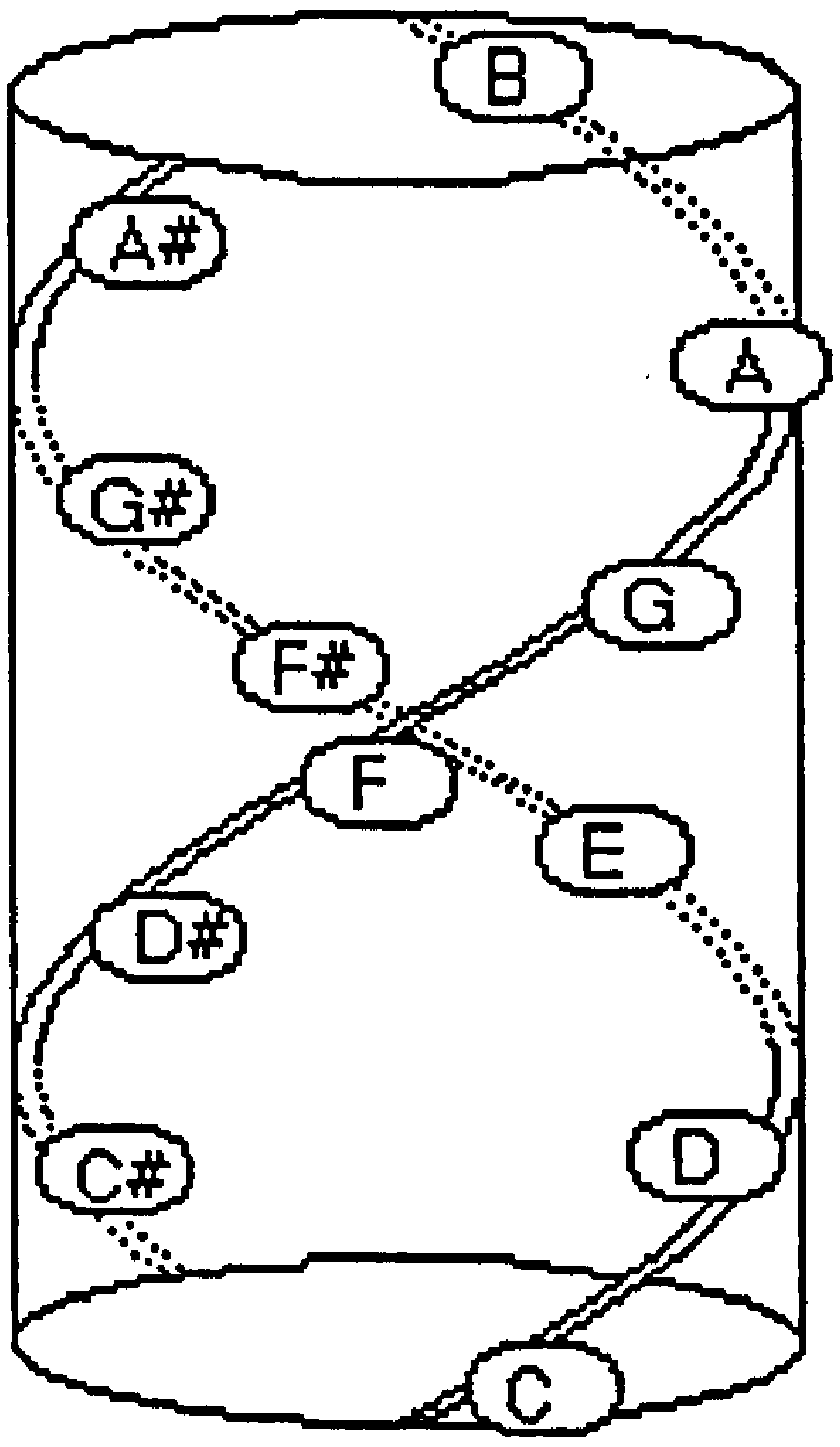
In fact, this refutation of Shepard's evidence for his model can be seen as at least partially undermining his differentiation between psychoacoustical and cognitive-structural theories of musical pitch; as is evident in the earlier discussion of Terhardt's theories, the "direct" one-to-one mapping between frequency and pitch that Shepard argues against no longer constitutes a significant feature of most current (two-component) psychoacoustical theories of pitch perception, and its rejection cannot of itself constitute a rationale for any multi-dimensional model of musical pitch being construed as necessarily "cognitive-structural" as opposed to psychoacoustical. However, there is considerable evidence - as presented in the previous part of this Section - for the perception of musical pitch relations that are not easily or well characterised in terms of recent psychoacoustical theories; despite the problems of this early version, the general constraints made explicit in the theory have enabled Shepard to develop more complex and explanatory versions.

The double helix representation of musical pitch

Shepard's recent (1982a, 1982b) version of the "cognitive-structural" model aims to account for many different types of perceived musical pitch relations in terms of the proximity of pitches within a complex spatial representation; spatial proximity is intended to model the degree to which pitches might be perceived as similar within a variety of musical contexts. This later model makes use of three components, of which one is unidimensional (pitch height), two being two-dimensional (the chroma circle and the circle of fifths), giving rise to a five-dimensional representation which has the form of a double helix wound round a helical cylinder. The component added to the earlier model, the circle of fifths, has its origins in conventional music theory where it functions as a means of indicating key-relatedness or key-signature; as used in Shepard's model it consists of the notes of the equal-tempered chromatic scale laid out in a circle so that each pitch forms the enharmonic musical interval of a perfect fifth with the notes on either side of it, and is intended to capture the notion that pitches separated by a fifth can be perceived as closely-related in some musical contexts.

Shepard points out that as well as representing octave-equivalence, this new model has "two related properties of some musical significance"; these are most easily described in terms of the three-dimensional structure formed by the pitch height and circle of fifths components. This structure takes the form of a double helix on the surface of a regular cylinder (Figure 4.1);

Figure 4.1: The double helix representation of musical pitch in three dimensions. The circular component corresponds to the circle-of-fifths, while the linear (vertical) component corresponds to pitch height (after Shepard, 1982a).

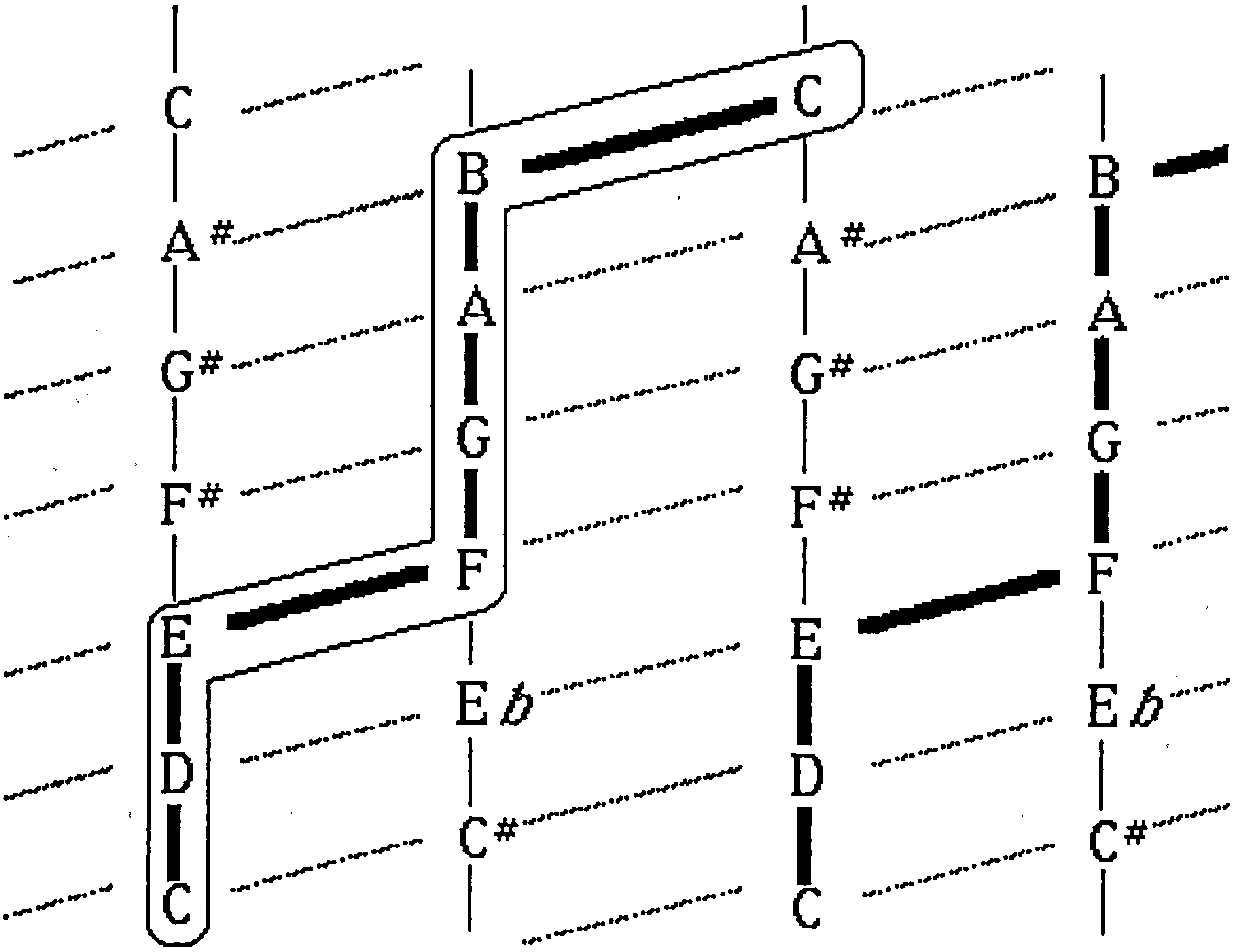


in this representation, the notes of any major diatonic key can be divided from the notes not in that key by passing a plane through the central axis of the double helix. Moreover, transposition into the most closely related keys are achieved by the smallest angles of rotation of the dividing plane about the central axis. Adding in the (two-dimensional) pitch chroma component, Shepard (1982a, p 364) states that the resultant five-dimensional structure can "account for perceived similarity of tones proximate in pitch height, heightened similarity at the octave and at the perfect fifth, as well as accounting for the separability of tones within a major diatonic key from non-key tones and the rotational proximity of closely-related keys".

In this way Shepard provides an "imagistic" cognitive-structural representation of musical pitch that integrates many of the features of perceived relations between pitches that experiments have shown to be significant; it would seem to be capable of explaining many of the observed judgmental and memory abilities of Western musical listeners. The model provides a coherent and logical basis for cross-octave chroma "identity", for the "privileged" status that notes within a key have in respect of one another, and for the idea of key-distance relations (which Dowling and Bartlett (1980) point up as important without outlining the structures that might underlie such relations). Moreover, the model could be extended if the experimental evidence so required, e.g. by the addition of two-dimensional thirds-space (Shepard, 1982a, p 365). The model can also be made to account for a range of further equivalences. For example, Dowling's (1978) finding of confusion between real and "tonal" transpositions - which seems to imply that semitones and tones can be perceived as functionally equivalent scalesteps within a key - can be accounted for by adopting Shepard's (p 357, p 384) suggestion that different components can be appropriately scaled or "weighted" so as to make the distance within the model representing an interval of a tone equal to the distance representing a semitone.

Shepard uses two topological transforms of the model to demonstrate the way in which it embodies significant aspects of musical pitch perception. One affine tranform - a two-dimensional projection of a one-octave segment of the helix (see **Figure 4.2**) - is used to show how the model can display the notes of the major diatonic scale as a connected but asymmetrical region of "melodic space". Shepard points out that this asymmetricality encapsulates the irregular intervallic structure of the diatonic scale, and can be thought of as "modelling" the fact that each note of that scale forms a unique set of intervals with all other scale notes, this remaining constant under transposition. In theory, this fact should enable any single scale-note to be identified in perception on the basis of the intervals that it forms with other scale-notes, irrespective of the order in which these occur - that is, to be identified on the basis of the particular alphabetic structure of the diatonic scale, a possibility which will be explored in a later section of this thesis. He then derives a further affine transform from the helix, displaying harmonic intervals (thirds and fifths) as steps between adjacent notes. This "harmonic" space has the property that any group of three adjacent (triangularly-related) notes forms a major or minor triad, while the diatonic scale remains

Figure 4.2:Two-dimensional structure obtained by "unwrapping" the helix, showing the asymmetrical structure of the diatonic scale (after Shepard, 1982a).



depicted as a connected spatial region. This latter space is closely related to both the charts of "key-regions" presented by Schoenberg (1969, p 54) and by Ellis (in the Appendices to his translation of Helmholtz, 1885/1954). It is also, as Shepard states, directly analogous to that proposed by Longuet-Higgins (1979) and by Balzano (1980).

The "cognitive-structural" model - experiments

Shepard, Krumhansl and others have carried out several series of experiments to uncover the precise forms of the "interpretative structures ... underlying the the cognitive organisation of musical pitch". These experiments used a range of measures to establish how pitch might be cognitively represented. Ratings of perceived similarity between pitches, or of the "fittingness" of pitches to complete short melodic phrases were elicited, or the time taken to discriminate between pitches or chords was measured. Almost all of these measurements were made in the context of melodies or chord sequences intended to evoke a tonal schema by providing an appropriate context for the perception of pitches and pitch relations.

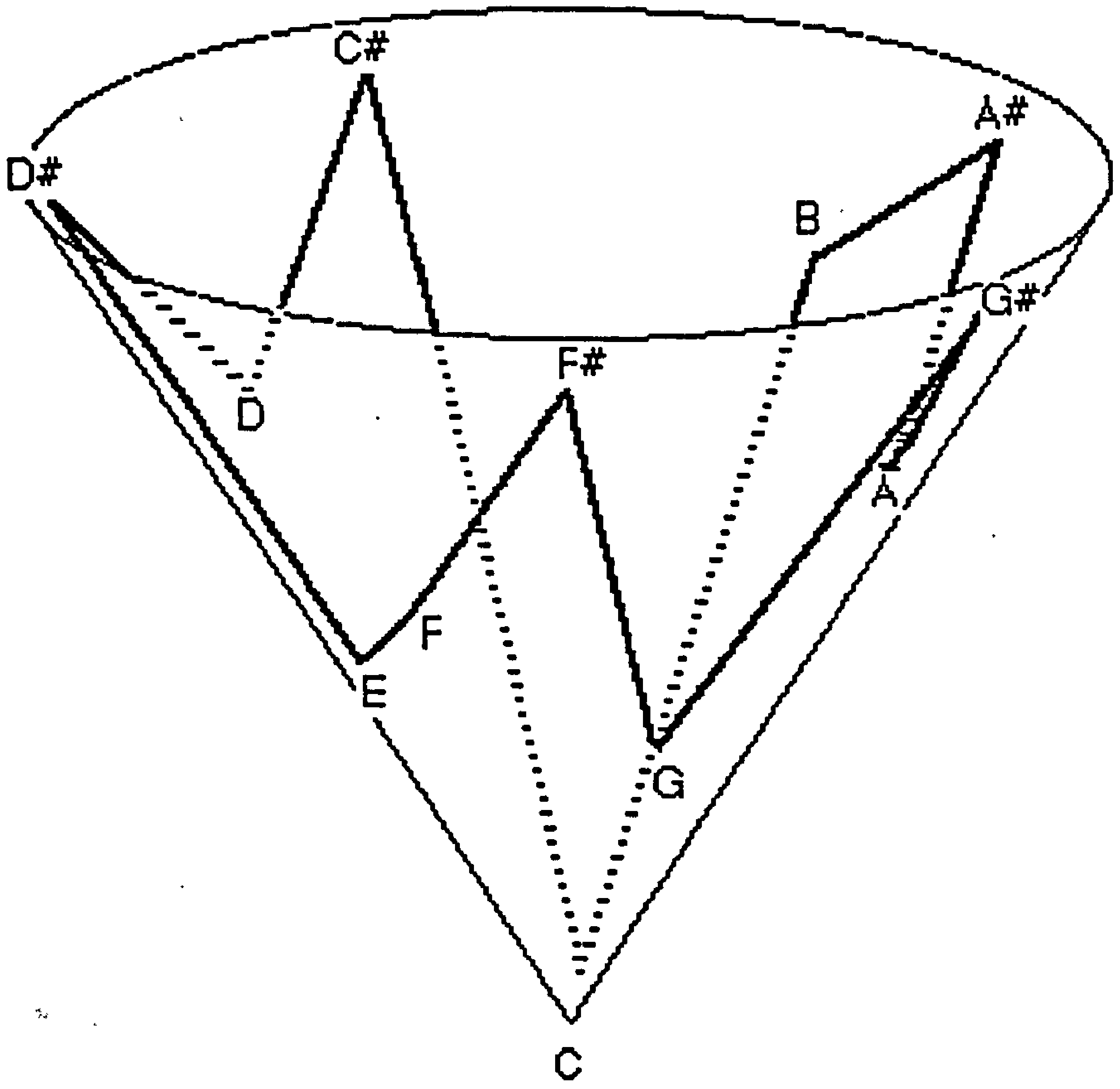
Krumhansl and Shepard (1979) asked listeners (divided into three groups on the basis of musical experience) to rate how well each chromatic note within an octave range completed a seven-note ascending or descending major scale (always starting on the tonic). They found that their least-musical group responded on the basis of pitch height (i.e. notes close in pitch height to the last note of the context were rated highly): moderately-experienced listeners rated notes most highly on the basis of their proximity within the chroma circle (i.e. the notes C, B, C#, A# would all be rated fairly highly as "completions" of a C major scale irrespective of octave, while F# received a low rating): more experienced listeners, however, tended to rate notes on the basis of their diatonicity irrespective of octave and on the basis of their relative "strength" or stability within a conventional tonal hierarchy (i.e. with the tonic most highly rated, followed by the dominant and the mediant, then the other diatonic notes). Shepard (1982a, p 365-369, 1982b) relates these findings to the multi-dimensional model, claiming that the responses of the most experienced group of listeners exhibit the same dimensions as are used in the model and can thus be construed as deriving from the listeners' use of a representation very similar to the model in the judgments.

Krumhansl and Kessler (1982) replicated and extended Krumhansl and Shepard's findings; they used a number of different contexts, including ascending/descending major and minor scales as well as major or minor chords and chord cadences (V-I). They found results similar to those of Krumhansl and Shepard, although their results for minor-key contexts showed a slightly higher rating (stability) for the mediant than for the dominant (in contrast to their major-key results). Again, these findings are directly expressible in terms of Shepard's model, in terms of the affine transform of the original spiral that yields "thirds-space"; in this representation, the "stable" notes of the tonal hierarchy can be seen to form a group of proximate elements within a key.

However, the results of Krumhansl's (1979) study are not so confirmatory. In her first experiment, she presented subjects with context sequences of either ascending or descending major scales or major triads, followed by two notes; all possible ordered pairs of chromatic notes within an octave range were presented, and listeners rated how similar the second note of each pair was to the first in the context provided. The results of this experiment were subjected to a multidimensional scaling solution, which in this instance took the form of a cone, with the notes of the tonic triad disposed in a circle around the surface at the vertex, other diatonic notes in a circle further from the vertex and non-diatonic notes in a circle furthest from the vertex (see **Figure 4.3**). Krumhansl takes this finding to indicate that the constant "cognitive distance" that musical intervals represent in Shepard's spatial model may not hold in a well-established tonal context. For example intervals of a major third - as between C and Ab, or between C and E - would all be represented by a similar distance in Shepard's model, whereas in Krumhansl's conical solution these particular notes would be separated by quite different distances; C and E would be relatively close, while C and Ab would be considerably further apart.

Moreover, Krumhansl found a regular pattern of asymmetries in the similarity ratings; as she states (1979, p 361) "tones less closely related to the established tonality [e.g., in the key of C, the note C#] were judged more similar to tones more central to the tonality [e.g. C] than tones central to the tonality were judged to tones less related to the tonality". She takes this to indicate that "the set of tones in a tonal context appears to contain certain members that act as reference points",

Figure 4.3:Multidimensional scaling of relatedness judgments in a C major context, showing members of tonal hierarchy closest to vertex of cone, other diatonic notes further from vertex and non-diatonic notes in a circle furthest from the vertex (after Krumhansl, 1979)



or, following Rosch (1975), "prototypes" for the set. Neither of these asymmetries are easily reconcilable with Shepard's model; it might be that they arise because of the difference in method from the previously-cited studies (using similarity judgments rather than rating notes for "fittingness") together with some susceptibility of listeners to effects of order in making similarity judgments over time. The difficulties that this asymmetry presents for a spatial, "imagistic" concept of pitch representation will be explored further in the next chapter.

Krumhansl and Keil (1982) carried out a further experiment to attempt to trace the developmental course of the acquisition of the tonal hierarchy. They played short melodic sequences to three groups of school-children of different ages, and to a group of adults; each sequence consisted of four-note "tonal contexts" (taking the form C-E-G-C) followed by either two in-key notes, one in-key and one out-of-key note or two out-of-key notes. Subjects were told that each sequence was the beginning of a melody, and rated each sequence according to "how good or bad it sounded". Children in the lowest age-group gave significantly higher ratings to sequences which terminated with in-key notes than to those which ended with a note-pair containing one or two out-of-key notes. Responses of children in the middle age-group showed the same pattern, although additionally exhibiting higher ratings for sequences which terminated with a note-pair of which both notes were members of the tonic triad. The highest age-group showed similar results to the middle age-group, while further distinguishing between sequence terminations containing one and two members of the tonic triad in their responses. The adult group showed a similar pattern of responses to those of the oldest children, while also showing a preference for sequences terminating on the tonic. There was a constant effect of pitch height in judgments of all groups, but this did not interact with the other effects found.

Overall, the experiment showed that the acquisition of the tonal hierarchy thus proceeded in an orderly manner, with the earliest developmental distinction to emerge between pitches in a tonal context being between in-key (diatonic) and out-of-key (non-diatonic) notes and a progressive refinement of judgment occurring in distinguishing between notes central to the key and those less central. Krumhansl and Keil explain this by stating (1982, p 250) that "it is essential to first develop an abstract internal framework that matches the most frequently occurring patterns of the scale. Once the framework is established, it is used to disambiguate the functions of the individual tones within the musical context". These results would seem to ally themselves with those of Krumhansl and Shepard (1979) and Krumhansl and Kessler (1982) in implying that listeners may make use of some representation similar to Shepard's model in the perception of musical pitch.

A "cognitive-structural" model for chords and keys - experiments

Krumhansl followed up these experiments on the perception of notes and intervals in a tonal context by testing whether or not a similar - or similar type of - cognitive representation might underlie the perception of chords and chord-relations. Krumhansl, Bharucha and Kessler (1982) again made use of the "tonal-context-and-fittingness" experimental paradigm; subjects heard an ascending or descending major or minor scale, followed by a pair of major or minor triadic chords, and rated how well the second of each pair of chords followed the first in the given context. They found in a multi-dimensional scaling solution that chords that could be construed as being on the tonic (I), dominant (V) and subdominant (IV) of a key clustered together, while chords on the supertonic (ii), mediant (iii), submediant (vi) and leading-note (vii°) were more distant. They also found an order effect similar to that obtained by Krumhansl (1979), in that when the second of the chords was drawn from "central core" (i.e. I, IV, or V) of the context key and the first was not, ratings were higher than when the order was reversed. They take these results to indicate that a hierarchy (in terms of relative centrality or stability) exists within the cognitive representation of the set of harmonies of a key, with the ordering of relative stability corresponding to that suggested in music-theoretic accounts (e.g. Schenker, 1906/1954) but differing from the hierarchy obtained for the corresponding root-notes of the chords. Their results also showed that chords belonging to different keys tended to be grouped accordingly, with chords that could be construed as "shared" by two or more keys falling fairly close together within the multidimensional spatial model, while chords unique to a particular key tended to fall further away from this "central" region. Thus chords which occurred (or fulfilled functions) in different keys - for example, the chords of C and G in C major and G major, respectively the tonic and dominant, and the subdominant and tonic - would be grouped relatively closely together, while the chords of F (in C major) and D (in G major) would lie much further apart, as F (C major's subdominant) does not occur in G major, nor D (G major's dominant) in C major.

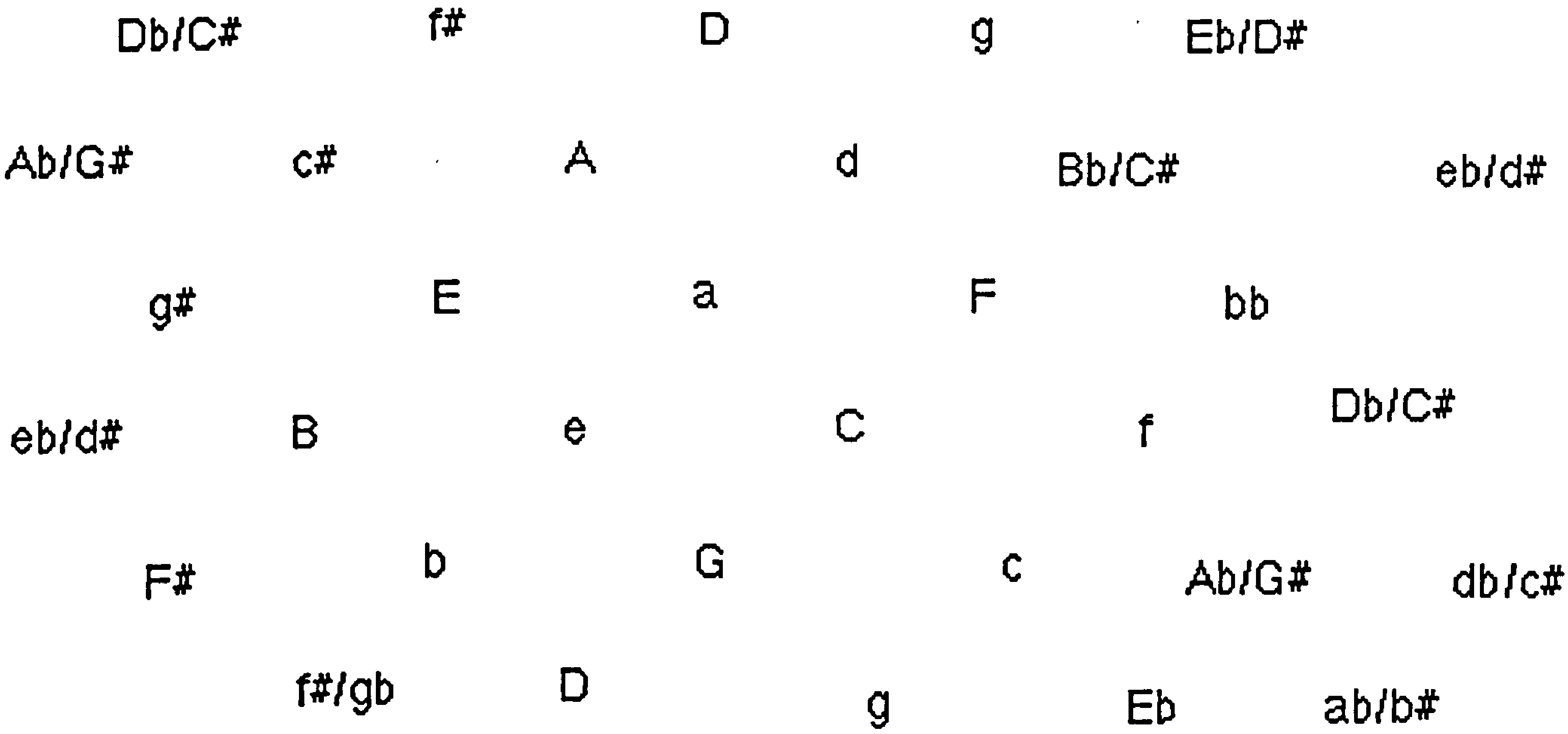
This model of the cognitive representation of intra- and inter-key chord relations has been extended by Krumhansl, Bharucha and Castellano (1982) and by Bharucha and Krumhansl (1983). They presented pairs of chords from the distantly-related keys of C major and F# major in the context of C major or F# major or with no context (Bharucha and Krumhansl, 1983) or in the context of G major, A major or B major (Krumhansl, Bharucha and Castellano, 1982); in both experiments the contexts (when provided) consisted of a IV-V-I cadence in the context key. The rationale behind the experiments was to test the limits of the key-distance effect, C major and F# major being as distantly-related as possible, as well as to test the effects of key-proximity; G major is close to C major and B major is close to F# major, while A major (according to conventional music theory) is equally distant from both C and F# major. Similar asymmetries in ratings of "core" chords (due to order of presentation) were found as in Krumhansl, Bharucha and Kessler's (1982) experiments, as well as strong effects of context on grouping of chords in multidimensional space. In the C major context, chords drawn from the key of C major were closely grouped, while F# major chords were separately and distantly dispersed; the converse held for F# major chords in the F# major context. In the G major context, C major chords were again closely grouped, F# major chords being dispersed, the converse holding for F# major chords in the B major context. Moreover, this separation in multidimensional space was maintained when both no context and when an A major context were used; the chords of C and F# were located in separate spatial regions, although exhibiting the typical core/other-diatonic-chord spatial layout somewhat more weakly than in the presence of appropriate contexts.

Krumhansl (1983, p 53) summarises these results as showing that in the cognitive-structural representation of pitches and keys, "for both tones and chords a central compact cluster of musical elements most central to the key was found, with other elements more distant from the core and from each other...key membership and membership in significant chords within keys were reflected in the results for single tones, structural relations between chords being governed by the harmonic function of chords in different keys...and being modified by the context in which the chords were embedded". She argues that this indicates that listeners have abstracted certain underlying principles of the functioning of tones and chords within musical keys and make use of this abstraction during listening to encode and remember individual elements, and that both key structure and chord-key relations are consequences of the hierarchy of structural stability that applies to the set of chromatic notes in tonal music.

Krumhansl (1983, 1985) presents a spatial model of "key-space" (geometrically relatable to Shepard's (1982) "thirds space" as well as to that of Balzano (1980) and Longuet-Higgins (1979)), derived from correlations between the rating profiles of single notes in key-contexts that were found by Krumhansl and Kessler (1982), and bearing a strong similarity to that proposed by Schoenberg (1969). In this spatial representation, adjacent keys are related by the interval of a perfect fifth in one dimension, and by a minor third in the other. Between each of the keys related by a minor third lies a parallel column of "parallel minors", so that, e.g., between C and Eb lies c, and between A and C lies a. Thus keys with a dominant relationship (V-I, I-IV) will be adjacent, as will relative minor keys (vi-I, i-IIIb) (see **Figure 4.4**).

Krumhansl (1983) uses this model to account for the results of Krumhansl and Kessler's (1982) experiment on how a sense of key develops and changes over time in a listener's perception of a series of chords which modulated to a close (e.g. from C to G) or to a distantly related (e.g. from C to F#) key. It was found that for close modulations, as soon as a chord in the new key was sounded a listener's sense of the new key was stronger than when the chord was heard in isolation (indicating that the listener had achieved a better realisation of where the chord was located in the "key-space"); for distant modulations, the sense of the new key was achieved only later in the sequence. However, in sequences comprising distant modulations (ending in a distant key), the sense of the initial key was more completely suppressed than was the case in sequences comprising close modulations (indicating that the listener's sense of harmonic identity and function was orientated towards a more distant region of key-space, and was "unconnectable" with the earlier established key-space region). Krumhansl (1983, p 57) takes this pattern of results to imply that "listeners integrate information over the successive chords of a sequence, dynamically generating hypotheses about likely key regions through interpreting chords in their multiple harmonic roles...the processing of these multiple harmonic functions is supported by an internal

Figure 4.4: "Harmony space" (after Krumhansl, 1983) showing fifth-related chords adjacent along one diagonal, and "minor-related" chords adjacent along the other. (upper-case=major, lower-case=minor)



representation in which each chord is closely associated with the various keys in which it functions, and keys sharing chords are perceived as closely related...the listener is thus able to achieve an initial key interpretation, and to assimilate modulations to new keys, particularly those that are closely related".

Cross-cultural evidence for cognitive-structural model

The idea that some cognitive-structural representation of a hierarchy of stability between the notes of a scale-system underlies perception of pitch relations receives support from the results of the study by Castellano, Bharucha and Krumhansl (1984) on the perception of pitch in the music of North India. Pitch organisation within this complex musical culture is most obviously manifest at the level of the raag, a term which is used to specify not only which notes should be used in particular compositions and improvisations but also which melodic formulae and which phrasal articulations are permitted and which notes should be more functionally significant within these phrases and melodic formulae. That is, a raag is a form of pitch organisation classifiable as a mode; within Powers' scale-melody continuum, it lies considerably closer to melody than to scale, taking the form of "generalised melody". In a piece of typical North Indian art-music, following a metrically-free alap, a fairly brief set composition (gat) will precede the main improvisatory body of the performance; the raag upon which the piece is based underlies all sections of the piece equally.

Underlying the raag level of pitch organisation is that based on the thaat; several raags can be derived from the same thaat, which functions to determine which notes will be available, and may also specify which of these should be more functionally significant in the music. A thaat can be conceived of (within Powers' continuum) as a "particularised scale", having seven notes of which two - Sa and Pa - are fixed while the remaining five (Re, Ga, Ma, Dha, Ni) may be chromatically inflected (though only in one direction - e.g. the note Re can appear as Reb, but not as Re#). Both raag and thaat are essentially modal levels of pitch organisation though occupying very different positions within the scale-mode-melody continuum, thaat being closer to scale and raag closer to melody. Indian music theory has developed its current model of inter-thaat relations only relatively recently (post-1930). In this model, 10 thaats can be laid out in a circle, each thaat sharing six of its seven notes with each adjacent thaat on the circle (although the theoretical nature of the "circle-of-thaats" is pointed up by the fact that one of these - No 7 - is not, in fact, used in Indian music). An eleventh thaat, Bhairavi, exists; this cannot be derived from any of the other thaats by simply altering one note, and is is held to occupy a position inside the circle. Although this configuration appears superficially similar to the Western "circle-of-fifths", movement from one thaat to an adjacent thaat in the circle should not be construed as a "transposition" of the original thaat to a new pitch level (as such a movement from one scale to an adjacent scale in the circle-of-fifths would be construed), but rather as a rearrangement of the intervals available within the original thaat. The "absolute" pitch of the fixed notes, Sa and Pa, does not alter from one thaat to another, and in any raag these two notes are generally functionally significant (being omnipresent in the drone which accompanies each North Indian piece as well as frequently starting or ending phrases). However, for each different raag, two notes - which may be other than Sa and Pa - are designated as vadi and samvadi; these also play a functionally significant role in the pitch structure of a raag, and may be used theoretically to differentiate between two raags that are in the same thaat. Although most of the notes of a piece in one particular raag will be those of its parent thaat, some will not; these, however, will function purely as pien (subsidiary) tones or grace notes, their function being ornamental rather than structural.

Castellano, Bharucha and Krumhansl (1984) made use of the "tonal-context-and-fittingness" experimental paradigm; in this experiment the "tonal context" consisted of the first part (sthayi) of a set composition or gat (which, at an average of 65 seconds, lasted considerably longer than those contexts used in earlier experiments of Krumhansl et al. on Western music), followed by a short theme or "exposition" of the characteristic features (i.e. melodic formulae) of the raag upon which the sthayi was based. This was followed by a note drawn from the chromatic scale, which listeners rated for fittingness with the preceding context. Each context was accompanied by a drone consisting of the notes Sa and Pa (analogous to C and G); this drone did not accompany the probe note. They made use of two groups of listeners; one group consisted of listeners with considerable exposure to and practical experience in North Indian music; the other group were listeners with some formal Western musical training but little experience of Indian music.

Castellano et al. found that a coherent structure underlies the fittingness ratings of the chromatic tones used. Sa was rated as best fitting, followed by Pa and then the vadi of whatever raag was implicit in the context; there seemed to be a definite preference for notes included in the thaat over non-thaat notes. By-and-large, few differences were found between the ratings provided by Indian listeners and those given by the other group. Given this lack of significant difference between the two groups' responses, Castellano et al. hypothesised that the ratings might be based on the different relative total durations that notes had within each context (i.e. derive from the proportion of the total context over which each note sounded); an analysis showed a very high correlation between both groups' ratings and the total sounding duration of each note. However, as this variable was found to be confounded with whether or not notes were members of the underlying thaat, a multiple correlation was performed to assess the relative contribution of thaat membership and duration to the ratings. Setting duration aside, it was found that thaat notes were preferred over non-thaat notes only by Indian listeners, while Western listeners' ratings could be accounted for purely on the basis of sounding duration. Similarly, when a multi-dimensional scaling of inter-raag correlations was carried out, the two-dimensional solution for the results of the Indian listeners conformed to the theoretical "circle-of-thaats" while the Western ratings did not.

From the results of this experiment together with the previous series, Castellano et al. conclude that, in respect of both Western tonal music and North Indian music listeners can abstract - as some form of tonal hierarchy, embodying the different "perceived stabilities" of notes - the principles of modal or tonal organisation that underlie note relations within individual musical contexts, and make use of this abstraction during listening to encode and remember individual elements. They also infer that longer-term exposure to music within some particular idiom enables the development of a structured representation of relations between these tonal hierarchies (intra- and inter-thaat relations in the case of Indian listeners, intra- and inter-key relations for Western listeners).

The "tonal hierarchy" sensitivity shown by both groups of listeners within Castellano et als. experiments was largely accounted for by the relative durations of the notes of that hierarchy within the experimental musical contexts used. Another experiment (Hansen, Kessler and Shepard, 1984) on Western and Balinese music, using similar methods and using Western diatonic and Balinese contexts appears partially supportive of this conclusion, although the pattern of results seems somewhat unclear (perhaps owing to the fact that differentiation between modes in Indonesian and Balinese music may rely more on contour rather than on sounding note duration, see Hood, 1954 and Becker, 1982).

The cognitive-structural model - implications for perception

Castellano et al. thus state that the aspect of pitch organisation within a piece of music that is most influential in a listener's abstraction of the tonal hierarchy - or assignation of different degrees of stability to different notes - is the relative sounding duration of the various notes in the piece; they add (p 409) that syntactic features - such as which notes begin and end phrases, occur in positions of rhythmic stress etc. - may also serve to instantiate the representation of the tonal hierarchy. Krumhansl (1985) amplifies the idea that relative sounding durations of notes is a strong determinant of the structure of the internalised tonal hierarchy. She states (1985, p 373) that "our experimentally quantified tonal hierarchy may be related to the statistical properties of music", that is, might be related to the relative frequencies of occurrence of different scale-notes in tonal music. In support of this theory she cites Hughes' (1977) study of the statistical distribution of notes in Schubert's Op 94, the results of which correlate very highly (0.89) with the structure of the tonal hierarchy found by Krumhansl and Kessler (1982), in which the tonic was highest-rated (occurred most often), and was followed by the dominant (next most frequent), the mediant, etc.

A set of papers produced in response to Castellano et al. follow up the question of how the tonal hierarchy might be abstracted and made use of in musical perception. Deutsch (1984) suggests that abstraction of the tonal hierarchy might in certain instances be underpinned by low-level sequential grouping processes (determined by the nature of short-term memory or by the the constraints of an encoding process such as that outlined in Deutsch and Feroe, 1981), with greater prominence being assigned to the last note of each grouping (typically, at least in Western music, a member of the tonal hierarchy of the piece in question). She also proposes that, for listeners experienced in an idiom, the tonal hierarchy or schema will be involved in "an elaborate bootstrapping operation", not only being abstracted from the piece in question but also being applied to it in a continual process of hypothesis-generation about which notes might fulfil which particular tonal functions.

Bharucha (1984) agrees with Deutsch's bootstrapping proposal. He goes on to suggest ways in which the tonal hierarchy may be actively used in the perception of a piece of music. He differentiates between two kinds of hierarchical representations of musical stability in cognition; the event hierarchy, or relative stabilities within the string of musical events that constitutes a piece of music, and the tonal hierarchy, which is an organisation of classes of musical events. He suggests (p 422) that tonal hierarchies facilitate the generation of event hierarchies by enabling the assignation of "prominent positions in the event hierarchy to events belonging to more stable classes [in the tonal hierarchy]", basing this claim on the fact that in the experiment of Castellano et al., Western listeners responded on the basis of note duration but Indian listeners' responses displayed in addition a tacit knowledge of the tonal hierarchy (of thaats and raags).

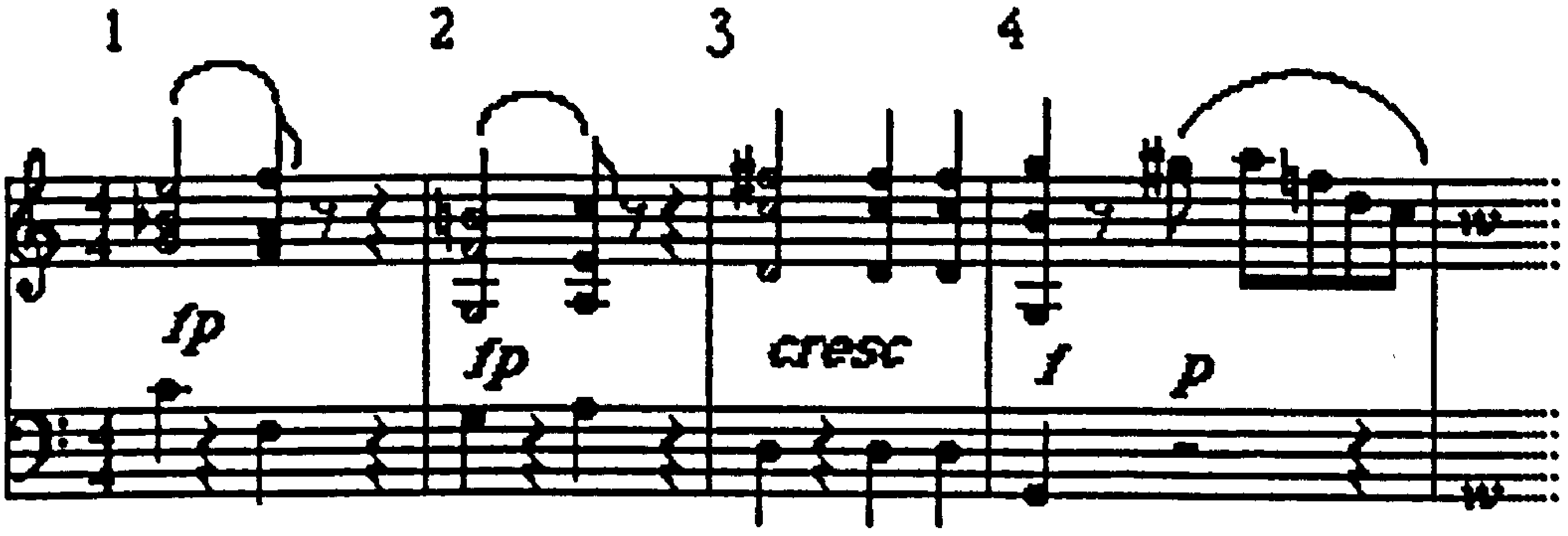
Bharucha goes on to present further evidence that pitch order (syntax) may serve to evoke a specific tonal hierarchy, showing that in the absence of durational cues an "unstable" note (one that is not a member of the tonal hierarchy in question) may be "anchored" to the tonal schema underlying the piece if a stable note which is a pitch neighbour (either diatonic or chromatic) follows it in temporal sequence. He concludes that a native listener has access to tonal hierarchies in long-term memory which, once activated, can facilitate the generation of event hierarchies in listening to a piece of music, using information contained in the temporal order and the overall duration of notes to activate the appropriate tonal hierarchy. He argues that once a tonal hierarchy is activated, the most stable event classes in the hierarchy (constituting the tonal schema) will be "most expected", thus rendering events that cannot be fit into this schema conspicuous unless they can be "anchored". The results of Bharucha and Stoeckig (1986, 1987) seem to support this claim; in reaction time tasks they found that triadic chords which would be construed as unstable within a specific tonal schema (i.e., within a context of the key of C major, the chord of F# major) are processed more slowly than are more related - and hence more expected - chords.

It appears that much of the experimental evidence on the perception of Western music can be directly related to the model proposed by Shepard for the cognitive-structural representation of musical pitch. As Shepard (1982a, p 373) states, "once notes [or chords] are categorically mapped into the discrete nodes of an internal representation that is functionally regular [by cues such as ordering within groups and the different relative sounding durations of notes], it is the structural properties inherent in the representation that are important". These significant structural properties include the fact that tones within a major diatonic key are separable from non-key tones, that closely-related keys are rotationally proximate; they also include (in the affine transforms) the depiction of the irregular intervallic structure of the diatonic scale and the possibility of representing the principal elements of the tonal hierarchy as proximate entities.

The fact that the "fit" of much of the data found in these experiments to his model (or to dimensions thereof) may be only approximate is seen by Shepard as no great obstacle; he proposes (op cit, p 383) that such findings can be explained by differentially "weighting" different dimensions of the model - e.g. giving more weight to note-similarity within the circle of fifths than within the chroma circle - so as to systematically "deform" the model to conform to the perceptions of individual listeners. Alternatively, he suggests (op cit, p 384) that these findings might be resolved by regarding them as constituting particularised "projections" of the helical model, that is, as constituting a "view" of the model as seen from particular "spatial points of view", leading to the same types of distortions of distance as would occur in the two-dimensional perspective projection of a three-dimensional scene.

The feasibility of the cognitive-structuralist model might be best investigated by applying it to a specific musical example, the "problematic" opening bars of Beethoven's Symphony No 1 in C (**Figure 4.5**). As Katz (1947, p 172) puts it, "this passage has been the subject of much controversy in the past, and still remains a problem for students of analysis and their teachers". Two questions in particular which have served as the focus for the controversy are "do the opening two chords of Beethoven's First Symphony define a key ?" and "how does C, the actual key of the work become apparent ?"

Figure 4.5: Beethoven, Symphony No 1 in C, bars 1-4



To the former question, the following answer might be offered (after Krumhansl and Kessler, 1982): a key - practically, of F major (though not necessarily identifiable as such by a listener who does not possess absolute pitch) - becomes defined in cognition because of the "provisional event hierarchy" that arises from the syntactical ordering of the two consecutive triadic chords having the roots C and F, leading to activation of an "F"-equivalent region of a listener's "key-space" representation.

An answer to the latter question might similarly be proposed, by suggesting that following this activation of the F region, a "C-ish" region is partially activated by the two triadic chords on G and A, and the G region is subsequently activated by the two triadic chords with roots D and G. Thus although the chord of C, the tonic, has not appeared in that role, its tonic function may be (fuzzily) hypothesised in bar 4 by the activation of regions of key-space that straddle it. This account rests on the listener's ability to make use of some "tonal-hierarchical" representation to infer an appropriate event-hierarchy and interestingly (in the light of the music-analytic controversy that has surrounded the passage) does not suggest that it may be encapsulated in terms of a single, completely appropriate event hierarchy; the ambiguity of the passage would be mirrored in the representations that listeners might form, either in terms of the "fuzziness" of the "C-as-tonic" hypothesis in bar 4 or in terms of the precise but dynamically changing functions assigned to different chords in the formation and discarding or confirming of key-hypotheses as the passage unfolds.

CHAPTER FIVE

Problems with the cognitive-structural model

And yet, in spite of the apparent success of the cognitive-structuralist approach in accounting for perceptions of musical pitch there are difficulties in accepting that the model and the associated experiments actually do provide an integrated theory of the cognition of musical pitch. These difficulties concern the relation of the experimental evidence to Shepard's initial model (as well as to Krumhansl's later "key-space"), the relation between experimental context sequences and both "real music" and music theory, and the theory's general neglect of certain aspects of pitch structure in proposing mechanisms for the instantiation of the tonal hierarchy in a listener's perception.

To illustrate the potential problems by means of a concrete example, consider the putative answers to the two question of key-identification above. The proposal that syntactic ordering alone serves to identify F as the initial tonic triad is questionable, in that it would seem to imply that if the order of chords had been reversed so that C followed F but their spelling and orchestration etc. remained the same, C would have been identified as the tonic. This is patently absurd, yet, of the mechanisms explicitly proposed within the cognitive-structural theory for inferring the tonal hierarchy syntax is the only one available, as frequency of occurrence cannot play a role in determining the relative stabilities of events at the very outset of the piece. Again, it could be proposed that the actual key of the piece, C, would be likely to be confirmed by the melodic configuration (an arpeggiated 9th chord on G) containing F natural in bar 4, as this configuration of notes could only occur within the key of C. The grounds for this intuition - based on knowledge of the possible notes and intervals that can occur within the key of C rather than on any syntactical or durational premises - do not seem to be captured by the rationales for key-identification that the cognitive-structural theory offers.

The problematic aspects of the cognitive-structural approach can be viewed as deriving from two sources. In the first case, the experimental literature provides results which, while consonant with conventional music theory, are simply not expressible in terms of any multi-dimensional model, far less in terms of that outlined by Shepard. In the second case, the approach adopted by the experimenters is so constrained as to leave unexplored aspects of musical pitch organisation such as within- and between-key interval structure that are incorporated in Shepard's model; at the same time the degree to which the experimental material used is representative of pitch structure in "real" tonal music is open to question.

Experimental results: lack of fit with theory

The locus of this first problem lies in the relation of the experimental evidence to the multi-dimensional, "imagistic" model. A major finding in the experiments is of some asymmetry between notes or chords presented in a "tonal" context in judgments of similarity or "fittingness as completions". Krumhansl (1979) finds these asymmetries in respect of inter-note judgments (in her words, "tones less closely related to the established tonality [e.g., in the key of C, the note C#] were judged more similar to tones more central to the tonality [C] than tones central to the tonality were judged to tones less related to the tonality"), and Bharucha and Krumhansl (1983) find a similar pattern of results in their Experiment 1 (where chords were rated as to their suitability in succeeding one another in time). Such asymmetries are referred (after Rosch, 1975) to the "prototypicality" within the context tonality of "central" notes or chords, these serving as cognitive reference points within a tonality or key; such notes or chords will be more "stable" than other notes which are either within or outwith the key in question, and will form the "apex" of the tonal hierarchy that serves to define a key.

However, these asymmetries are not capable of being expressed within Shepard's model, or indeed within any multi-dimensional model. As Tversky (1977) points out, asymmetrical judgments of similarity in any "cognitive domain" violate the constraints upon which multi-dimensional models are founded, the constraints of positivity, symmetry and triangle inequality. If the "distance" between two points (representing the adjudged similarity) within a multi-dimensional space varies depending on which point the measurement is made from (as is implicit in the asymmetrical similarity judgments found), then the space is simply incoherent: the similarity between two points is intransitive, and cannot be represented by a single metric. Although Krumhansl (1979) does manage to represent her experimental findings on the similarity relations between notes in a tonal context in terms of the conical structure described earlier it should be pointed out that this is achieved at the expense of sacrificing the asymmetries in the data. Distances in the conical model represent the average similarity ratings of two notes to each other rather than the consistently different ratings that were actually found.

This asymmetry in listeners' judgments would appear to be central to any cognitive representation of a tonal hierarchy, but it is entirely incompatible with any multi-dimensional solution. It might be, however, that this problem is more apparent than real, deriving as it does from Shepard's insistence on the imagistic nature of the cognitive representation of musical pitch; as Johnson-Laird points out (1983, Ch 7), imagistic representations can always be restated in propositional terms, while the converse does not necessarily hold. Some such restatement - perhaps in terms of the framework advanced by Tversky (1977) - might suffice to capture the types of of structural relations that Shepard's model expresses whilst providing the means to represent prototypical relations based on asymmetries. Indeed, Bharucha and Krumhansl (1983) appear to attempt this type of propositional restatement of the their experimental findings. It would also dispose of the difficulty of potential infinite regression that Shepard poses for himself when he suggests that Krumhansl's (1979) findings might be accommodated within his model by regarding them as constituting a "view" of the model as seen from a particular "spatial point of view"; if this were indeed to be the case, something uncomfortably homuncular must be doing the viewing. As will be demonstrated below, it might be better to conceive of Shepard's model as a geometrical representation embodying some of the formal properties of structural pitch relations proposed by Balzano (1980, 1982), an aspect of the model which, to be fair to Shepard, he does acknowledge (although in his view this aspect of the model always seems subsidiary to its status as an imagistic, spatial representation).

Experimental materials and their relation to theory

The second problem rests on several different issues concerning the design of the experiments and the interpretation of their results. Throughout all the experiments referred to, context sequences were used to "set" subjects appropriately for the similarity or fittingness judgments, the contexts being intended to activate some schema based on an overlearned representation of pitch relations - the tonal hierarchy. Given that a tonal hierarchy is presumed to be derived or abstracted from repeated exposure to pieces of tonal music (at least for Western listeners), the utility of the context sequences in activating a tonal hierarchy must depend on how well the contexts embody the significant structural features of pitch usage in "real" tonal music, or on the extent to - and manner in which - the event hierarchies that constitute individual context sequences embody cues as to the tonal hierarchy. The degree to which the context sequences that are used actually do embody a representative range of possible cues as to the tonal hierarchy is open to question, as is the manner in which they encapsulate significant structural features of tonal music.

Perhaps the most immediately apparent difficulty in accepting the context stimuli as representative of pitch usage in "real" tonal music lies in the objections which can be raised to the physical and psychophysical characteristics of some of the experimental stimuli used. Krumhansl and Kessler (1982), Krumhansl, Bharucha and Kessler (1982), Krumhansl, Bharucha and Castellano (1982) and Bharucha and Krumhansl (1983) all make use of the same type of stimuli in the experiments on perception of inter- and intra-key chord relatedness. Indeed, these studies are among the very few which explicitly seek to address the question of the perception and cognition of Western music in its harmonic - rather than melodic - aspect; all of the experimental studies outlined in Chapters 2 and 3 (with the exception of those of Terhardt and Parncutt) have based their accounts of the perception of pitch relations on purely melodic stimuli, with all the limitations on the generalisability of their findings that this implies.

However, the significant advance that the cognitive-structuralists' experiments seem to represent may have less substance than is claimed for it, in view of the nature of the stimuli they used. These consisted of "chords with an organlike sound, without any clearly defined lowest or highest component frequencies" (Krumhansl and Kessler, 1982, p 341). While one may sympathise with the attempt to rid chord successions of any messy voice-leadings that might be perceived by producing more-or-less "fused" entities, stimuli such as these can surely not be taken as wholly representative of chords and their implications as encountered in "real" tonal music. It is likely that chords can and will be heard as fused entities in real music; but, in general, some melodic or linear pattern will also be heard. However, this complex percept - chord and line - is legislated out of "key-space" by this attempt to use "unanalysable" chord sounds. In fact, it would appear that the cognitive-structuralists' view of what constitutes the musical entity "chord" in perception is considerably more naive than that of the psychoacousticians whom Shepard castigates for their decontextualised use of experimental material; Terhardt, after all, does suggest that both synthetic and analytic listening modes can be operational in the perception of chords.

The attempt to constrain voice-leadings - which could be regarded as "computationally-explosive" in terms of the number of voice-leadings that can be construed within even a brief succession of chords - is understandable, but the means adopted - attempting to remove voice-leadings entirely - leads to the creation and use of highly artificial stimuli in the key-space experiments. Subjectively, the stimuli appear to be almost single notes at times, with a pitch that possibly corresponds to that of the putative "fundamental frequency" of which the chord components are "harmonics" although at other times they are heard as triadic chords. It would appear that Krumhansl et al., in constraining their experimental material in this way, are attempting to use "pure" chords - stripped to their "harmonic essence" - to elicit the cognitive representation that seems to underlie the perception of harmonic structure in Western music. This, surely, is to confuse the entity "chord" as it functions in harmonic theory with the acoustical entity formed by the simultaneous sounding of different notes; in the former context a chord is a unitary element in a highly abstract symbolic representation of one theoretical domain of musical structure, while in the latter context a chord may be equally a unitary entity or a contingent confluence of melodic lines, as Dahlhaus (1980) points out[1].

Even if it is accepted that the "harmonic" forms of stimuli may be chimeric, a substantial body of experiments remains to be considered which uses melodic stimuli and appears to offer powerful support for the existence of the tonal hierarchy. Even within these experiments, however, the three-cornered relation between music theory, the cognitive-structuralists' experimental materials and conclusions, and "real" music requires further analysis. Several problematic aspects of the cognitive-structuralists' assumptions and experimental methodology will be discussed, and it will be proposed that certain of these are common to most of the previously-discussed experimental approaches to the cognitive representation of musical pitch.

<u>Tonal hierarchy - acquired by learning or acculturation ?</u>

First of all, the basis for the internalisation of the tonal hierarchy remains unclear, despite the processes suggested by Bharucha (1984) and Krumhansl (1985). These processes rely on listeners being responsive to two specific factors in forming a representation of the tonal hierarchy as they are exposed to a wider and wider range of tonal music: the differences in the relative sounding lengths of notes (duration) and in the greater

---

[1] It would be of considerable interest to use the cognitive-structuralists' chordal stimuli as inputs for Terhardt's or Parncutt's algorithm, in order to judge the possible root relations that they might have implied for subjects as well as to assess the degree to which chordal components might segregate in perception.

or lesser tendency of notes to terminate note-groups (syntax). That is, the abstraction of the tonal hierarchy is seen as the result of a process of <u>acculturation</u> rather than as arising through a <u>conscious</u> learning of the music-theoretic "functions" of different diatonic notes within a key. Evidence to support this theory is found in the fact that subjects in several of the experiments (e.g. Krumhansl and Kessler, 1982) are stated to have had no formal instruction in music theory but to have exhibited sensitivity to the tonal hierarchy in their responses.

However, in some respects this evidence is conjectural, and conflicts with the results of other cognitive-structural studies. For instance, Krumhansl and Shepard (1979) found that untrained listeners differentiated between diatonic and non-diatonic notes in their responses, but showed no consistent differentiation <u>between</u> diatonic notes (as would be expected if listeners had access to some tonal hierarchical representation), a pattern also found by Krumhansl (1979). These results might be expected in the light of Krumhansl and Keil's (1982) results on children's acquisition of the tonal hierarchy which imply that the formation of some schema based on diatonic/non-diatonic differentiation precedes any apparent awareness of the tonal hierarchy[2].

---

[2] A more recent study, that of Cuddy and Badertscher (1987), appears to show that a tonal-hierarchical representation is present in the responses of even very young children. However, the same objections as will be raised later in this Chapter may be applied to these results; for instance, their data could be interpreted as deriving from short-term memory considerations, with subjects responding on the basis of note saliency within the context sequences rather than on the basis of a long-term abstracted representation

Krumhansl's (1979) and Krumhansl and Shepard's (1979) findings could be taken as indicating that awareness of the tonal hierarchy arises as a result of formal musical training rather than being a concomitant of acculturation, a possibility that is in fact not contradicted by the results of Krumhansl and Kessler cited above. In Krumhansl and Kessler's experiments, "subjects with no formal training in music theory" had a mean number of 8.6 years formal instruction on a musical instrument. In classifying these subjects as "non-musicians", they appear to overlook the fact that even elementary instrumental instruction usually incorporates and embodies many music-theoretic concepts including those of differential degrees of stability and key-relation. Accordingly, the responses of their "non musically-trained" subjects cannot be claimed in support of the "acquisition-by-acculturation" hypothesis that underlies the model of abstraction of the tonal hierarchy based on note duration and syntax.

Separability of syntax and duration

Moreover, the experimental treatment of syntax and duration as possible cues as to the tonal hierarchy raises problems. The melodic context sequences used in the experiments by Krumhansl (1979), Krumhansl and Shepard (1979) and Krumhansl and Kessler (1982) consist of scale passages that ascend to or descend from the tonic by step, or of triads that ascend from the tonic. The notes that make up these stimuli are not presented isochronically; the tonic of whichever key they are intended to evoke generally has twice the sounding duration of any of the other notes used.

In this way, the cues that are adduced as instantiating the tonal hierarchy in its cognitive representation - duration and syntactic ordering - are present in the stimuli; the tonic will presumably have some enhanced relative salience in perception because of its longer duration as well as because of the fact that it occupies the initial or terminal position in each sequence. Thus the factors of syntax and of frequency of occurrence (or proportion of elapsed time occupied by each note) are explicitly incorporated in each melodic tonal context. The effects of each factor within each experiment, however, necessarily interact; no attempt is made to counterbalance the factors in order to assess their potential separability.

The "representativeness" of contexts

Not only the processes whereby the cognitive representation of the tonal hierarchy arises, but the actual existence of the tonal hierarchy as a fully-developed representation in long-term memory are open to question. Given that the cognitive-structuralists claim that duration and syntax are the operational factors in the acquisition, establishment or recognition of the tonal hierarchy, what must - by their account - be happening when listening to one of the context sequences ? To recap, the contexts embody cues (temporal order and overall duration) as to the relative stabilities of the notes that make it up; they thus embody cues to "event hierarchies" from which listeners may abstract tonal hierarchies. Simultaneously, in a sort of "bootstrapping" operation (Deutsch, 1984) the tonal hierarchies to which listeners have access in long-term memory will be activated by the cues and will facilitate the generation of some event hierarchies (Bharucha, 1984).

It can be assumed that as the each context sequence is played, the longest-duration note(s) and initial and terminal notes compete for perceptual stability; the listener will then judge a following note for goodness-of-fit with the sequence, or will judge the similarity of note (i) to note (ii). It is therefore likely that the adduced tonal hierarchy will be based around the longest-duration note(s) or initial and/or terminal notes. Their sequences - scales and triads - embody cues of duration and syntax as to the "tonal hierarchy" (^1-^3-^5), and they find these notes rated higher for "goodness-of-fit". That is, they find that subjects rate single notes as being better fitting or more similar to one another on the basis of a) whether or not those notes occurred in the context sequences (scalar contexts thus differentiate between diatonic and non-diatonic notes, while triadic contexts differentiate between notes of the triad and other diatonic notes), b) whether or not those notes occupied a proportionately longer timespan within the context sequences than other notes (given the anisochronic nature of the context sequences, the tonic - starting note - generally lasts the longest) and whether or not those notes occurred at the end (or perhaps at the start) of the context sequences. Again, the tonic usually ends and starts scalar sequences while the dominant may end triadic sequences; these are thus more likely to be highly rated than are other notes. In this way it can be seen that the "event hierarchies" that may be derived from the melodic contexts are functionally similar to the putative "tonal hierarchy". As this is the case, it could be argued that listeners' responses are being shaped by <u>short-term</u> memory processes rather than being determined by some schematic representation in long-term memory, as Butler (1989) suggests.

The cognitive-structuralists could, however, argue that their context sequences are fully representative of the significant features of pitch structure in tonal music, and are thus capable of instantiating schemata that listeners have abstracted from long-term exposure to tonal music. Nevertheless their sequences are all very similar and highly constrained in melodic shape and manner of melodic progression; they cannot be said to constitute a fully representative sample of the figurations that may occur in tonal music. While one obviously finds examples of such passages as ascending and descending major and minor scales and major triads in tonal music, these are typically ornamental or figurational rather than being structurally integral.

The only experimental results which seem unequivocally to indicate a representation of a tonal hierarchy in long-term memory are those of Castellano et al. (1984). The rating patterns that they found seem to indicate that native Indian listeners do have access to some long-term representation of a tonal hierarchy, and that this is positively a long-term representation rather than a short-term effect of the context sequences used.

However, several caveats can be entered against taking these results to indicate that something of the same is necessary going on in the case of the western-musical studies. The contexts used in Castellano et al. were more extensive than those used in the western studies, and were taken from "real" music (rather than being intended to be in themselves embodiments of theoretical abstractions). As Dowling (1984) comments, of all the contexts used in the cognitive-structural experiments, those made use of by Castellano, et al. appear to be the most satisfactory as contexts, perhaps because of their extensiveness as well as because of their existence as compositions and themes in "real" music (although Dowling (1984), however, is probably wrong in equating thaat with scale or tuning system, and Bharucha is right in consigning it to mode; the _jati_ would be the scale-equivalent level of pitch organisation in Indian music). Moreover, the theoretical principles and practical usages of pitch organisation in Indian music are _not_ the same as those underlying western tonal music; some of the apparent similarities between western and Indian music theory - at least in the form outlined by Bhatkhande in the 1930s - may, in fact, be due to a conscious attempt on the part of the Indian theorists to produce a theory modelled on at least the manner of presentation of western music theory. Finally, it is conceivable that the non-transposable nature of the derivation of the different thaats may enable differentiations between thaats on the basis of pitch structure to be predicated on absolute pitch organisation rather than on relative pitch. In this way, for example, absolute pitches and their disposition within the [fixed] octave or tessitura may play a role in Indian music in determining the function of each note which relative pitch structure is called on to do in western music.

In respect of the experiments conducted using "Western" melodic contexts, it can thus be argued that the "tonal hierarchy" found is simply an artifact of the particular class of "event hierarchies" that these context sequences embody. If the context sequences had contained cues as to other "event hierarchies", it may be imagined that the "tonal hierarchy" that resulted would have been quite different. By this argument, the results that indicate the existence in long-term memory of a representation of the tonal hierarchy are simply artifactual and occur because listeners are in fact using cues of duration and syntax to assign different "stabilities" to notes in working or short-term memory.

The "informational" approach - objections

A further problem arises with one of the factors held to instantiate the tonal hierarchy - the relative sounding durations of notes within a piece or pieces of tonal music. It can be argued that a quantification of these different sounding durations - in essence, equivalent to the production of a "weighted scale" (see Nettl, 1977) - does not suffice to establish a basis for the cognitive organisation of musical pitch. This sort of "note count" will correlate with a measure of the "redundancy" of different notes within a given musical context. Hughes (1977), in the chapter which Krumhansl (1985) cites as providing support for the "differential-duration" basis of the tonal hierarchy, explicitly refers to this fact in his information theory-based analysis of Schubert's "Moment Musicale", Op 94. Although in Bharucha (1984, after Deutsch) this focus on redundancy as a means of differentiating between notes of the "tonal hierarchy" and other notes is broadened by the inclusion of considerations of note orderings or syntax (and hence, implicitly, of consideration of inter-note transitional probabilities), note "redundancy" is a key factor in the cognitive-structuralists' theories: the higher the "redundancy" of a given note in a musical context, the more "stable" or referential" it is likely to appear in perception.

However, such purely "informational" measures as the redundancy of single notes in a given context are of only limited use in determining a basis for the perceptual encoding of serial structure, or in evaluating the criteria for judgments about aspects of that structure in cognition. As Simon (1972, p 381) points out in his critique of the model proposed by Vitz and Todd (1969) for cognitive encoding of serial structure, an "informational" approach to the encoding of sequential structure in cognition is incapable of capturing the full complexity of the processes and representations involved. That is, it is not a good indicator of the method whereby a pattern might be encoded, and hence is not a good indicator of referential value - or stability - of different elements within the pattern. A major factor which must be taken into account (op. cit., p 371) is the fact that "anything recognizable by a subject ... as the result of previous training or experience is assumed to be codable into a symbol...the available coding alphabet then consists of the set of such symbols...stored in long-term memory". Only then can any measures based on information theory usefully be applied in order to ascertain degrees of pattern complexity or to estimate the likely referential status of elements or symbols in respect of some particular coding alphabet.

It could be argued, on the basis of the results presented in Krumhansl and Shepard (1979), Krumhansl (1979) and Krumhansl and Keil (1982), that the internalisation of an appropriate musical "alphabet" precedes the establishment of the tonal hierarchy. As stated above, these results show that musically untrained listeners differentiated between diatonic and non-diatonic notes and that children differentiate between diatonic and non-diatonic notes before they acquire the ability to differentiate among these notes in "targeting" those of the tonal hierarchy. The appropriate musical alphabet could thus be something akin to the diatonic scale. The nature and structure of that alphabet, and hence the potential significance of elements and relations within that alphabet in the establishment of a tonal hierarchy would need to be assessed. The question can be restated as one which concerns the relationship between the tonal hierarchy and the putative coding alphabet. In particular, it would appear necessary to examine the ways in which diatonic structure is characterised and manipulated in the cognitive-structural experiments so as to ascertain whether or not these experiments shed light on the role - if any - played by diatonic intervallic structure in the abstraction of the tonal hierarchy.

The neglect of scale structure

The cognitive-structuralists do not entirely ignore this question; Krumhansl and Keil assert, in their discussion of children's acquisition of the tonal hierarchy (1982, p 249), that "once the possible [diatonic] pitches of the tonal system were isolated, finer differentiation was made within the set of scale pitches". However, no consideration is given in the further cognitive-structuralist experiments to how this "isolation" of diatonic pitches might bear on the ways in which "finer differentiation was made within the set of scale pitches" in the establishment of the tonal hierarchy. That is, no explicit account is taken in their theories of how the structure of the diatonic scale might serve as a cue in the abstraction of the tonal hierarchy from the context sequences.

For example, Krumhansl (1985, p 372) states that each context sequence used was a "musical unit, such as a scale...that unambiguously established a major or minor key". In using the term "scale" in this way, Krumhansl appears to be conceiving of key and scale in a way that is more consonant with a pedagogical music-theoretic usage rather than with Powers' and others' more specific and generalisable framing of these terms. While it appears that the cognitive-structuralists are attempting again - as is the case with the "harmonic" stimuli - to present the stripped-down "essence" of scale and key as experimental contexts, scale here seems to have the meaning that it has for musicians within a pedagogical rather than a theoretical framework. As Butler (1988) points out, the cognitive-structuralists' "scale" contexts are highly specific pieces of "passage-work" rather than presentations of the abstract structural features - essentially, possible formable intervals - that characterise the theoretical definition of scale.

In order to test the nature of the relationship in cognition between diatonic structure and the tonal hierarchy, a wider range of note patterns than were actually used in the experiments should have been used; for instance, the experimenters could have adopted Deutsch's suggestion (1982, p 287, where she says that "it is to be expected that different contexts would give rise to different patterns of similarity relationships using these techniques") and presented several different step-wise patterns which started and finished equally on all possible scale notes. Otherwise the possible effects of <u>diatonic scalar interval structure</u> - which constitutes a very powerful theoretical means whereby scale-notes may be differentiated one from another - are completely confounded in these experiments with the possible effects of duration and syntax.

It should be recalled that the depiction of asymmetric diatonic scalar interval structure - in particular, depiction of the separability of tones within a major diatonic key from non-key tones as well as of the fact that each note of such a scale forms a unique set of intervals with all other scale notes - is a feature which Shepard (1982a, 1982b) claims to be of considerable significance within his model. It may be that the tonal hierarchy and diatonic asymmetricality are entirely separable, independent factors. If this is the case, then the context sequences are acceptable. However, the relationship remains untested because the context sequences are not explicitly designed to factor out effects of duration, syntax and interval structure.

If, for example, the sequences had consisted of the notes of the diatonic scale laid out in a different temporal order (e.g. starting on the supertonic so as to outline the Dorian mode) and the initial tonic hierarchy (tonic-dominant-mediant etc; was still found to be present in listeners' judgments, this could indicate that something independent of syntax and duration was acting to determine relative stability. As the only invariant cue would be interval structure, this would argue for this structure playing some role in establishing the tonal hierarchy. It could be that, given the notes of the diatonic scale, listeners were capable of abstracting an overlearned tonal hierarchy that remains constant because of the invariance of the structure of the diatonic scale even when re-ordered.

If the re-ordering resulted in a primacy for the supertonic and its triad, it could indicate that the "new" tonal hierarchy is predicated on its "embodiment" in the new type of context sequence. This is the conclusion at which Krumhansl and Kessler (1982) arrive in presenting the results of their study using minor-scale contexts; they found that listeners appeared to abstract a tonal hierarchy that differed from that found when major-scale contexts were used. Alternatively, the results could indicate that the tonal hierarchies found in the experiments are derived principally from syntax and durations within the context sequences. In this latter case, the new tonal hierarchy found would be a short-term effect specific to the melodic configuration used and not generalisable to a wider sample of tonal melodic configurations. However, the issue cannot be resolved one way or the other; as stated above, the context sequences are not designed to factor out possible effects of duration, syntax and interval structure.

As stated, the possible effects of diatonic interval structure in the perception of tonal relations are given considerable weight by Shepard. He states (1982a, p 378) that "As has been emphasized [by music-theorists] it is the property of each diatonic scale degree having a unique relation to all other tones in that scale that confers on each scale degree its unique 'dynamic quality' or 'tonal function'". It is to expected, then, that such possible features of diatonic interval structure would be taken into consideration in exploring the nature of the tonal hierarchy's representation in cognition. And yet, as has been shown, the cognitive-structuralists do not take this factor into account in their experiments.

Problems with experimental approaches to musical pitch cognition

In fact, the nature of diatonic scale structure and the question of its relationship to the tonal hierarchy has not been explored in depth in any of the previously-discussed experimental approaches to the cognitive representation of musical pitch, although the differentiation between scale and the tonal hierarchy has been made. As stated previously, Cuddy in her earlier papers (e.g. Cuddy, Cohen and Mewhort, 1981) takes an essentially pragmatic approach to this question, using the concepts of scale and key much as they appear in traditional music theory and focusing primarily on characteristics of temporal structure. Although in a more recent paper based on the cognitive-structural paradigm (Cuddy and Badertscher, 1987) she initially differentiates between scale and key in terms that are clearly structural, she subsequently confuses the issue by relying on psychoacoustical and music-theoretic premises (e.g. Patterson, 1986: Schenker, 1906/1954) to argue for the triad as the "progenitor" of the diatonic scale. Deutsch (in Deutsch and Feroe, 1981, and Deutsch, 1982) again differentiates between the cognitive representation of diatonic structure and the triadic components of the tonal hierarchy in treating these as different representational alphabets; however, she offers no means of relating these alphabets (other than noting that one will contain the other).

Of all the previously-discussed experimenters, Dowling (1978) presents perhaps the most coherent scheme for the cognitive representation of musical pitch. To recap, his scheme has four levels, the psychophysical pitch function, the tonal material, tuning system and mode; he refers to diatonic scale-structure and the tonal hierarchy as, respectively, the level of tuning system and the level of mode. Unfortunately, Dowling is not particularly specific about the role that these different levels might play in musical cognition, nor does he indicate how these representations might arise. He does state (Dowling, 1984, p 418) that most studies have neither found nor aimed to test a possible representation of pitch at the tuning system level of organisation in cognition, having focused on representation at the modal level; the implication seems present that both levels of representation are inseparable and arise simultaneously in the acquisition of pitch schemata. This, as has been shown, conflicts with the evidence of Krumhansl (1979), Krumhansl and Shepard (1979) and Krumhansl and Keil (1982) for diatonic rather than tonal differentiation in the responses of some subjects; that is, the tuning system level would appear to have some cognitive reality, even if it is not separable from mode.

To summarise, all of the experimental approaches to the cognitive representation of musical pitch that have been discussed leave the question of the possible significance of diatonic scale-structure in cognition unexplored. Despite the consistency of the results that the cognitive-structuralists present, and notwithstanding the coherence of their interpretations of those results, they also appear to overlook such considerations. This consistency and coherence may in part be due to the relatively formal nature of the theories of pitch organisation that form the basis of at least their initial studies, those proposed by Shepard. His model offers a lucid means of characterising both the structure of the diatonic scale and of tonal relations within the tonal hierarchy, as well as demonstrating an admittedly abstract relationship between diatonic structure and the tonal hierarchy via the affine transformation of the double helix that results in "thirds space".

As outlined above, there are problems with Shepard's model (as, for example, in its lack of fit with some of the experimental results). However, not least among these problems is the fact that it appears to be more confirmatory than explanatory, and thus unlikely to be able to comply with Erickson's (1982) proposal that "the findings of cognitive psychology ...should be the substrate upon which theories of various aspects of music can be erected". That is, Shepard's derivation of his model is explicitly founded on music-theoretic schemes of pitch relations and neither the empirical studies nor his theoretical expositions reveal or propound a coherent extra-musical rationale for the structure of the cognitive representation; in this respect, the cognitive-structural approach seems to result in an incomplete psychological model.

In fact, the significant musical-structural properties displayed by Shepard's model would appear to be no more than the logical consequences of his acceptance of complete enharmonic equivalence between flats and sharps and the incorporation into the model of the circle-of-fifths. That is, features such as the capacity to display the diatonic scale as a connected region, diatonic "asymmetricality" and the depiction of the differential proximity of keys in "key-space" appear to derive firstly from Shepard's treatment of notes such as Eb and D# as identical and secondly from the particular proximities displayed by notes and groups of notes within the resulting "equitempered" circle of fifths. The fact that such "special properties" necessarily follow from these initial assumptions is not particularly explicit in Shepard's exposition of his model; that this is indeed the case becomes evident from consideration of Balzano's group-theoretic account of musical pitch (Balzano, 1980, 1982). Despite the group-theoretic representation's omission of certain features of musical pitch that are definitely operational in music (such as the pitch-height domain), it has advantages over Shepard's model as a basis for the exploration of musical pitch cognition; it is formally definable and (to a certain extent) empirically neutral; relationships within and between its constituent parts (pc sets and collections of pc sets) and between its whole and its parts can be completely formally described, and clearly formulated hypotheses and predictions about the perceptual status of these relationships can be empirically tested.

The group-theoretic representation of musical pitch

One of Balzano's primary concerns in treating musical pitch in group-theoretic terms is to elucidate the special properties that inhere in the sets of pitches (numbers) analogous to diatonic scales within this representation. He suggests (1982) that the fact that these properties are intrinsic to the group-theoretic representation of pitch constitutes a strong argument for assuming that the cognitive representation of musical pitch embodies group-theoretic principles to some extent. Balzano (1980) treats the notes of the chromatic scale - or the chromatic set - as analogous to a cyclic group of order 12, taking octave-related notes as equivalent and differentiating between notes solely on the basis of chroma. This enables pitches - or more properly, pitch-classes - to be represented numerically, as in **Figure 5.1**, analogous to the chroma circle.

This representation has the properties of complete group representation and interval equality. It can be produced by selecting an identity element (0) and an operation (+1) and applying the operation recursively on the identity element so as to generate further set members. All 12 possible pitch-classes (henceforth **pcs**) available within mod 12 arithmetic will thus be produced, and the differences or intervals between adjacent pcs are all equal (in mod 12 arithmetic, only 1 step separates 11 and 0 as 0 is equivalent to 12). The only other possible representation of elements of the cyclic group of order 12 which retains these properties is shown in **Figure 5.2**, and is analogous to the circle-of-fifths, each pair of adjacent notes being separated by 7-mod-12 minimum intervals (semitones). The two representations are thus homomorphically equivalent; that is, mod 12 relations holding between elements in one representation also hold in the other representation.

These relations are founded on the idea that sets of pcs can be viewed as equivalent; for example, the set of pcs (0,1,2) is equivalent to the set (1,2,3) in that the same interval

Figure 5.1:The chroma circle - pitch class numbers and note names.

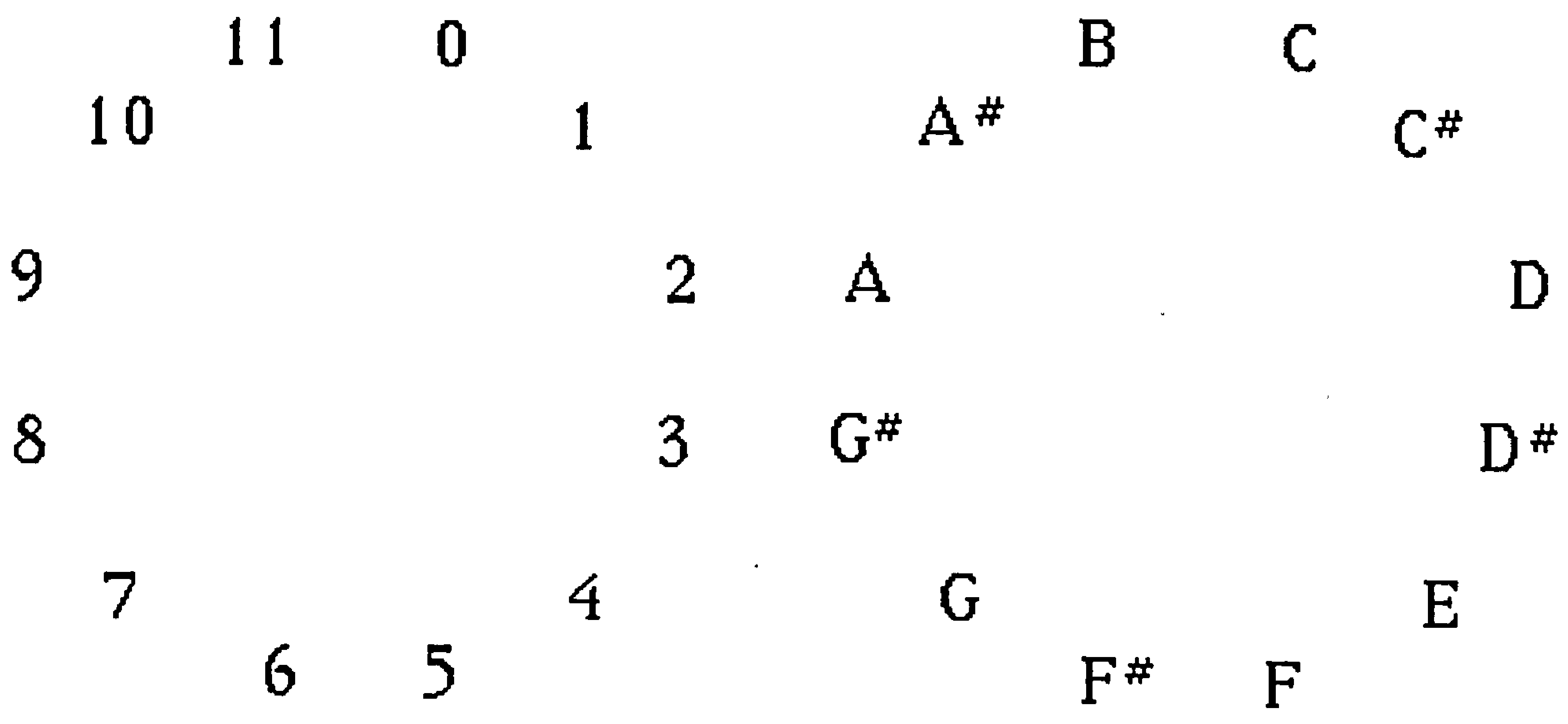
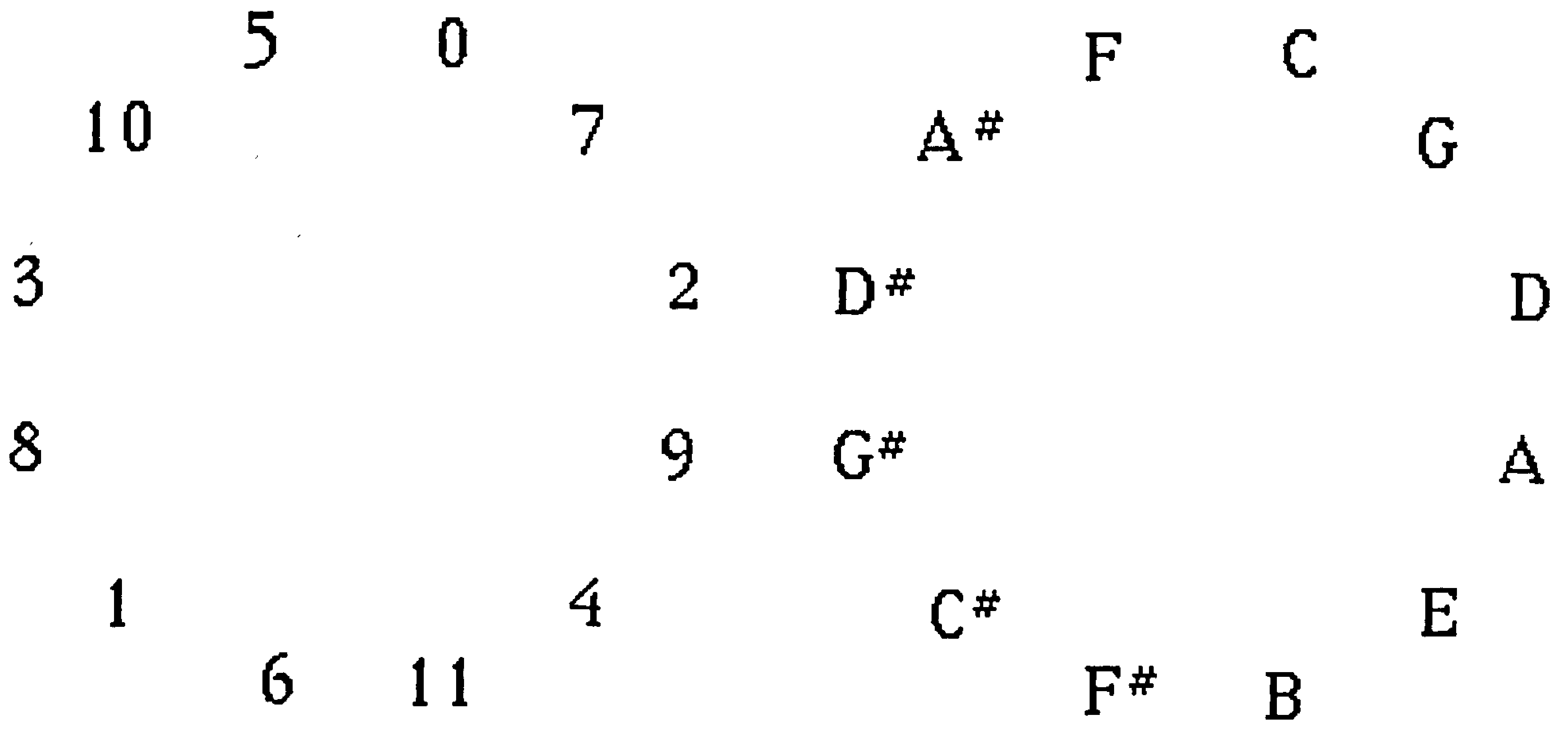


Figure 5.2:The circle-of-fifths - pitch class numbers and note names.



relationships are formable within each set, and one set may be mapped onto the other by one simple operation, "transposition" (consisting of the addition or subtraction of the same mod 12 number to or from all members of the pc set). Thus pc sets having the same number of elements and the same layout on the chroma circle or circle-of-fifths are equivalent under transposition. The term has here a subtly different meaning from its musical usage, whereby a series of intervals performed at two different pitch levels can be said to the same melody, the one being a transposition of the other; transpositional equivalence between pc sets is order independent, so that the pc set (0,1,2) would still be transpositionally equivalent to (1,2,3) even if its elements were presented in some other order. That is, a pc set will retain its identity no matter the order of occurrence of its constituent elements.

The total number of different pc sets that may be formed from the chromatic set is 4095; this is reduced to 351 by treating transpositionally-related sets as the "same" pc set. From this range of 351 non-transpositionally-related pc sets, Balzano picks out the set (0,2,4,5,7,9,11) - which corresponds in structure to the diatonic scale, and will be referred to as the <u>diatonic set</u> - as having particular and singular characteristics. First, it has the property of <u>uniqueness</u>; each pc can be differentiated from every other pc by the set of intervals it forms with each other pc of the set. Although this property is common to a few pc sets, the diatonic set possesses not only uniqueness but <u>coherence</u>. Coherence means that the sum of any two intervals formed between three adjacent pcs of the set in normal order (roughly speaking, arranged so that pc numbers increase from left to right, with the smallest intervals between adjacent pcs grouped to the left in the set description; see Forte, 1973) will be greater than any single interval occurring between adjacent pcs in a pc set so ordered. Thus although the diatonic set has unequal-sized intervals or scale-steps between adjacent pcs (as it must, to possess uniqueness), the sum of any two consecutive scale-step intervals is always larger than any single scale-step interval. Coherence and uniqueness can be thought of as conferring possible advantages in perception, in that the notes of a diatonic melody will be differentiable simply by the intervals that they form with other notes irrespective of their order of occurrence, and any major or minor interval within the melody will be constituted of the same number of scale steps, enabling a listener to judge the size of movement that it constitutes within the underlying scale (with the one exception of the tritone).

In addition to coherence and uniqueness, the diatonic scale has the property of simplicity. This is the property that each diatonic set may be transformed into another such set by changing only one element, the relation between the initial set and the transforming element being the same for all equivalent forms of the diatonic set. This can be seen most simply in **Figure 5.3**, corresponding to the circle-of-fifths. Given that seven adjacent pcs here form the diatonic set, any such set can be transformed (in terms of pc content) into an equivalent set simply by shifting it round the circle. For example, the set (5,0,7,2,9,4,11) is transformed into an equivalent set by replacing 5 with 6 to produce (0,7,2,9,4,11,6). All diatonic

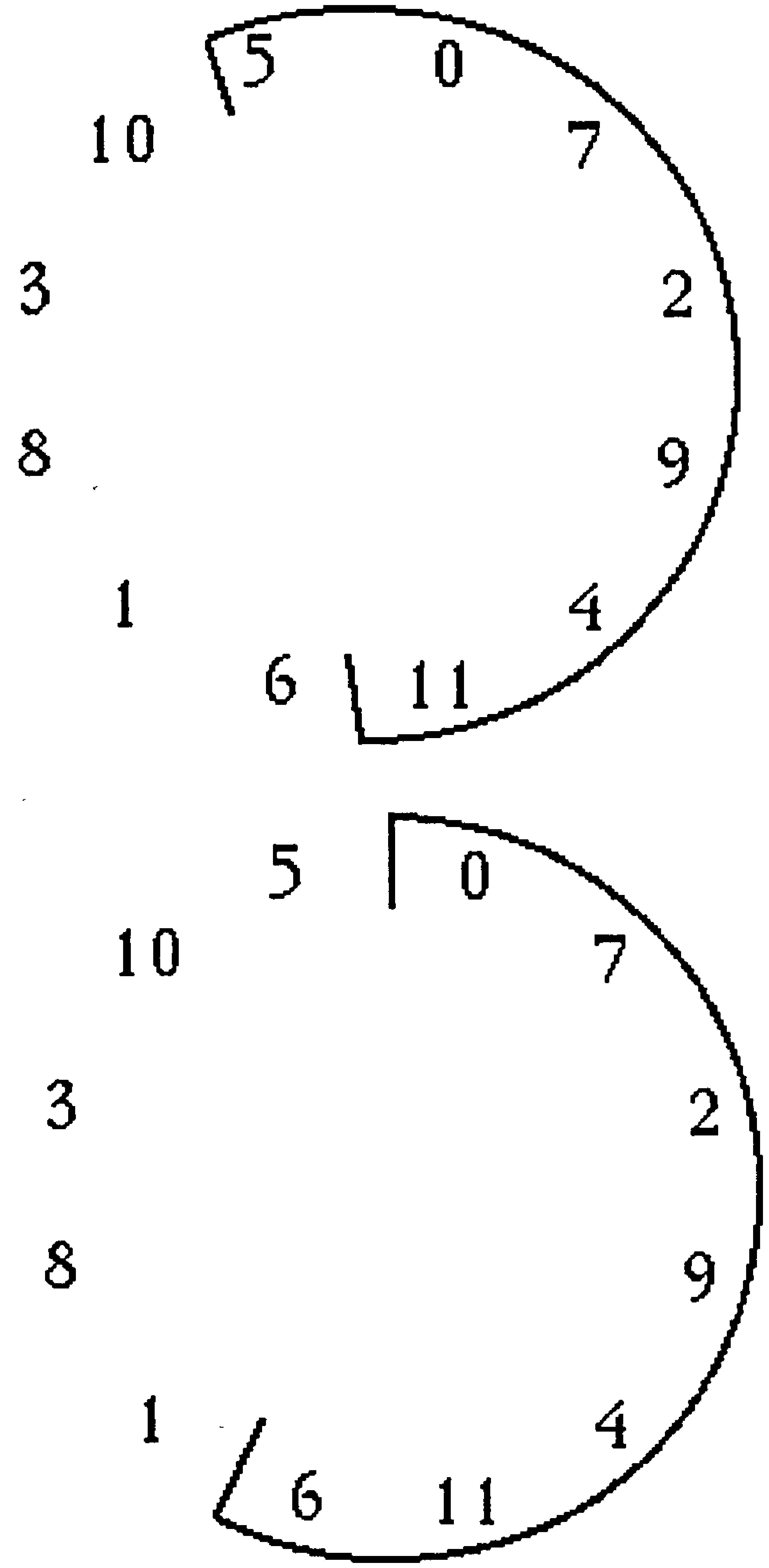


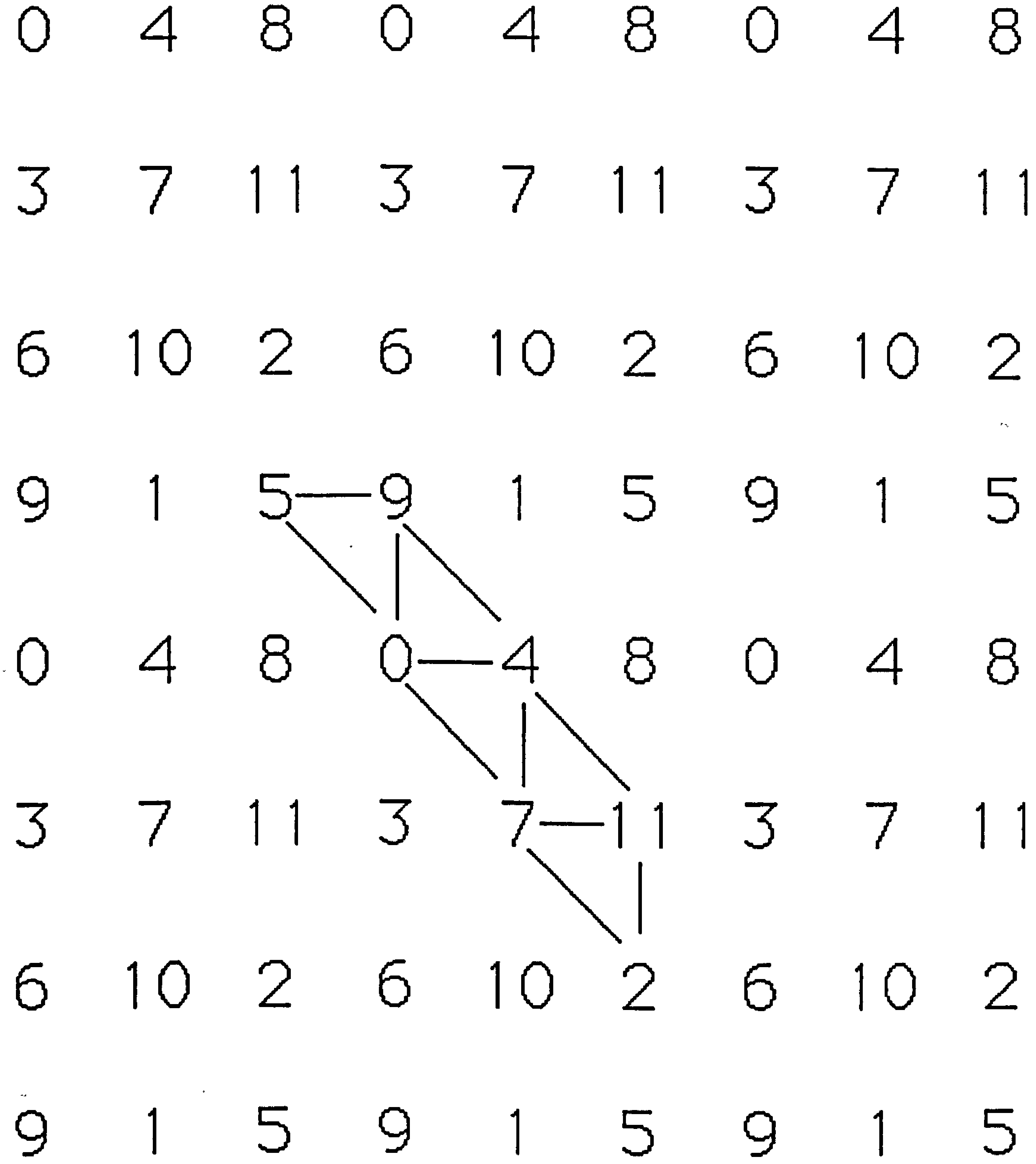
Figure 5.3: Simplicity relationships between diatonic scales within the circle of fifths.

major scales have similarly adjacent neighbours within which they share all but one of their pc content. Only one other pc set of cardinality 7, the pc set (0,1,2,3,4,5,6) has the property of simplicity, and although it has the property of uniqueness it lacks that of coherence.

Balzano (1980) uses one other isomorphism of the cyclic group of order 12 to produce a representation in which elements separated by intervals of 3 and 4 (corresponding, in the pitch class domain, to intervals of a m3 and a M3 respectively) are adjacent; in other words, major triads are represented as proximate and compact entities. The isomorphic representation that Balzano uses is the direct-product group, in which two operations, +3 and +4, are applied to a reference element to generate new set members; this produces the representation shown in **Figure 5.4**, which is directly analogous to Shepard's (1982a) thirds-space, to a topologically-transformed version of Longuet-Higgins' (1979) "representational matrix for tonal music" and to a simple transform of Krumhansl's (1985) "harmony-space". This representation is formally equivalent - isomorphic - to the two forms described above (the chroma circle and the circle of fifths), yet captures concisely the fact that the diatonic scale may be represented in terms of a set of interlocking major and minor triads.

Balzano (1980,1982) derives the three properties of the diatonic set from its representation within the cyclic group of order 12, in which equivalence between pc sets is based solely on set cardinality (number of members) and transposition. However, in contemporary music theory (see Rahn, 1980: Forte, 1973) a more

Figure 5.4:The direct-product group representation of the chromatic set. The operator in the X-domain is 4, while that in the Y-domain is 3.



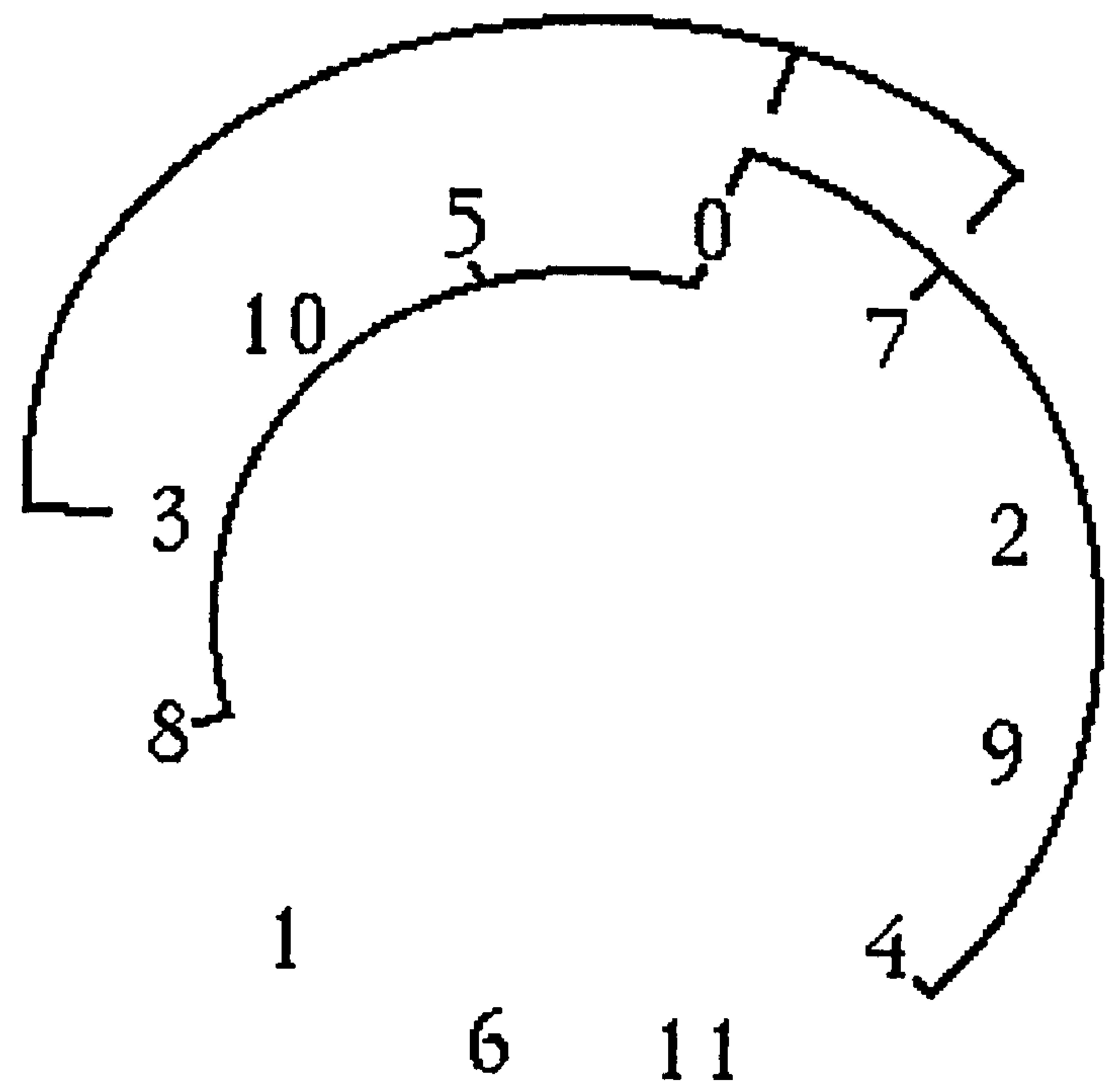
0	4	8	0	4	8	0	4	8
3	7	11	3	7	11	3	7	11
6	10	2	6	10	2	6	10	2
9	1	5	9	1	5	9	1	5
0	4	8	0	4	8	0	4	8
3	7	11	3	7	11	3	7	11
6	10	2	6	10	2	6	10	2
9	1	5	9	1	5	9	1	5

widespread group-theoretic approach makes use of the dihedral group of order 24 (Balzano, 1982) to represent pitch organisation, using equivalence relations founded on cardinality, transposition and inversion. Inversional equivalence is best defined as complementary equivalence within mod 12 arithmetic; the inversion of pc (8) would be (4), while the inversion of the pc set (0,7) would be (0,5), this last relation being analogous to the musical statement that a perfect fifth has as its inversion a perfect fourth. To say that two sets are equivalent under transposition and inversion is to say that two sets would "map" onto one another if reflected around a diameter of the chroma circle or circle-of-fifths and shifted round the circle to fit. The set (0,4,7) can be inverted to give the set (0,8,5), of which the normal order would be firstly (5,8,0) and then (0,3,7) (see **Figure 5.5**); thus the sets (0,4,7) - analogous to the major triad - and (0,3,7) - the minor triad - are equivalent, according to Forte (1973), as subsets of the dihedral group of order 24.

The unexplored question

Within Balzano's formalisation, the pc set equivalent to the diatonic scale has some properties which would seem to be of potential perceptual significance; in justification of his multidimensional model, Shepard (1982a) explicitly points out that these properties should, for example, enable any single scale-note to be identified in perception on the basis of the intervals that it forms with other scale-notes, irrespective of the order in which these occur - that is, to be identified on the basis of the particular alphabetic structure of the diatonic scale. However, in the empirical studies by Krumhansl, Bharucha et al.,

Figure 5.5:Inversional equivalence of set (0,4,7) to set (0,3,7).



such properties are not called on to account for ways in which an internalised scale schema plays a role in the perception of musical pitch. This is despite statements such as (Krumhansl and Keil, 1982, p 249) "once the possible [diatonic] pitches of the tonal system were isolated, finer differentiation was made within the set of scale pitches" and (Krumhansl and Keil, 1982, p 250) "it is essential to first develop an abstract internal framework that matches the most frequently occurring patterns of the scale..once the framework is established, it is used to disambiguate the functions of the individual tones within the musical context". Indeed, as has been seen, no attempt is made to control for the possible effects in perception of the structure of the underlying "tuning system" or scale in designing the context sequences used in the cognitive-structuralist experiments, despite the evidence that the diatonic/non-diatonic distinction may well precede the development of sensitivity to the tonal hierarchy.

If the diatonic scale structure is the "framework" within which notes can be differentiated, then the implications of scale-structure for the cognitive representation of musical pitch, and the contribution of that particular framework to the establishment of a hierarchy of tonal functions must be investigated (i.e. the contribution in terms of both the initial acquisition and the inference of the appropriate event hierarchy within a piece of music). As has been shown, the cognitive-structuralist experiments do not explore these implications; the experiments can only be said to test the effects of duration and syntax, conflating alphabetic scale-structure and melodic temporal structure and concentrating thereby on effects of temporal structure at the expense of alphabetic structure. So the question of whether the structural properties of the diatonic scale actually do play any role in cognition remains to be answered. The question may be restated as "is alphabetic diatonic structure actually privileged in cognition ?" (given that most studies that hint at this conflate alphabetic and temporal structure). In a sense, this seems to raise the question of whether Dowling's "tuning system" has any psychological reality. To answer this question it would appear necessary to test whether or not the responses of listeners exhibit a sensitivity to diatonic scale structure in melodies that can be said to be independent of any cues that syntax and duration may provide to the creation of an event hierarchy.

CHAPTER SIX

Experimental Series 1

The first problem that has to be confronted in testing this hypothesis is how to construct such melodies, as almost all "real" music that conforms to diatonic scale structure is likely to incorporate modal cues. According to Krumhansl and Bharucha, these modal cues as to the different referential stability of notes are likely to derive from the different sounding durations of notes (or probability of occurrence of notes), consistent syntactic structure (or different probability of occurrence of ordered sets of notes). They could also depend on some more "global" features; for example, Divenyi and Hirsh (1978) showed that notes constituting the "pitch boundaries" (extremes of tessitura) of a melodic figure were likely to be perceived as more salient - and hence, perhaps, more likely to act as reference elements - than other notes in the melody. Factors such as these are likely to constitute extrinsic cues to the "event hierarchy" or modal/tonal structure within a melodic sequence.

These factors are largely dependent on some degree of repetition, whether of individual notes or of ordered sets of notes. That is, they are dependent on the different redundancies, or, conversely, on the probabilities of occurrence, of the different components of a melodic sequence. In order to eliminate these yet still conform to diatonic structure it would appear sufficient to produce notes randomly with intervals conforming to those that are formable within a diatonic scale. This should act against the formation of any particular set of modal characteristics. Sequences thus produced would conform to a scale structure but not to any single consistent and specifiable mode, as there would be no extrinsic basis for assigning a given note more centrality than any other. However, defining "diatonicism" solely with respect to the notes of one diatonic scale provides an "all-or-nothing" definition with little scope for experimental manipulation of the stimulus material as well as causing difficulties in creating and constraining experimental material that is _non_-diatonic yet non-modal. Moreover, such a strategy would require a vast number of experimental sequences to be generated and tested in order to ensure that the experimental melodies did not embody non-obvious modal cues (such as might arise from the application of pseudo-random constructional procedures). In addition, any melodic sequence of more than seven notes using notes solely drawn randomly from one diatonic scale must contain repeated notes; this necessary repetition might confer modal attributes on the sequence.

As what is to be tested is the listeners' sensitivity to _diatonic intervallic structure in general_, the experimental material must not be tied to any one specific manifestation of diatonic structure, i.e., not tied to any one seven-note diatonic collection. Some more powerful means of constraining the production of experimental sequences is obviously required. The solution adopted here is to produce melodic sequences that vary in the number of consecutive notes which can come from any one diatonic scale; this results in the production of sequences of which the _interval structure_ approximates in an ordered way to that of the diatonic scale or set.

A further consideration in testing the above hypothesis is the nature of the experimental response task. Most experiments on the cognitive organisation of musical pitch have used recognition-based paradigms. However, the cognitive-structural approach - which, for all its failings, appears to be the most successful approach yet in delineating the "fine detail" of musical pitch representation - seems to owe its apparent success to its use of contextualised rating tasks. A rating task could prove useful in testing the current hypothesis; if diatonic scale structure does serve as a privileged set of relationships (Posner, 1973; Reed, 1978) for adult Western listeners, the degree of conformance of a sequence of notes to diatonic structure could serve as a basis for ratings of aesthetic preference (Garner, 1970, and Krumhansl, 1979).

This could stem from a sense of familiarity (Zajonc, 1980), as there is some evidence that people can abstract typical structures from components of structures previously perceived and regard these as familiar even though they have never seen them before (as in Solso and McCarthy, 1981). Note that diatonic structure as embodied in the diatonic scale would be a form of typical structure that has never been experienced and can never be observed directly but only abstracted from sequences of notes in particular combinations. It must be stressed that the stepwise progression of notes often referred to as a "scale" - going from tonic to tonic up the octave - is a particularisation of the scale and strongly implies a particular mode with the tonic being well-defined as starting and end points. Apart from familiarity, another basis for aesthetic judgments related to scale structure per se may be the redundancies, symmetries or "informationimparting" properties inherent in the diatonic system (Balzano, 1980, 1982). Whatever the basis, ratings of preference would seem to be valuable in determining how far subjects are able to derive "sense" from a series of notes. Though aesthetic preference is itself a highly complex and unstable phenomenon, it has appeal as a task domain in the context of an experiment involving music in approximating to a form of behaviour generally associable with real-life musical experiences (differing, in the experimental context, largely in the explicitness of the evaluation required). Preference would here be regarded as primarily dependent on stimulus structure and probably also on degree of relevant experience or training of the subjects (Munsinger and Kessen, 1966).

Throughout the experiments to be described in Chapters Six, Seven and Eight, the emphasis will be upon using subjects who have received little or no formal musical training. This is because it is likely that the categories and concepts of western music theory - which serve as a substantive element of formal Western musical training - are to some extent active in the perceptions of formally-trained western musicians. Any investigation of the cognitive organisation of musical pitch as displayed by trained musicians thus runs the risk of being confirmatory rather than explanatory in respect of music-theoretic ideas; the results of such studies cannot be used in the way that Erickson (1982) suggests, when he states that "the findings of cognitive psychology should be the substrate upon which theories of various aspects of music can be erected". It would seem more appropriate to examine whether or not subjects with no formal musical training can display evidence of of some coherent cognitive representation of musical pitch structure and if so, what form is taken by such a representation. The use of non-musically trained subjects would thus be conditioned by the idea that while the processes and products of musical acculturation are typically unamenable to conscious introspection (contrary to those of formal musical training) they are likely to reflect the functioning of very general cognitive principles and thus constitute a major constraint on the ways in which even highly-trained musicians organise their perceptions of music. Thus the few experiments described in the succeeding Chapters which make use of trained musicians do so in order to test the generality of the cognitive principles that appear to emerge from the experimental results.

The adoption of this approach means that it is likely that any coherent experimental findings will have positive value for music theory. If non-musicians can behave in a coherent fashion in respect of material ordered according to simple and specifiable premises - and if these specific behaviours are also exhibited by trained musicians - then it is possible to claim that the discovery of principles that arise from acculturation is being approached, and that these principles are to some extent meta-theoretical (or at least are beyond the descriptive power of traditional music theory) and reflect the operation of general cognitive processes. The principles may, however, be capable of absorption into music theory.

To return to the question which ended the previous section, now restated in an empirically accessible form within the terms of the preceding discussion: how far might conformance to diatonic structure (independently of mode and any additional symmetries and redundancies) influence aesthetic preferences for melodic sequences ? The question is now addressed by generating sequences of notes that conform to diatonic structure in quantifiably different degrees (but which are random in other respects), and by investigating the aesthetic preferences of groups of subjects for these sequences. Before describing the experiments, it is necessary to describe the sequence generation procedures in some detail.

Procedures for generating experimental melodic sequences

There are numerous ways in which one may realise the concept of degree of "conformance" to scale structure. One way would be to generate sequences comprising notes within a single scale and simply to insert controlled numbers of non-scale notes. However, just because the notes are not in the designated scale does not mean that they, together with other notes within the sequence could not conform to some scale or other. In addition, problems arise over the creation of very non-diatonic sequences in that the non-scale notes could easily begin to form their own unquantified and variable diatonic or tonal structure. The realisation of conformance chosen for the present series of experiments was therefore based on a much more tightly constrained and controlled procedure. Approximation to diatonic structure is here defined according to the number of consecutive notes anywhere in the melodic sequence that could come from a single (major) diatonic scale.

Any pair of notes of the chromatic set must come from at least two diatonic scales so the least diatonic sequences (termed first order conformant) would be such that no matter where one examined the sequence one would never find more than two consecutive notes that could possibly come from the same diatonic scale. The next least-conformant sequence (termed second-order conformant) would always have three and only three consecutive diatonic notes, never more (or less). Thus notes 1, 2 and 3 would fit into one or more diatonic scales whereas notes 1, 2, 3 and 4 would not. Similarly, notes 2, 3 and 4 would fit into a diatonic scale but not together with note 5, and so on. Third-order conformant sequences always had four and only four consecutive diatonic notes but never five, and fourth-order conformant sequences always had five but never six. No "n+1th" order segments can occur in an nth-order conformant sequence. Fifth- and sixth-order conformant sequences are possible, but are highly restricted in their constituent intervals. For example, in sixth-order sequences, which have seven consecutive diatonic notes, the intervals between consecutive notes must always be either five or seven semitones. This is because the only way of ordering the notes of the chromatic set serially so that seven and only seven consecutive notes can come from the same scale is in the form of the circle of fifths.

This realisation of diatonic conformance is not intended to prejudge the level(s) of organisation on which subjects might identify diatonicism, nor is it intended as a model of sequence perception. Neither is it related (other than superficially) to Miller and Selfridge's (1953) "orders of approximation" to English. Their sequences conformed to purely stochastic constraints, permitting, for example the occurrence of n+1th-order sequences in nth-order approximations, but conformance to target structure in the sequences here in question is fully determined (i.e. by the closed nature of diatonic inter-relationships).

The notes were also constrained within a frequency range of 20 semitones (between F=174.61 Hz and C=523.25 Hz). The largest interval permitted between consecutive tones was seven semitones in an attempt to reduce the possibility that subjects might treat temporally discontiguous notes that fell in narrow pitch ranges as separate streams (Bregman and Campbell, 1971; McAdams and Bregman, 1979) having variable and uncontrolled diatonic conformance. In an attempt to make explicit any possible effects of stream segregation three different presentation rates were used, as there is evidence that temporally discontiguous notes would be more likely to stream at faster rather than slower rates of occurrence (van Noorden, 1975). The three rates used were three, six and nine notes per second, these rates being chosen to provide a range of tempi representative of those that might occur in western music.

Other than being constrained by order of approximation and range, the notes that constituted the experimental sequences were chosen at random. Therefore the scope for inferring a mode in a given sequence or part of a sequence by virtue of the different sounding durations of different notes, the probability of certain note combinations, the starting or final notes, or the absolute pitch boundaries in relation to the scale were not confounded with the variable of interest. This does not rule out the possibility that subjects might themselves ascribe modal properties to the sequences or to parts of the sequences. The important thing is that there should be no extrinsic cues for such ascription.

In the experiment described here subjects were presented with sets of sequences, each having a specific order of conformance to diatonic scale structure. After each sequence subjects made a judgment about it (expressed aesthetic preference). It was expected that aesthetic preference would be related in some way to degree of diatonic conformance. The simplest relation between the two would be an increase in preference with increasing conformance to scale structure.

Experiment 1.1

Method

Subjects: Twenty subjects were recruited for the experiment. Ten of the subjects were postgraduate students or staff in the Department of Music of The City University, London. This group of students is referred to as the musician group (estimated mean years of formal musical training was 9.4). The remaining ten subjects were students enrolled on courses other than music at the University of London. Some of the latter (nonmusician) group had had some formal training, though none had had as much as had the "least-experienced" member of the musicians group.

Stimulus material: First-, second- and third-order sequences were composed by hand. Five sequences of each order were generated. Each of the sequences consisted of 20 notes within the previously specified range. The sequences were realised on and recorded on to tape from a Fairlight CMI. Each note consisted of a fundamental and its third and fifth harmonics, all in phase and of the same amplitude. Rise and fall times were always 17-23 msec. Each of the sequences was played at three different rates - 3, 6 and 9 notes per second. The time between offset and onset of consecutive notes was constant at 5 msec for each rate.

Order of stimulus sequences: The five sequences in each of the four diatonic orders were randomised and recorded on to audiotape at a rate of 3 notes per second. The same procedure was used for rates of 6 and 9 notes per second. Thus rate of presentation was blocked and diatonic order randomised within each block of a given rate. Whatever the rate of presentation, there was a pause of 4 seconds between each sequence and a pause of 10 seconds between each block at different rates. To provide a check on possible order effects of rate half the subjects in each group (musician and nonmusician) heard the blocks in the order 3, 6 and 9 notes per second, while the other half heard the blocks in the order 9, 6 and 3 notes per second.

Procedure: The subjects were instructed that they would listen to short tunes. They were told that their task was to indicate their degree of liking or disliking for the different tunes on a 5-point scale. A rating of 5 indicated "strong liking", of 4 "quite liking", of 3 "neutral", of 2 "quite dislike" and a rating of 1 indicated "strong dislike" (giving a range from maximum to minimum preference). Subjects were told that they could give a tune two ratings if they were unsure of which category of rating best represented their preference; in that case, the mean of the two ratings was taken.

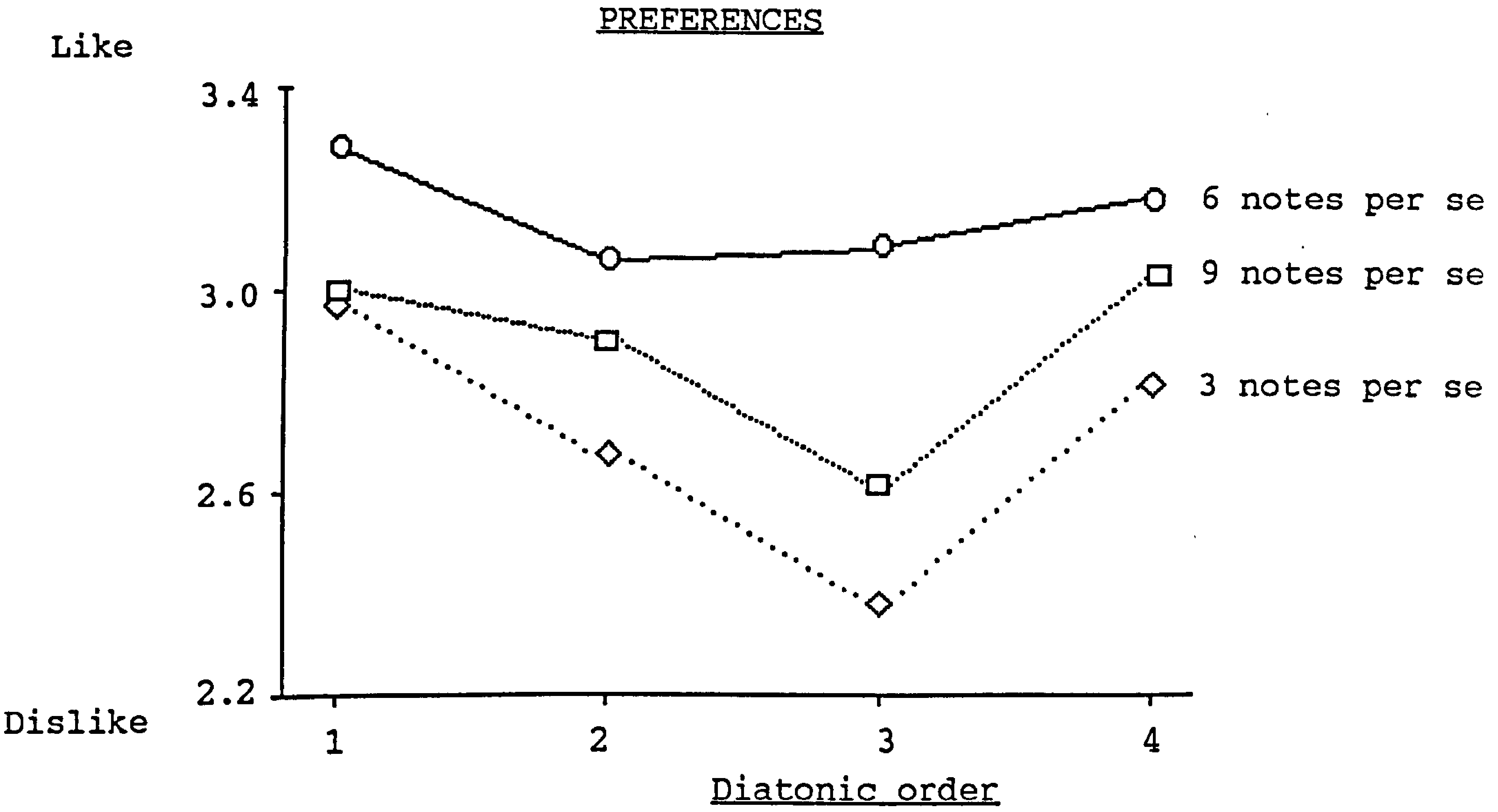
Results and discussion

The mean rating for each of the diatonic orders at the three presentation rates were calculated subject by subject. The mean values across subjects are presented in **Figure 6.1**.

At each of the three rates, preference for first-order sequences is higher than for other order sequences. Preference drops for second- and third-order sequences and increases again for fourth-order sequences. The form of the preference function across diatonic order is thus an inverted U-shape at each of the three presentation rates. The preference for six notes per second was always higher than for 9 notes per second, which in turn was always higher than for 3 notes per second.

These observations receive statistical support from an analysis of variance. The factors included in the analysis were group (musicians vs nonmusicians), rate (three levels) and order of conformance (four levels). There was a significant difference in preference for the different rates, $F_{(2,36)}=9.97$, $p<.001$ and a significant difference across the different orders, $F_{(3,54)}=7.54$, $p<.001$. There was no interaction between rate and order. It is important to note that more-or-less the same results were obtained for the group of musicians and the group of nonmusicians: there was no effect of group, nor did this interact with rate or order. An a posteriori test (Newman-Keuls) showed that ratings for first-order sequences differed significantly from those for second-order sequences ($p<0.05$) and from those for third-order sequences ($p<0.01$), while ratings for fourth-order sequences differed significantly from those for third-order sequences ($p<0.05$) and second-order ratings differed

Figure 6.1: Preference ratings versus diatonic order for each of the three presentation rates in Experiment 1.1.



Mean preference ratings for each diatonic order at each presentation rate

Diatonic Order

1	2	3	4	Rate
3.28	3.04	3.05	3.15	6
3.02	2.96	2.75	3.03	9
3.01	2.68	2.40	2.76	3

significantly from those for third-order sequences ($p<0.05$). Differences between ratings for first-order and fourth-order sequences were non-significant, as were differences between ratings for fourth- and second-order sequences (see **Table 6.1**).

Diatonic Order	Third	Second	Fourth	First
Third	-	$p<0.05$	$p<0.05$	$p<0.01$
Second	-	-	NS	$p<0.05$
Fourth	-	-	-	NS
First	-	-	-	-

Table 6.1

Results of a posteriori (Newman-Keuls) test carried out between preference ratings for sequences of different diatonic orders.

If anything, a monotonically increasing function with respect to diatonic order had been expected, but preferences for the least diatonic sequences (first order) were higher than for other diatonic orders. The obvious question to consider was whether the composed sequences included some bias; the sequences had proven to be particularly difficult to generate by hand, and although I had tried to override any musical preferences that arose during their composition, some personal predispositions may have helped shape the sequences. In subsequent experiments an attempt was made to avoid this possibility by generating the sequences by computer.

Various algorithms were employed, the first being an "exhaustive-search" algorithm (written by Robert West). This relied on a non-cyclic representation of all diatonic scale collections within a twenty-note tessitura. A "seed" of several notes was selected at random from one scale (the number of notes chosen depending on which diatonic order was desired). A new scale was then chosen at random, and tested to ensure that it contained the previous n-1 notes but not the nth note back (for third order sequences, the previous three notes must be included in the new scale but not the fourth note back). A note was then chosen randomly from the new scale, and tested to ensure that it, together with the previous n notes did not occur in any possible scale. The process was iterative, and was effective in producing second, third and fourth order sequences; however, it was not so efficient in producing first-order sequences, often taking a very long time or even failing completely.

Recall that first-order sequences require two and only two consecutive notes to come from one scale, the third note coming from a different scale. It was realised that the exhaustive-search program's difficulty in producing first-order sequences derived from the fact that there are very few types of three-note sets - or more accurately, pc sets of cardinality three - which do not fall into one diatonic scale or another. These sets are of the form (0,1,4) and its inverse (0,3,4), as well as of the (non-invertible) forms (0,1,2) and (0,4,8); using note-names, these are sets of the form (C,C#,E), (C,D#,E), (C,C#,D) and (C,E,G#). Accordingly, an "interval-pair exclusion algorithm" was developed by Peter Howell and myself to generate first-order sequences only. This algorithm simply "chained" (at random) sets of the pitch intervals formable between the available first-order - or non-diatonic - pc-sets (these interval-pairs taking forms such as +1+3, +1+1, -4+1 etc.).

However, it was felt that this solution - using different algorithms for first-order sequences - was unsatisfactory, and a further algorithm was developed for the generation of all sequence-types. This algorithm was based on the circle-of-fifths. Recall that each diatonic scale is a region constituted of seven adjacent elements on the circle-of-fifths. The circle-of-fifths algorithm simply selected n notes at random from within one diatonic region; it then chose each extra note at random so that the new note and the previous n-1 notes were within a span of seven elements, but that the nth note back was always outside the range. This algorithm (written by Robert West) proved to be fast and equally effective in generating sequences of each order.

In the next experiment, the number of diatonic orders was extended by the inclusion of completely diatonic sequences, of which the notes were randomly drawn from those available within a single diatonic scale. An attempt was also made to distinguish between the possibility that subjects' preference for first-order sequences arose from these sequences being perceived as somehow more "musical" than the others used, or whether they arose in spite of first-order sequences being judged <u>less</u> "musical" (subjectively, the sequences seemed somewhat "jazzy" but at the same time jagged and complex). To address this question, judgments concerning both preference and musicality (see Krumhansl, 1979) were obtained for each subject.

Experiment 1.2

Method

Subjects: Twenty subjects were used in this experiment. All were students at the University of London and none of the subjects were musicians as no effect of formal musical training had been observed in the Experiment 1.1.

Stimulus material: First-order sequences were generated by the interval-pair exclusion algorithm; second-, third and fourth-order and diatonic sequences were generated by the exhaustive search algorithm. As noted above, the difference between the two algorithms was primarily for computing convenience and had no effect on stimulus construction. Other details of stimulus construction were as for Experiment 1.1.

Procedure: The preference judgments were obtained from the subjects first. Preference and musicality judgments were not counterbalanced in this experiment as musicality judgments were obtained only after explaining the way in which the stimulus sequences were constructed (i.e. that different sequence types contained different numbers of consecutive notes that could come from the same scale). It was felt that if musicality judgments were obtained first, subjects' preference ratings might be affected. A 5-point rating scale was used for the musicality judgments. The order of presentation rates was counterbalanced as in Experiment 1.1.

Results and Discussion

Mean preference and mean musicality judgments were obtained for each subject. The mean preference judgments across subjects are presented in **Figure 6.2** and the mean musicality judgments across subjects are presented in **Figure 6.3**.

For the preference judgments, a U-shaped function was obtained at all the presentation rates. First-order sequences were not least-preferred, but were preferred over the second-and third-order sequences This finding replicates that in Experiment 1.1. Also, as found in Experiment 1.1 the most preferred rate was 6 notes per second, followed by 9 notes per second then 3 notes per second. An ANOVA showed that there was a significant difference in preference with respect to diatonic order, $F_{(4,76)}=4.0$, $p<.01$ and with respect to presentation rate, $F_{(2,38)}=3.6$, $p<.05$. There was no interaction between presentation rate and diatonic order.

The results for the musicality judgments on diatonic order appear similar to those reported for preference. Thus, the first-order sequences were not only most preferred but they sounded more musical to the subjects. This effect was significant in the ANOVA, $F_{(4,76)}=3.7$, $p<.01$. The ordering of rate with musicality was different from that with preference (the most preferred rate was 9 notes per second, followed by 6 notes per second then 3 notes per second). This latter finding might indicate that subjects were judging rate differently when asked about sequences' musicality rather than about relative preference. On the other hand, because musicality judgments were always obtained after preference judgments, this finding might arise through judgments changing with exposure to sequences played at different rates, perhaps indicating some increasing familiarity on the part of the subjects with the types of sequences used. In the next

Figure 6.2:Preference ratings versus diatonic order for each of the three presentation rates in Experiment 1.2.

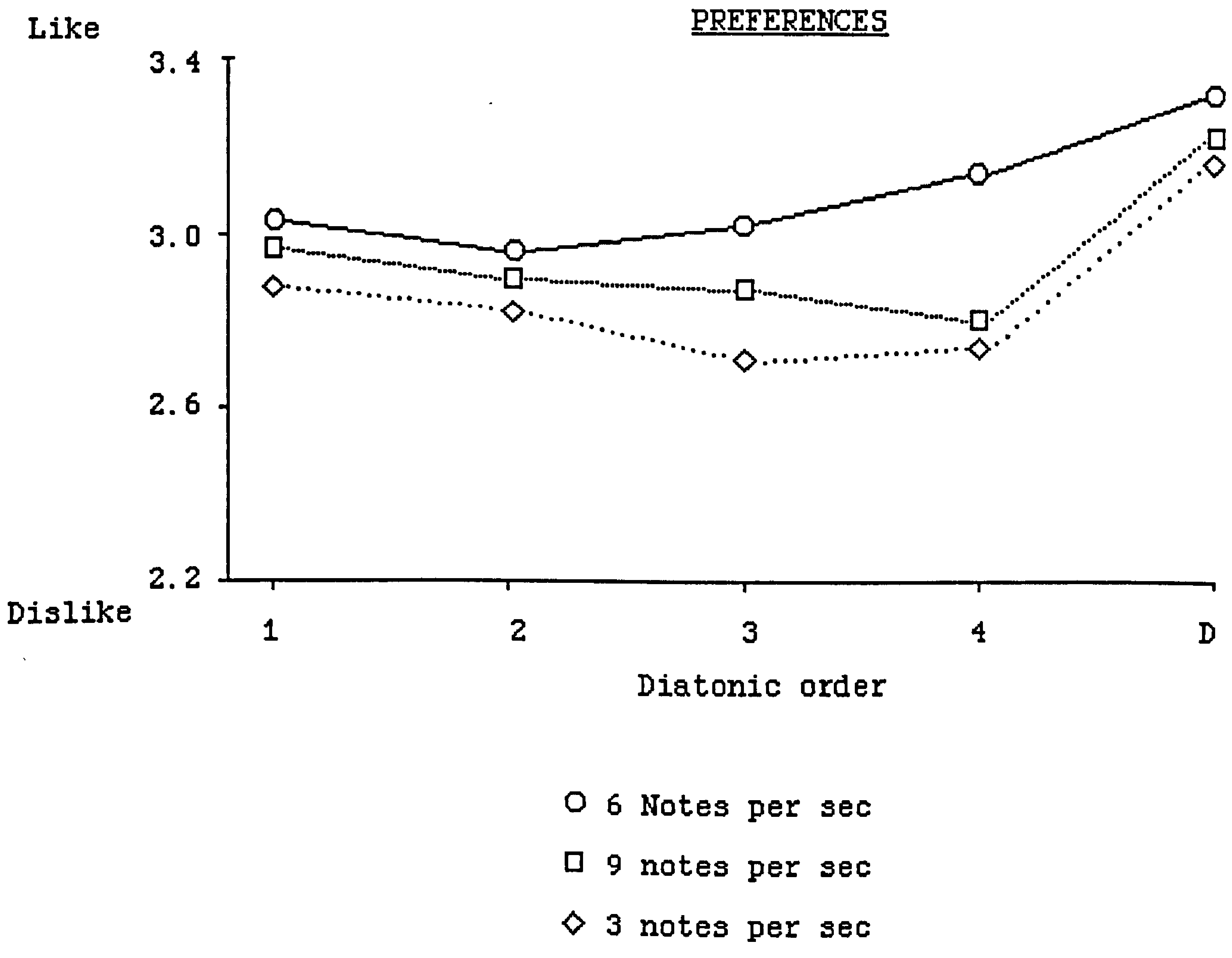
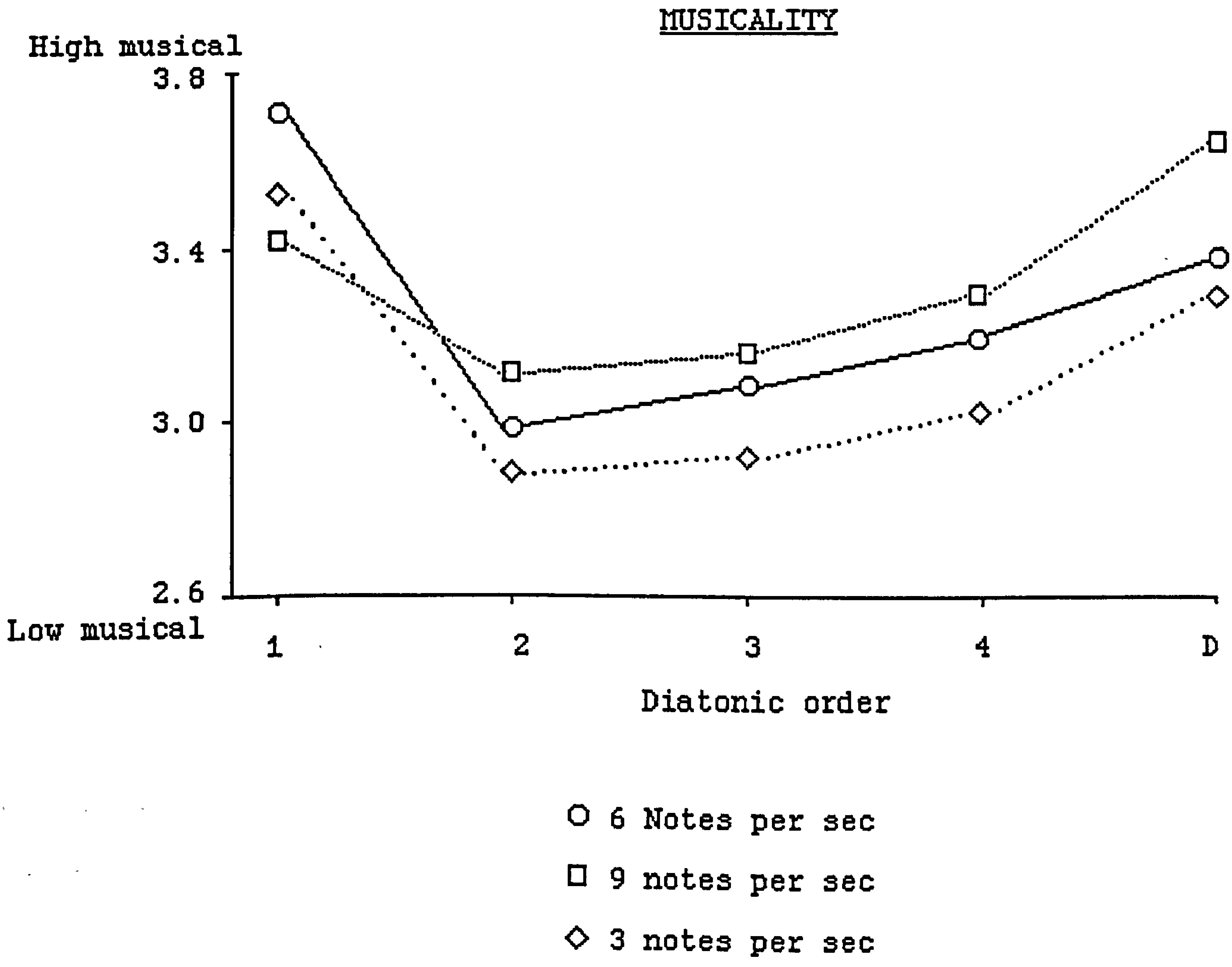


Figure 6.3 :Musicality ratings versus diatonic order for each of the three presentation rates used in Experiment 1.2.



experiment, preference and musicality judgments were counterbalanced across subjects to prevent exposure to the different rates affecting musicality and preference judgments with respect to rate. For the next experiment, implementation of the exhaustive-search algorithm had been made more efficient, and it became practical to generate first-order sequence by means of it.

Experiment 1.3

Method

Subjects: Four groups, each consisting of four subjects were used in this experiment. None of the subjects was taking a degree in music, although some had had some elementary musical training.

Stimulus material: The second-, third- and fourth-order and diatonic sequences used in Experiment 1.2 were used in this experiment, together with new first-order sequences generated using the exhaustive-search algorithm. The orders were randomised and recorded separately for each rate of presentation.

Procedure: The procedure was essentially the same as in Experiment 1.2: all subjects performed both preference and musicality judgment tasks. Order of rates (3, 6 and 9 notes per second and 9, 6 and 3 notes per second) for preference and musicality was counterbalanced across groups of subjects, as was the order of which judgment they performed first.

Results and Discussion

The preference and musicality judgments were calculated as in Experiment 1.2. The U-shaped functions obtained in previous experiments were again obtained here for the preference and musicality judgments. There was no clear separation between presentation rates as found in Experiments 1.1 and 1.2. This would, of course, be expected if judgments about rate change with length of exposure to the different rates as speculated above. Counterbalancing the order of the judgment conditions (preference or musicality) removes any marked effect of rate on judgment but leaves the effect of diatonic order unaffected.

As confirmation of these observation, ANOVAs were performed separately on the preference and musicality judgments. The between-groups factors distinguished in each analysis were order of judgment (preference or musicality) and rate orders (3, 6, 9 for preference, 9, 6, 3 for musicality and vice versa). Rate (three levels) and diatonic order (five levels) were again the within-groups factors. Although rate had a significant effect in the musicality judgment task, $F_{(2,24)}=5.8$, $p<.01$ it was not significant in the preference task. Conversely, effects of diatonic order did not achieve significance in the musicality task (though tending towards it), whereas the diatonic order effects were highly significant in the preference task, $F_{(4,48)}=7.5$, $p<.0001$. The U-shaped function is, however, still present on the musicality ratings that were obtained. These results seem to show that whereas judgments of musicality in this experimental context involve some operations differing from those involved in preference judgments, musicality is not sufficiently separable from preference to serve as a distinct judgmental dimension.

Structural bases for first-order preference

The results of the first three experiments showed that there was something special about first-order sequences that made them aesthetically more pleasing to musicians and non-musicians alike than higher orders of conformance to scale structure; moreover, on the whole these sequences were also judged to be more musical than orders other than completely diatonic.

One possible explanation for these results might be that specific structural features of the sequences were confounded with diatonic order. **Table 6.2** shows a statistical analysis of the sequences with respect to six particular note- and interval-characteristics (derived from a sample of 50 sequences of each type), and it is clear that certain certain structural features of the sequences did differ systematically. Of particular interest might be the mean interval size between consecutive notes. Small-intervals - such as are here shown to be prevalent in first-order sequences - might confer a coherence on these sequences that the other orders lacked; as Dowling (1978) points out, western melodies are typically composed of series of small pitch intervals, and it has been shown (see McAdams and Bregman's review, 1979) that small pitch differences between consecutive notes in a sequence contribute very significantly to the perception of such a sequence as a coherent "linear stream" or melodic line.

	(a) No. of reversals		(b) Pitch range		(c) Pitch midpoint		(d) Interval size		(e) Var. of occur.		(f) Var. of int. size	
Order	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
1st	13.4	6.89	13.6	**8.80**	8.7	8.49	2.3	**.09**	14.4	**71.36**	1.7	**.41**
2nd	14.5	5.03	17.8	3.29	9.7	2.48	5.0	.01	23.4	21.44	3.3	.13
3rd	13.4	6.17	17.1	5.59	9.9	3.38	3.9	.25	22.3	42.94	3.1	.35
4th	15.3	3.27	18.2	2.03	9.7	2.16	5.8	.01	24.0	13.10	1.3	.16
Dia	13.1	4.50	17.4	3.23	9.7	2.99	3.7	.12	22.0	45.47	3.4	.48

Table 6.2

Means and standard deviations of contour (a), range (b), pitch midpoint within overall experimental range (c), interval sizes (d), variance of occurrence of notes (e) and variance of occurrence of different interval sizes (f) in experimental sequences. Numbers in **bold** type indicate p<.01. The sequence characteristics are derived from a sample of 50 sequences of each order (250 sequences in all). The p values are for first-order sequences versus all others.

A further explanation might lie in the different range of specific constraints over permissible intervals between consecutive notes (and between groups of notes) necessary to produce sequences of different orders. As noted above, the sixth order of conformance only allows two possible intervals - a perfect fourth and fifths. The fourth-order sequences certainly had a greater predominance of these interval than had other sequences and were thus to some extent structurally parsimonious in the range of interval types of which they could be composed. First-order sequences similarly displayed a considerable degree of structural parsimony, not so much in terms of permissible single intervals but in terms of permissible interval sequences. As mentioned above, first-order sequences could be generated by means of the interval-exclusion algorithm because of this constraint, while second-, third- and fourth-order sequences could not. As this feature of different types of pattern parsimony (of first- and fourth-order sequences) are necessarily confounded with diatonic order, it cannot be ruled out as a possible explanation for the U-shaped preference and musicality functions found in the three previous experiments. Note, however, that neither of these two explanations can account for the high ratings given to the completely diatonic sequences (unless one assumes that the U-shape arises from the operation of different sets of judgmental criteria applied to the diatonic sequences and to those of other orders).

The third explanation derives from the possibility that subjects may have imposed their own subjective metres on the sequences, although there were no explicit cues - such as duration, or variation in intensity level - to such metres; metre is here taken as a temporal grouping of note sequences into segments of nominally equal duration (bars), each comprising a specifiable numbers of "beats" or rhythmic units (Shaffer, 1982). The first beat of a bar is often indicated by increased intensity, by the length of the note occurring on that beat, by the imposition of some "structural accent" (see Lerdahl and Jackendoff, 1983) or by a number of other cues. In the experimental sequences, all notes were produced with similar intensities and durations. Moreover, there should have been no systematic constraints that would have provided cues for any specific subjective metre. However, in the absence of explicit cues subjects can and do impose their own metres where possible (Povel and Okkerman, 1981; Thomassen, 1982), perhaps to "make sense of" the sequence by temporally segmenting it in an attempt to integrate it into some larger-scale linear or hierarchical structure. Jones, Kidd and Wetzel (1981) provide evidence that listeners can attend to temporally discontiguous notes of a sequence on the basis of metrical regularity. If this is the case, then the prime candidates for a subjective metre would be two, three or four beats in the bar. There are other common metres used in Western music such as six, eight, nine or twelve beats in a bar, but these are fairly long (requiring greater memory span to allow them to be sustained) and are anyway divisible by at least two of the shorter metres.

If the first-order sequences are examined, it becomes apparent that a sequence formed of every fourth note is necessarily a fully diatonic sequence; that is, notes 1, 4, 7, 10, ... or notes 2, 5, 8, 11, ... or note 3, 6, 9, 12, ... of any first-order sequence form, in their own right, fully diatonic sequences. As noted in the account of the interval-pair exclusion algorithm, first-order sequences can be generated by "chaining" pitch-class sets of the types (and <u>only</u> of the types) (0,1,2):(0,4,8):(0,1,4):(0,3,4). These can be thought of as forming "triangles" within the circle of fifths (or as concatenating to form regions analogous to diatonic scalar sets <u>but on the opposing diagonal</u> in Balzano's (1980) direct-product group representation of the chromatic set). In any chain of such triangle configurations used to construct a first-order sequence (at least, of the length used here), every fourth note must fall within a region of the circle of fifths spanned by a single diatonic major scale. Thus if subjects were imposing - consciously or otherwise - a subjective metre of three beats in a bar on the first-order sequences, the first (and presumably metrically most significant) note of each temporal group (i.e. <u>1</u>, 2, 3, <u>4</u>, 5, 6, <u>7</u>, 8, 9, <u>10</u>, ...) would form a diatonic sequence. The sequence could thus be "heard" as an essentially diatonic sequence with two intervening nondiatonic notes between each diatonic note, and would be functioning as an example in perception of the principle of "reduction" advanced in differing forms by Schenker (1935/79) and by Meyer (1973).

Second-order sequences, in which every sixth note conforms to a diatonic sequence, require an unusual (in Western music) subjective metre of five beats in the bar for subjects to perceive diatonicism in this way. In third-order sequences a metre of six beats in the bar is needed, and with fourth-order sequences the required metre is seven beats in the bar (see **Table 6.2** below). Thus, only first order sequences permit the identification of scale structure in periodic discontiguous notes by means of what could be termed a simple or typical subjective metre. In sum, the preference for first-order sequences together with their adjudged musicality may be based on their potential for enabling subjects to identify or extract conformance to scale structure via an accessible metric structuring.

Order	No of contiguous diatonic notes	Period of discontiguous diatonic notes	Implied metre
First	2	3	3-beat
Second	3	5	5-beat
Third	4	6	6-beat
Fourth	5	7	7-beat
Diatonic	All	-	-

Table 6.3

Number of contiguous and period of discontiguous diatonic notes (plus implied metre) in each order.

Of these competing explanations for the preference for first-order sequences, the first two are difficult to test directly as there is no scope for independently varying the factors involved. The third explanation is easily tested, however, and such a test could at the same time clarify or even eliminate the two other hypotheses. This test simply involves imposing cues to different metres on the sequences. This could be done by increasing the intensity of, e.g., every fourth note (to give a three-beat metre). If the subjective metre hypothesis is correct, a three-beat metre should leave the preference for first-order sequences intact, whereas a four-beat metre should prevent subjects from imposing their own three-beat metre, thereby reducing preference for first-order sequences by "blocking-out" the conditions under which diatonicism might be identified. The next experiment was designed to test this idea.

Experiment 1.4

Method

Subjects: Ten subjects with musical backgrounds similar to those in Experiments 1.2 and 1.3 participated in this experiment.

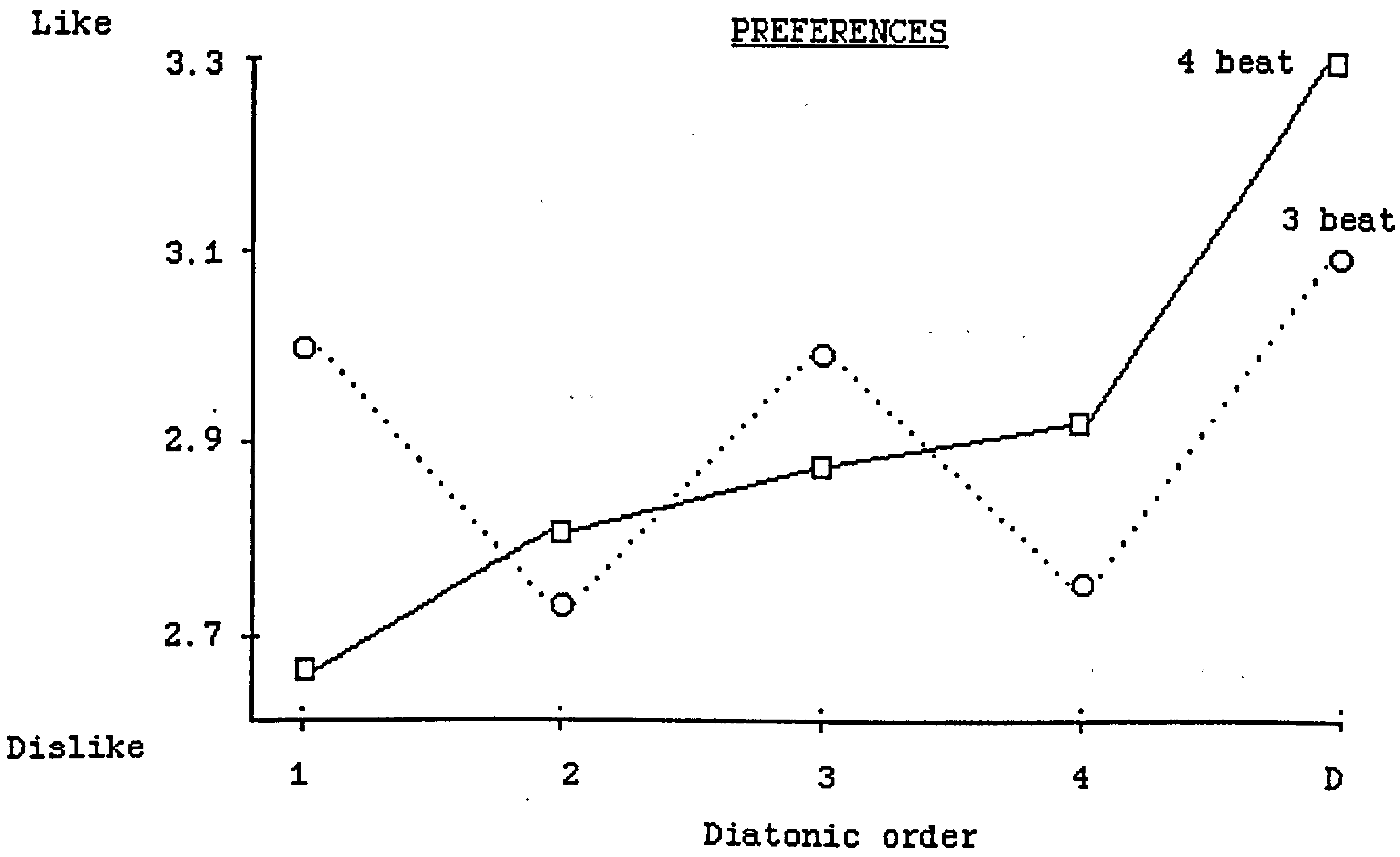
Stimulus material: The circle of fifths algorithm was used to generate all sequences. This algorithm produced exactly the same results as those used previously, but was much quicker. Each sequence consisted of 24 notes rather than twenty as used previously. An accent structure was introduced to the sequences on every fourth or fifth note by doubling the intensity of the appropriate note: these accents produced metres of three beats and four beats in the bar. To ensure that there were no unlooked-for interactions between sequence type and where the metrical unit started, the notes on which the accents occurred were varied across sequences. Thus, one one sequence might have the notes 1, 4, 7, ... accented, while another would have accents on 2, 5, 8, ... All the sequences were recorded at a rate of 6 notes per second.

Procedure: Subjects rated the sequences for preference as in the earlier experiments. Subjects listened to 175 sequences in all, made up of five sequences in each of the five orders used, with the metrical unit starting on each of the three different positions available within the three-beat metre and on each of the four starting positions in the four-beat metre (making a total of seven starting positions). All subjects heard the same randomised set of 175 sequences.

Results

Preferences for each diatonic order, each metre and for each starting position of the metrical unit were calculated. In an ANOVA, diatonic order, metre (three or four beats) and starting position in each metrical unit were distinguished. None of the main effects were significant, but the interaction of diatonic order and metre was highly significant, $F_{(4,36)}=8.9$, $p<.001$. These results show clearly that the pattern of preferences across diatonic order differs depending upon what metre occurs. The interaction is presented in **Figure 6.4**. Most significant is the fact that under the three-beat metre there is a peak at the first-order sequences that is completely abolished under the four-beat metre. Also of interest is another peak at the third-order sequences.

Figure 6.4:Preference ratings versus diatonic order for each of the two different imposed metres.



Discussion

The last experiment showed that with a four-beat metre imposed on the sequences an increase in preference with diatonic order emerged, with first-order sequences being the least preferred. Despite the fact that noise might have been expected in the data arising from fatigue effects due to the overall duration of Experiment 1.4, the results appear clear and unequivocal. With the three-beat metre, there was a greatly increased preference for first- and third-order sequences. This provides clear support for the subjective metre hypothesis, which was proposed as an explanation for the first-order effect found in the previous three experiments. This experiment also rules out the two other explanations, as the structural features of first-order sequences remain unchanged in both metrical conditions. The peak at the third-order sequences in the three-beat metre was unexpected, though it should in fact have been predicted from the subjective metre hypothesis. Recall that identification of diatonicism from periodic discontiguous notes requires a metre of six beats in a bar (see **Table 6.3**). Without any imposed metre, it was suggested that this might require too great a span to form a likely subjective metre. However, with an imposed metre of three in a bar the identification of diatonic relationships between nonadjacent accented notes should be facilitated.

It would be tempting, on the basis of this discussion and of **Figure 6.4** to suggest that subjects used a kind of "default" metre of four beats in the bar, which in the case of first-order sequences, was modified to a three in a bar metre because this permits the identification of diatonic relationships in these sequences. (Joining the three-beat rating for the first-order sequences to the four-beat ratings for the others provides the U-shaped curve found previously.) This would suggest an active attempt on the part of subjects to "make sense of" the sequences by seeking out diatonic relationships, making use of shifting metrical-attentional strategies. Such a hypothesis may be correct, but at this stage there is not sufficient evidence to determine with the higher-order sequences that other subjective metres were in operation, or whether there was any systematic subjective metre at all. What is clear is that a three in a bar metre was in operation for the first-order sequences, and that this was the basis for the increased preference shown for these sequences.

A further issue that remains unresolved is how far the subjective metre imposition with the first-order sequences was a conscious strategy or even whether there is any consciousness of the results of this process. In the preceding discussion the term metre has dominated, but this has been mainly for clarity of exposition. There is no reason why the identification of scale structure from periodic elements within the sequences may not occur via some more abstract identity between these periodic elements, involving not only the first note in each subjective group but all the other notes as well. For example, as mentioned earlier, not only do the notes 1, 4, 7, 10, ... form a fully diatonic sequence, but so do the notes notes 2, 5, 8, 11, ... and the notes 3, 6, 9, 12, ... It is therefore possible that subjects would consider themselves to be adopting no specific metre, whereas at some level this grouping was going on and creating its effects on preference.

The results, however, are consonant with the hypothesis that subjects in some way segmented the sequences in perception in accordance (where possible) with the metrical structures indicated by the experimental findings. It was felt that subjects might develop representations of metrical subgroups of the sequences, so that, for instance, first-order sequences might be representable in memory as a series of three-note groups. This was tested by means of a recognition memory experiment making use of only first-order and completely diatonic sequences. Each sequence was followed by a three- or four-note probe set. If subjects indeed had access to sequence segments in memory the size of probe set should make no difference to recognition of probes following diatonic sequences, but three-note probes following the first-order sequences should be better recognised than should four-note probes.

Experiment 1.5

Method

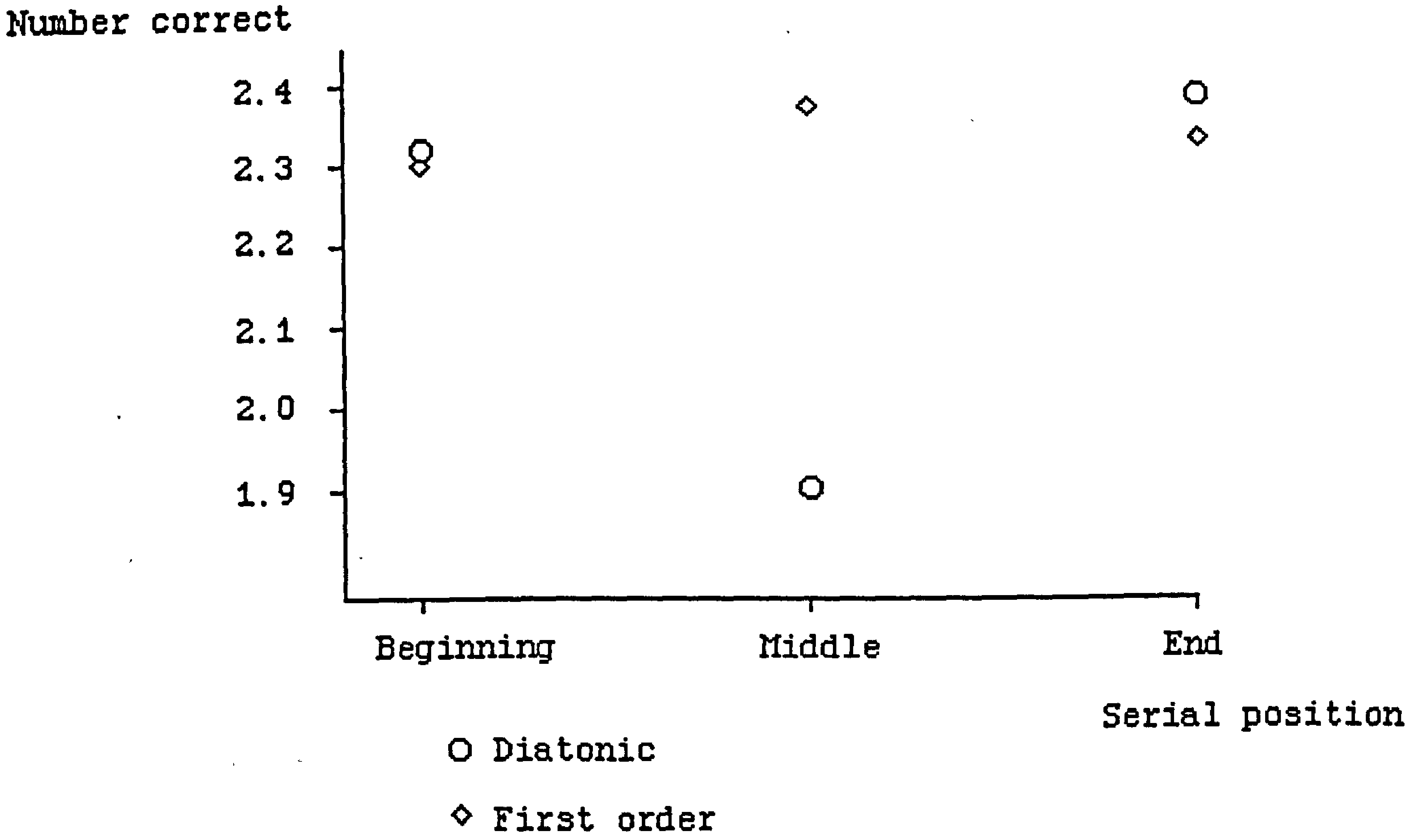
Subjects: Thirty-two subjects with musical backgrounds similar to those in Experiments 1.2 and 1.3 participated in this experiment. Stimulus material: An algorithm similar to that used in Experiment 1.4 was used to generate all sequences, which were either fully diatonic or were first-order conformant. Each sequence consisted of 12 notes presented at a rate of four notes per second, followed by a pause of one second and then either a three-note or four-note probe. The four-note probe was either notes 1-4, notes 5-8 or notes 9-12 of the preceding sequence, or was a distorted version of one of these blocks. The three-note probe was either notes 1-3, 4-6, or 10-12 of the sequence, or was a distorted version of one of these blocks. The distortion (for three- and four-note probes) was produced by transposing one of the notes of one of the original probes by a semitone.

Procedure: Subjects listened to each trial (sequence plus probe) and indicated whether they thought the probe had occurred in the preceding sequence. All subjects listened to 60 sequences in the three-note probe condition and 60 sequences in the four-note probe condition. Half the subjects heard the three-note block of trials first, followed by the four-note block, while half the subjects heard the four-note then the three-note block. Both blocks of trials (three-note and four-note) contained five sequences in each of twelve possible conditions; these conditions were derived by crossing sequence type (diatonic or first-order), position of probe (beginning, middle or end), and probe type (correct or distorted).

Results and discussion

The experimental findings were not simply relatable to the anticipated segmentation of the sequences; three-note probes following the first-order sequences were not significantly better recognised than were four-note probes. In fact, no interaction was observed between probe size (three- or four-note)and diatonic order. However, it was noted that while a typical serial position effect was observed for the completely diatonic sequences (with recognition better for probes drawn from the beginning and end of the sequences than for probes drawn from the middle), probes following the first-order sequences seemed to be equally well recognised wherever they had occurred within the sequences (see **Figure 6.5**). In the ANOVA this interaction between sequence type

Figure 6.5: Number correct versus serial position of probe for each of the two diatonic orders used.



and position of probe proved to be highly significant, $F_{(2,42)}=6.56$, $p<.01$.

There are several possible explanations for these findings. The lack of a serial-position effect for the first-order sequences might be explained by the fact that as subjects could not easily relate the first-order probes to diatonic structure in the case of the first-order sequences they were forced to retain the probes in memory in the form of "raw" interval information. As Dowling (1982b) seems to indicate, storing melodic fragments in memory in this form can in some circumstances lead to more accurate retention than might an attempt to fit or encode the fragment in terms of some diatonic structure.

In addition, subjects' performance in Experiment 1.4 was accounted for by their inferring diatonic structure from periodic and discontiguous notes in the experimental sequences. It might then be expected that no specific advantage in memory would be conferred on three-note probe sets in the first-order sequences, consisting as they did of blocks of three consecutive notes. As Jones, Boltz and Kidd (1982) demonstrated, listeners seem to be more sensitive to relationships between notes of a melody which are discontiguous than they are to relationships between adjacent notes when the former conform to some rule-structure and the latter do not seem to conform to such a structure. Probe sets might thus have been generated more appropriately from discontiguous rather than consecutive sets of notes (for example, an appropriate probe following a first-order sequence might have been constructed from the notes 1, 4, 7 and 10). This, however, would have been problematic. One option would have been to present the probe with the same note inter-onset interval as occurred between consecutive notes of the sequence, thus presenting the probe at three or four times the rate at which it would have occurred in the sequence. Alternatively, the probe could have been presented with the same note inter-onset interval as that with which it occurred in the sequence (i.e. at one-third or one-quarter the rate of the preceding sequence). This change in rate might have acted as an extraneous cue as to sequence segmentation for the subjects, and would have been acting in addition to any cues supplied purely by diatonic order, necessarily being confounded with such cues.

Notwithstanding the failure of Experiment 1.5 to illuminate subjects' strategies, the series as whole showed that adult Western listeners can use diatonic scale structure independently of any higher-order temporal patterning. That is, simply on the basis of the pitch intervals in a melodic sequence, listeners can infer whether or not - or perhaps even to what degree - the sequence conforms to diatonic structure, independently of any overt saliency-imparting cues (such as repetitions of notes, use of particular scale notes or sets of notes to start or end sequences, etc.) as to which notes of the melody might be considered as the modal centre or tonic.

This sensitivity to diatonic interval structure would seem to indicate that whatever the form taken by the cognitive representation of musical pitch, it embodies diatonic interval structure in some way. In other words, alphabetic diatonic structure does appear to be privileged in cognition. However, this does not necessarily imply that Dowling's "tuning system" has any psychological reality that is separable from modal levels of organisation, as the results obtained do not indicate clearly the nature of any relation that might hold between a representation of diatonic interval structure and a tonal hierarchical representation; for instance, listeners could have inferred some event hierarchy - and hence some specific mode or key - in the course of listening to each sequence, and made use of this inferred modal structure in their judgments. Moreover, the fact that this sensitivity to diatonic intervallic structure exists is not particularly informative - except in a very general way - about how a cognitive representation of diatonic intervallic structure might come into play in listening to a melody.

CHAPTER SEVEN

Even given that the cognitive representation of musical pitch embodies diatonic interval structure, significant questions remain to be answered. What form does our representation of diatonic interval structure take ? How does this representation relate to any possible representation of the tonal hierarchy ? Are we sensitive to the types of properties that Shepard and Balzano describe as inherent in diatonic interval structure ? Are we, for example, sensitive to the diatonic property that different scale degrees may be identified by virtue of their unique intervallic contexts ? What role might the representation of diatonic interval structure play in listening to a melody ? What types of perceptions - or rather, musico-perceptual judgments - does it enable ? For instance, does it enable us to judge when notes belong in the same diatonic collection or scale, and if so, how ?

All of these questions address facets of a broader question: what specific contributions - if any - to a listener's experience of tonal relations are made by their cognitive representation of diatonic interval structure ? These contributions could take several forms, some of which are suggested by the diatonic properties of uniqueness, coherence and simplicity outlined by Balzano. However, the results of Experimental Series 1 suggest one obvious influence of diatonic interval structure on tonal perception; sequences of notes that fall within one scale seem to be heard as "belonging" or "fitting" together better than sequences of notes which continually roam from diatonic scale to scale. That is, listeners seem to sense that sequences of notes which conform to diatonic interval structure are somehow connected. Investigation of just how this sense of connectedness comes into being in the course of listening to a melody should provide clues to the way in which diatonic structure is represented in cognition, and should help to clarify the roles played by that representation in the perception of tonal relations. Accordingly, a further series of experiments was carried out to examine how this sense of connectedness - or diatonic conformance - might arise in melodic perception.

Experimental Series 2

It appears that judgments of the diatonic conformance of a melody could be carried out in several ways, depending largely on the nature of the listener's representation of diatonic structure. A listener may match notes and the pitch intervals that separate them against some representation of the sets of notes or intervals that may occur within a diatonic scale, and judge conformance according to whether or not a match is achieved. Or a representation of a specific scale or set of scales may be activated as the melody unfolds, leading to judgments of conformance based on adherence to that scale or set of scales; a representation of a specific diatonic scale might incorporate some differentiation between scale degrees in the form of a tonal hierarchy, or it might not. Alternatively, a listener's representation of diatonic structure may embody some as yet unspecified properties of diatonicity, which may directly influence their judgments of diatonic conformance.

If the first of the above hypotheses were to hold, a listener would have to match each new note or interval together with its predecessors against diatonic or non-diatonic interval combinations. For example, the notes C, D, F and G may all be found within at least one scale, so a listener can simply judge it as diatonic-conformant without necessarily making any further inferences; similarly, the notes C, Eb and E cannot fall within the same scale and so any sequence which contains those notes can be considered non-diatonic by a listener without needing to consider which scale or set of possible scales is violated. Note that the property of diatonic conformance or non-conformance is intrinsic to sets of notes <u>irrespective</u> of the order in which they actually occur in time.

If the second hypothesis were to hold, the listener would not need to remember the individual notes or intervals of the sequence. As each new note is presented it may help the listener to activate some representation of a scale or set of scales which is consistent with the sequence up to that point. Listeners would then build up this diatonic schema and match subsequent notes or intervals with it. Thus as each new note is presented it is "fitted" against the evolving diatonic schema; if it is within the same scale as its predecessors it may serve to strengthen the diatonic schema, otherwise it will be judged as ill-fitting. Alternatively, subjects might assume that certain notes of the unfolding sequence fulfilled specific tonal functions, and make their judgments as to diatonic conformance on that basis. In this case, a listener would be likely to set up and to maintain a strong and specific hypothesis in the course of a sequence as to which scale or key its notes were derived from (i.e. judge the sequence as wholly diatonic or tonal), only forming judgments about degree of diatonic conformance on the occurrence of notes which disconfirmed the hypothesis.

A question arises as to what factors might influence the strength of a particular diatonic schema. It could be that the number of different notes presented from a single diatonic scale has an effect in itself. Alternatively, or in addition, pitch relations among the notes presented (assuming that they are all drawn from the same scale) could play a role. For example, a sequence of notes which presents a "plausible tonic" early on in its course might give rise to the hypothesis that a specific scale or key underlies the sequence; the identification of a note *as* the "tonic" could arise through an awareness of the likelihood that a member of the tonal hierarchy will occur fairly early on in most tonal melodies or perhaps - as Shepard and Balzano hint - because the listener's representation of diatonic structure enables the "tonic" to be identified by means of the intervals that it is capable of forming with the other notes of a major key.

A series of experiments was conducted in order to determine how people's sense of diatonic conformance arises in the course of listening to a sequence of notes. It would not have made sense simply to ask subjects to rate the degree of diatonic conformance of melodies as they progressed because many of them would not have understood the task (unless expressed in terms of preference or musicality as in the previous experiments) and secondly because too many variables with no necessarily direct connections with diatonic structure - such as contour - would have to be manipulated in the experimental design. Subjects in the following experiments were presented with sequences of notes that contained one out-of-scale note and were asked to indicate as soon as they heard a note which they felt did not fit in with its predecessors. The likelihood of subjects indicating an out-of-scale note as ill-fitting in a given experimental condition should give an indication of subjects' abilities to detect violation of diatonic scale conformance in that condition.

All sequences were generated randomly by computer, subject to the experimental constraints outlined below. Accordingly, the results of the experiments should not be confounded by any consistent well-formedness of experimental sequences with respect to a particular musical style. They should indicate whether aspects of diatonic structure considered as a formal system can account for judgments of diatonic "belongingness" without reference to criteria derived from observed melodic principles in particular musical styles.

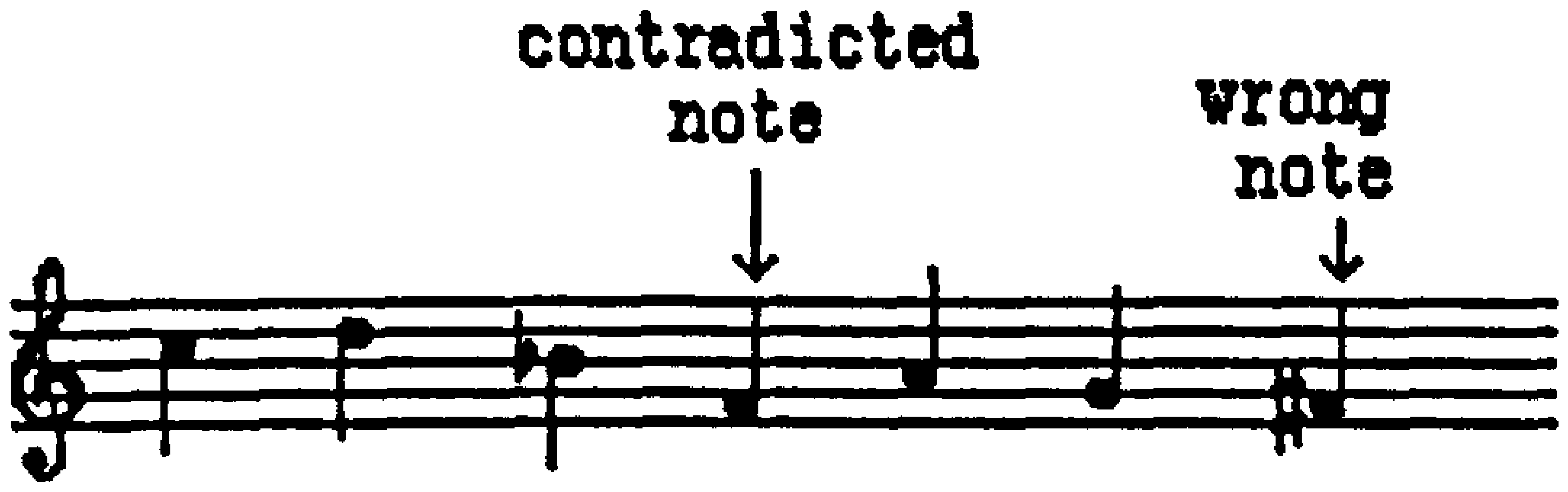
Experiment 2.1

Subjects listened to note sequences and were asked to indicate when a note occurred which they felt did not fit in with its predecessors. The sequences consisted of all the notes from a single scale with no repetitions. At some point in each sequence a note was introduced which could not be accommodated within the same scale as its predecessors. Subjects' performance in identifying that note as the non-fitting note was measured. The sequences were randomly generated by computer, subject to the constraints imposed by the experimental conditions; these involved varying four parameters of the sequences prior to the non-diatonic - "wrong" - note.

One parameter was simply the number of scale notes presented prior to the wrong note. Thus the wrong note might occur after the fourth, fifth or sixth note. This was intended to test the possibility that simply increasing the number of diatonic - within-scale - notes prior to the non-diatonic note might assist listeners to "build up" a diatonic or tonal schema and might positively enhance their ability to detect the "wrong" note.

A second parameter was whether or not the notes preceding the wrong note specified a single scale. As any seven adjacent notes on the circle of fifths constitute a single diatonic scale, this parameter depended on the spread within the circle of fifths that the notes prior to the wrong note were drawn from. This spread was either seven notes (giving only one possible scale) or six notes (permitting two possible scales). For example, if there were four notes prior to the wrong note, these might be F, G, A and B (giving a spread of seven notes, and allowing of only one possible scale), or F, G, A and E (giving a spread of six notes, and allowing of two possible scales). Alternatively, if there were five notes prior to the wrong note, these might be C, B, F#, A and G (giving a spread of seven notes, and allowing of only one possible scale), or C, B, E, A and G (giving a spread of six notes, and allowing of two possible scales). The parameter may be referred to as "number-of-scales" <u>or</u> as "scale specificity", these two terms being directly and inversely related. It should be noted that the value of this parameter is <u>not</u> dependent on the order in which the notes actually occur.

Figure 7.1:Melodic sequence as used in Experiment 2.1



The third parameter, describable as "distance to the contradicted note", has elements in common with the nature of sequence construction in Experimental Series 1, and requires some explanation. Figure 7.1 shows one of the experimental sequences. The first six notes fall into one diatonic scale, but the seventh - F# - cannot be fitted into the same scale as <u>all</u> of its predecessors together. However, this "wrong" note can belong to the same scale as its immediate predecessor, G, and both of these can belong to the same scale as their predecessor A. It is not until one goes back three notes to the note F that the non-diatonic nature of the F# becomes apparent. Thus F#, G, and A can fall within a scale, but F#, G, A and F cannot. In this case the distance to the contradicted note is three, the contradicted note being F. Distance to the contradicted note, therefore, reflects the number of notes which a subject would have to retain in memory in order to identify a non-diatonic note-combination involving the out-of-scale note.

A fourth parameter was the nature of the contradicted note itself. This could function either as a plausible tonic for the sequence of diatonic notes preceding the wrong note, or as a plausible mediant. The word "plausible" is used here as the possibility of assigning tonal function to specific notes is governed in these sequences by their "intervallic contexts" (as suggested by Balzano) rather than by syntactic or durational cues. Moreover, in the case of a sequence which could derive from two possible diatonic scales, two plausible tonics are available. The contradicted note - F - in the sequence shown in **Figure 7.1** is a plausible tonic, although the preceding Bb is also a plausible tonic. However, in the case of a sequence which could derive from only one scale, only one plausible tonic is available (e.g., in the sequence B, A, C, F, G, D preceding the wrong note Db, the contradicted note - C - is the only possible tonic).

The effects of the sequence parameters on wrong-note detection can provide information about how judgments of diatonic conformance are made, and hence give some clues as to the nature of the diatonic representation. If a greater number of diatonic scale notes prior to the wrong note (independent of other factors) improves detection of that note, it is likely that presentation of different scale notes is helpful in building up a diatonic scalar schema against which subsequent notes may be judged. If the "scale-specificity" variable influences wrong-note detection, than the diatonic schema which listeners derive must be based at least in part on information of this kind. If distance to the contradicted note influences wrong-note detection, then it is likely that listeners are attempting to retain specific information about notes and intervals and to match this against what is permissible within the diatonic scale; if this is the case, then the further back that subjects have to go to realise that a note does not fit with its predecessors, the worse should be their performance. The fact that the contradicted note could be either a plausible tonic or plausible mediant was intended to test for the possibility that subjects' representations of diatonic interval structure were integrally bound to a representation of a tonal hierarchy. If this were the case, subjects' performance on wrong-note detection might be expected to be better when the contradicted note was a tonic (and therefore more "anchored" to a strong representation of a specific key) than when it was a mediant.

Subjects also performed a note-repetition detection task in order to provide a comparison with wrong-note detection. The sequences used for this task were identical to those used in the wrong-note detection task, except that the out-of-scale note was replaced by a repetition of the contradicted note. Recall that the contradicted note was the note at which, counting back in the sequence from the non-diatonic - wrong - note, the non-diatonicity became apparent. For example, in the sequence B, A, C, F, G, D, Db, the wrong note is Db. One needs to go back to the C before the non-diatonic quality of the note Db becomes apparent. The analogous repetition detection task sequence would be B, A, C, F, G, D, C.

With the repetition detection task, all the parameters which were manipulated for the wrong-note detection task were varied similarly. Thus the number of notes prior to the repeated note, the number of notes between the first and second occurrence of the repeated note, the scale-specificity of the sequence of notes prior to the repeated note and whether or not the repeated note was a plausible tonic were varied. It was expected that increasing the distance between the first occurrence of the note and its repetition would impair performance, because of increased memory load. Performance should also worsen with increased number of notes prior to the repeated note, because of proactive inhibition from irrelevant notes. It was not considered likely that the scale-specificity of prior notes would affect this task. Differences and similarities in the performance on the note-repetition and wrong-note detection tasks should provide information about the role of memory for specific notes or intervals in judgments of diatonic conformance.

Method

Subjects: Eight university undergraduates or postgraduates took part in the experiment (five female, three male). None had undertaken or or was undertaking a university course in music. Four had received no formal musical training: four had at one time played a musical instrument, of whom two had studied elementary music theory at school. None reported being "tone deaf" or having "absolute pitch".

Stimulus material: The sequences were generated on a PDP-11 computer; the algorithm used to generate the material was programmed by Robert West. The output of the program consisted of eight numbers specifying the order of a sequence of notes drawn from the same range as that used in Experiments 1.1 - 1.5 (from F=174.6 Hz to C=523.23 Hz) and having the same spectral and amplitude envelope characteristics. All notes across the range used were synthesised on a Fairlight CMI and recorded digitally into the PDP-11 at a 10 KHz sampling rate and low-pass filtered at 4.5 kHz (48 dB/octave).

Two types of sequence were generated. These differed with respect to whether they contained a wrong note or whether a note was repeated. A wrong note was defined as a note that could not exist in any diatonic scale together with all the preceding notes.

Wrong-note sequences varied in a specified manner according to the scale-specificity of the notes (i.e. their spread within the circle of fifths) prior to the wrong note (either one or two possible scales), the number of notes in the sequence prior to the wrong note (4, 5 or 6), the distance to the contradicted note (3 or 4) and whether or not the contradicted note was a plausible tonic. Repetition-detection sequences were similarly varied, except that the distance to the contradicted note became the distance between the first occurrence of the repeated note and its repetition. Thus five factors were varied independently: task (repetition-detection versus wrong-note detection), number of notes prior to the contradicted note (4, 5 or 6), distance to the contradicted note (2 or 3), scale-specificity of the prior notes (one or two possible scales) and status of the contradicted note (tonic or mediant). Five sequences were presented for each of the 48 conditions thus created. Trials were blocked according to task (repetition-detection versus wrong-note detection), and task order counterbalanced across subjects.

Subjects were instructed to indicate which note in the sequence they thought was either ill-fitting (wrong-note detection) or was repeated (repetition-detection). It should be noted that subjects were offered no specific criteria upon which to based their judgments. They were not informed that judgments of ill-fittingness involved the detection of an out-of-scale or non-diatonic note. There was no attempt to define the term "ill-fitting" in terms of diatonicity for the subjects, who were left to determine their own criteria.

The subjects had two response keys. Depressing one of the keys "fed" them a note 500 ms after it was depressed; this self-pacing was adopted after an initial pilot study indicated that subjects were more-or-less unable to respond at all to experimenter-paced sequences. When the subjects thought that the last note they had been played was either ill-fitting or was a repetition of a previous note (depending on which condition they were performing), they depressed the other key, the selection was recorded by the computer and no more items from that sequence were played. If subjects did not detect what they thought to be a wrong or repeated note after all eight notes had been played, the computer recorded that no selection had been made. Two seconds after the subject's response (or non-response), another sequence was played. A short burst of narrow-band high-frequency noise served as a signal for subjects of the onset of each new sequence. The process was repeated until the subjects had heard all the sequences.

Results

The number of times a subject correctly identified the non-diatonic note as the wrong note was counted for each of the 48 sequence types (maximum=5). Similarly, the number of times a note repetition was correctly identified was counted. An ANOVA was conducted to determine the effects of the four factors (number of notes prior to the wrong/repeated note, distance to the contradicted/repeated note, scale-specificity and task type) on subjects' performance.

Table 7.1 gives the ANOVA summary table for all main effects and two-way interactions involving the task. It can be seen that there were main effects of: task (repetition detection versus wrong-note detection) and distance to the contradicted or repeated note. Significant interactions arose between: task and distance to the contradicted or repeated note, task and number of notes prior to the wrong or repeated note, and task and scale-specificity.

Source	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Task	1	2.7	2.7	14.6	0.007
Distance	1	0.66	0.66	19.2	0.003
No of prior notes	2	0.03	0.02	0.14	NS
Scale specificity	1	0.01	0.01	0.14	NS
Nature of c'dicted note	1	0.04	0.02	0.13	NS
Task X dist	1	0.69	0.69	6.1	0.05
Task X prior	2	0.52	0.26	5.0	0.025
Task X scale spec	1	0.53	0.53	7.1	0.03

Table 7.1

Summary table for analysis of variance on results of Expt. 2.1

Repeated notes were identified significantly more often than were non-diatonic notes. Subjects were able to identify on average 41% of the note repetitions but identified as ill-fitting only 23% of the out-of-scale notes. Even so, wrong-note detection occurred more often than might have been expected by chance, given that there were seven opportunities for judging a note to be ill-fitting in each sequence (chance rate of 14%). The greater the distance (i.e. number of notes) between the out-of scale/repeated note and the contradicted/original note, the worse the performance. However, this was entirely due to the repetition-detection task. **Table 7.2** shows the mean number of correct responses with varying distance for the two tasks. As the number of notes intervening between the first occurrence of the repeated note and its second occurrence increased, performance worsened (2.46 versus 1.62). However increasing the distance between the wrong note and the contradicted note had no effect on wrong-note detection (1.19 versus 1.20). This underlies the significant task by distance interaction.

	Distance to the contradicted/repeated note	
	2	3
NRDT	2.46	1.62
WNDT	1.19	1.20

	Number of prior notes			Prob.*
	4	5	6	
NRDT	2.25	1.99	1.87	NS
WNDT	0.89	1.36	1.33	NS

	Number of possible scales		Prob.*
	2	1	
NRDT	1.86	2.16	$p<0.05$
WNDT	1.41	0.98	$p<0.01$

Table 7.2

Mean numbers of correct responses in the note-repetition detection task (NRDT) and the wrong-note detection task (WNDT) (maximum=five).

*Significance values for the variables "Number of prior notes" and "Number of possible scales" in **Table 7.2** are derived from the application of an a posteriori studentized range statistic.

Increasing the number of notes prior to the repeated note worsened repetition detection in a monotonic fashion. In contrast, wrong note detection improved with increasing the number of notes prior to the wrong note between four and five, although not between five and six. This led to a significant task by prior notes interaction. Decreasing the number of possible scales (i.e. increasing scale-specificity) made little difference to repetition detection, but worsened performance on the wrong-note detection task from 1.41 to 0.98 correct (out of five). The task by number of scales interaction was significant.

Discussion

Results show that the repetition detection task and the wrong-note detection task involve very different processes. Repetition detection is worsened by increasing the distance between the two occurrences of the repeated note and the number of notes prior to the repeated note but not by the number of possible scales. This indicates that performance of the task is strongly dependent on the demands which it makes of short-term memory (as would be expected from results such as those of Deutsch (1970)) but is largely independent of relatively subtle gradations of diatonic context. Wrong-note detection is unaffected by distance to the contradicted note, is improved (up to a point, see **Table 7.2**) by increasing the number of prior notes and is improved by increasing the number of possible scales from which the preceding sequence of notes could be derived.

The fact that wrong-note detection was unaffected by distance to the contradicted note, while repetition detection deteriorated with distance to the repeated note suggests that wrong-note detection does not depend on memory for specific notes or intervals. Thus subjects apparently did not base judgments on note or interval combinations which sound non-diatonic or peculiar. However, distance was only manipulated within a very small range, and it remains possible that over a wider range of values an effect would be observed.

The improvement in wrong-note detection with increasing number of notes prior to the wrong note only occurred between four and five notes. Performance did not improve from five to six notes. It seems that presenting different notes from a scale in itself strengthened the diatonic schema up to a point, but no further. After five of the seven possible notes had been presented (note that which notes is here independent of how many notes), new notes did not in themselves add information. Moreover, no significant effect was found of whether or not the contradicted or repeated note could serve as a plausible tonic. It would appear either that subjects were unable to identify possible note function from the intervallic context provided or that subjects simply did not differentiate between the potential tonal functions of the notes presented in carrying out either of the tasks.

Wrong-note detection improved with an increase in the number of scales from which the preceding sequence of notes could be derived. That is, as scale specificity increased, performance on wrong-note detection deteriorated. On the surface this result is paradoxical in that it means that subjects apparently gained a stronger sense of diatonicity from notes which were more ambiguous as to the scale from which they were drawn. However, it is important to remember that there is a distinction to be made between scale specificity and scale conformance - that is, between sequences of notes which unambiguously characterise a scale and sequences of notes which may be drawn from a range of scales. Although both the set of notes C, F and G and the set B, F and G occur within the same scale it is only the latter which unambiguously specifies which scale, consisting as it does of the the notes bounding the diatonic scale on C on the circle of fifths (notes B and F) and one other note which falls within the scale on C, the "tritone-plus-one" set. However, given the relative rarity of the tritone as a linear pattern in tonal music the latter set might (in the context of this experimental paradigm) not lend itself to creating a strong sense of diatonic conformance. In view of the paradoxical nature of the scale specificity results two further experiments were carried out in which scale specificity was varied over a wider range of values, as was the distance to the contradicted note.

Experiments 2.2 and 2.3

Problems arise in attempting to vary (i) scale specificity (ii) distance to the contradicted note and (iii) number of prior notes all independently of each other. Recall that to provide an increased range of possible scales from which prior notes can be drawn, the range of spreads of the prior notes within the circle of fifths has to be broadened. Thus to provide an upper boundary to this range of four possible scales and a lower boundary of one scale, the notes used can only be drawn from a spread of from four to seven adjacent notes within the circle of fifths, thus fixing the maximum number of prior notes at four. However, this limits the range of possible distances to the contradicted notes. Similarly, the provision of an increased range for distances to the contradicted notes sets broader limits on the number of prior notes, which interacts with the the range of spreads of the prior notes within the circle of fifths, and hence range of possible scales. This interdependence is not, however, simple and monotonic.

In the following experiments, in order to widen the range of the specificity and distance factors the number of prior notes (notes preceding the wrong note) was held constant. Unfortunately, this meant that the wrong note always occurred at the same point in all sequences, and would have enabled subjects to develop strategies for detection based simply on number of notes rather than on any structural variables. Accordingly, "two experiments in one" were conducted. Each experiment used a fixed but different number of prior notes and the sequences for the two experiments were mixed together in a single session.

Subjects listened to eight-note sequences and indicated when they thought one of the notes did not fit in with its predecessors. As in Experiment 2.1, the eight notes consisted of the seven notes of a given scale and one note which could not be fitted into a given scale together with its predecessors. In Experiment 2.2 there were always four notes prior to the wrong note. Distance to the contradicted note varied between two and three notes. The spread of prior notes within the circle of fifths varied between seven and four (allowing one to four possible scales). In Experiment 2.3 there were always five prior notes. Distance to the contradicted note varied between two and four notes. The spread of prior notes within the circle of fifths varied between seven and five (allowing one to three possible scales). The effect of the number of prior notes on wrong note detection was gained by looking only at sequences of which the parameters were common to the two experiments, i.e. distance varying between two and three notes and spread of prior notes varying between seven and five (allowing one to three possible scales).

Method

Subjects: the subjects were eight undergraduate and postgraduate students. None were studying music at University level. Five had received no formal musical training. Three had at one time played a musical instrument, of whom two had studied music theory at school. None reported being tone deaf or having perfect pitch.

Stimulus material: The note sequences were constructed in an identical manner to that used for Experiment 2.1 except that the parameters were varied differently and "tonal status" - tonic or mediant - of the contradicted note was not specified, as no effect of this parameter had been observed in the previous experiment. For Experiment 2.2 the number of prior notes was four, distance to the contradicted note varied between two and three, and spread of prior notes over circle of fifths varied between seven and four. For Experiment 2.3 the number of prior notes was five, distance to the contradicted note varied between two and four, and spread of prior notes over circle of fifths varied between seven and five.

Procedure: All subjects were presented with all sequences. Sequences of which the parameters were common to both experiments were analysed to determine the effect of number of prior notes. The procedure was identical to that used in Experiment 2.1. Subjects indicated when they thought that a note did not fit in with its predecessors. No repetition detection task was performed.

Results

Table 7.3 gives the ANOVA of the results of Experiment 2.2. It is clear that only one factor had any effect on performance - spread of notes within the circle of fifths. As expected from the results of Experiment 2.1, performance improved with decreasing spread (the mean numbers correct out of a maximum of five were 0.5, 1.1, 1.1 and 2.0 for spreads of 7, 6, 5 and 4 respectively).

Source	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Distance to c'dicted note	1	0.00	0.00	0.03	NS
Spread in circle of fifths	3	1.36	0.45	9.27	0.0005
Distance X spread	3	0.05	0.02	0.43	NS

Table 7.3

Summary table for analysis of variance on results of Expt. 2.2

Table 7.4 gives the ANOVA of the results of Experiment 2.3. As before, the spread of notes within the circle of fifths has a significant effect. The mean numbers correct out of a maximum of five were 0.8, 1.2 and 1.5 for spreads of 7, 6 and 5 respectively. However, distance to the contradicted note also had an effect. There was a monotonic decrease in wrong note detection with increasing distance to the contradicted note. The means of the number of correct identifications in each condition were 1.3, 1.2 and 0.9 for distances of 2, 3 and 4 respectively.

Source	Degrees of freedom	Sum of squares	Mean square	F-ratio	Probability
Distance to c'dicted note	2	0.23	0.11	3.81	0.05
Spread in circle of fifths	2	0.30	0.15	4.29	0.03
Distance X spread	4	0.26	0.06	2.54	NS

Table 7.4

Summary table for analysis of variance on results of Expt. 2.3

Analysis of the combined results of the two experiments revealed that performance was slightly better with increasing number of notes prior to the wrong note (2.5 versus 1.8 out of a maximum of five), but that the difference was not significant, $F_{(1,7)}=2.36$.

Discussion

Experiments 2.2 and 2.3 showed that, as anticipated on the basis of the result of Experiment 2.1, judging diatonic notes as ill-fitting increased as the spread within the circle of fifths of notes prior to the non-diatonic note decreased. In other words, the greater the number of scales within which the prior notes could occur, the more likely that a non-diatonic note would stand out. Judgments of non-diatonicity were affected by how many notes one needed to go back before the non-diatonicity became apparent only when the range of this variable was extended so that it varied between two and four notes (as in Experiment 2.3). There was no consistent effect of the number of notes prior to the non-diatonic note.

The effect of distance to the contradicted note suggests that subjects might be sensitive to non-diatonic note combinations per se (a somewhat different proposition to their sensitivity to similar note combinations in the context of a wide range of stimulus types such as occurred in Experiments 1.1-1.5). However, this effect only appears in Experiment 2.3, over the wider range of values (the distance effect in Experiment 2.1 was due almost entirely to the note repetition detection task), and further work would be needed to determine over what range of values it might have some effect. The effect of the number of prior notes on wrong note detection was inconsistent, and again further work would be necessary to clarify the situation.

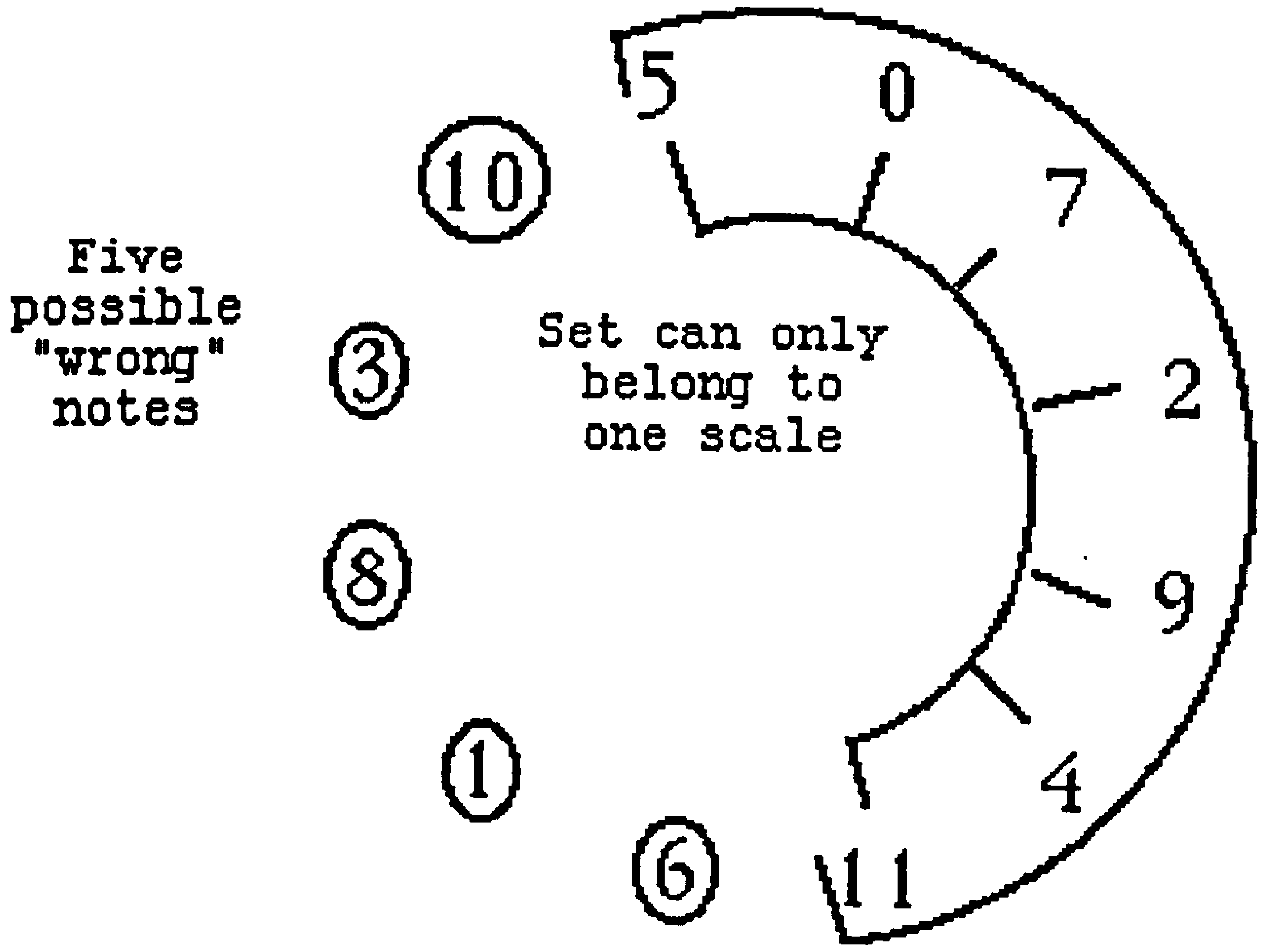
The most striking and consistent finding, and the only finding that appears to depend solely on experimental manipulation of diatonic structure, is that non-diatonic note detection improved with a decreasing spread of notes in the circle of fifths, or decreasing specificity as to which particular diatonic scale the prior notes derived from. It appears that in making judgments of diatonic conformance listeners were not relying on a representation of a single diatonic scale or mode, but were calling on knowledge of some more abstract properties of diatonic interval structure. This runs counter to what one might have intuited; if listeners have access to some representation of diatonic interval structure (as Experimental Series 1 indicates) it would seem likely that the more a sequence conformed to diatonic structure as exemplified within a single diatonic scale, the better listeners would be able to judge out-of-scale notes as ill-fitting. Moreover, the results of Experimental Series 2 appear to be diametrically opposed to the results of a series of experiments conducted by David Butler and Helen Brown (Brown and Butler, 1981: Brown, 1988).

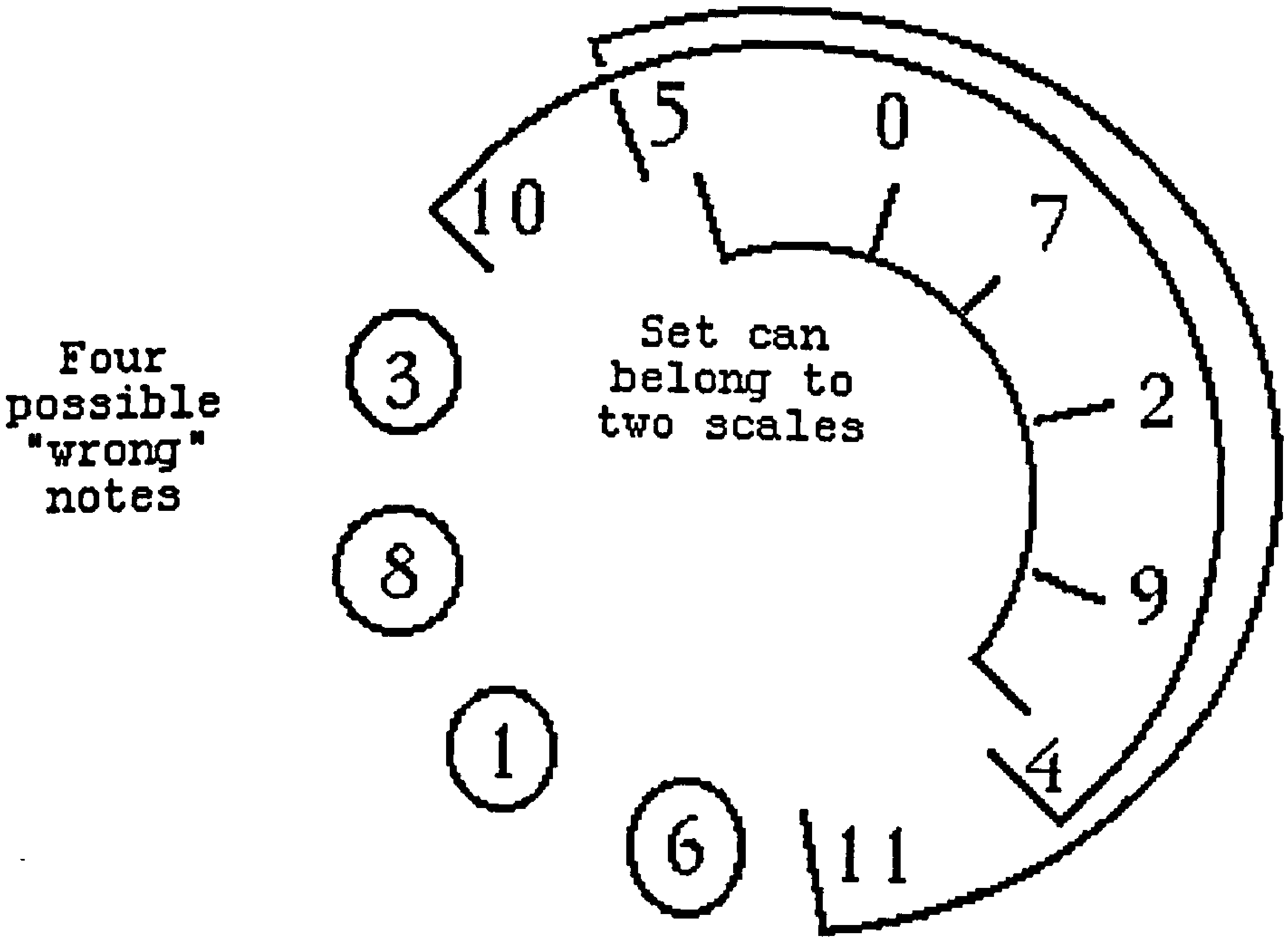
Brown and Butler (1981) showed that subjects exhibited a high degree of accuracy (87%) in producing a note which was the tonal centre or tonic of a key on the basis of hearing a set of three notes - or pc set of cardinality three - which consisted of two notes separated by a tritone plus one other note. As mentioned earlier, such a pc set (e.g. B, F and C) uniquely specifies a single major diatonic scale, and thus has the highest possible degree of scale specificity. In other words, Brown and Butler appear to show that <u>high</u> scale specificity is associated with a sense of "within scale-ness". This is entirely consistent with listeners making use of some cognitive representation of the interval structure within a single diatonic scale.

However, Brown and Butler's subjects showed an even higher level of identification of <u>possible</u> tonics (97%) when presented with a three-note set which had a lower scale specificity, i.e., which could come from more than one scale. This finding could be taken to imply that the cognitive representation of diatonic interval structure can be to some extent modelled by a circle-of-fifths-type representation which encapsulates relations between notes within diatonic scales as well as relations between scales in a compact and lawful way. If the circle-of-fifths is examined, it can be seen that a set of notes which has a narrow "spread" on the circle-of-fifths has a low scale-specificity and may thus be related to a number of plausible tonics; it might, then, be expected that as the number of plausible tonics increases (as scale-specificity decreases) the better a subject is able to select the tonic of a key to which the set of notes is appropriate (see **Figure 7.2**).

A tentative explanation similar to the above could be advanced in the case of the results of Experimental Series 2. As shown in **Figure 7.2**, as the spread of notes within the circle-of-fifths is decreased - that is, as the the scale-specificity of a set of notes is lowered - not only does the number of possible tonics increase, but the number of possible wrong notes <u>decreases</u>. That is, there is a simple and logical relationship between the number of diatonic scales from which a set of notes might derive and the number of notes which cannot fit into any of those scales: as the number of possible scales increase, the number of available wrong notes <u>has</u> to decrease. It might be that this decrease in the number of <u>possible</u> wrong notes with decreasing scale-specificity assists subjects in judging when a note "does not fit". This explanation, like the above, relies on the cognitive representation of diatonic interval structure encapsulating in a fairly detailed way the "simplicity"-modelling characteristics of the circle-of-fifths (i.e., the way in which different scales share different numbers of notes and exclude others).

Figure 7.2:Relationship between spread of a set of notes on the circle-of-fifths, the number of possible scales from which it may derive, and the number of available "wrong" - out-of-scale - notes.





However, there are a number of important differences between the experimental methods employed in Experiments 2.1-2.3 and in Brown and Butler's study. Firstly, their study required subjects to identify the tonal centre or "key note" of a scale, or more properly, a <u>mode</u>. Experimental Series 2 differed in that subjects were required to detect out-of-scale notes. These notes had to be out-of-scale with respect to any possible diatonic scale permitted by the preceding notes, and so did not necessarily test subjects' derivation of a <u>single</u> unique scale or mode; it is worth noting that subjects in Experiment 2.1 above seemed insensitive to manipulations of the potential tonics of the modes implicit in the sequences used. That is, while the present study was aimed at testing the degree to which diatonic interval structure could cue a sense of diatonic conformance (posing the question of how a set of notes might give a sense of being connected with other notes in some broader set), Brown and Butler's study was aimed at testing how interval structure could cue a sense of <u>specific</u> mode (i.e. asking how a set of notes might be related in perception to <u>one</u> particular note within a broader set). The question asked of their subjects is not one that simply or necessarily concerns a generalised sense of diatonic conformance.

Secondly, Brown and Butler used formally-trained musicians in their study; these subjects would be more likely to have developed conscious strategies for the tasks required than would the musically-untrained subjects used in Experiments 2.1-2.3, perhaps on the basis of having encountered the circle of fifths as a representation of tonal relations in the course of their formal training. Indeed, it would not have seemed sensible to have required the subjects used in Experimental Series 2 to identify the "tonal centre" implied by the preceding notes as the very concept of "tonal centre" is likely to be encountered explicitly only in formal musical training.

Despite these differences of method and aim. it is of considerable interest to note that Brown and Butler's results show that manipulations of diatonic interval structure may influence subjects' identification of mode. As described earlier, mode or key is a level of pitch organisation generally associable with aspects of temporal rather than purely alphabetic structure. In fact, Brown's (1988) study constitutes a thorough examination of the ways in which temporal order and interval structure may interact in the identification of a tonal centre. The relationship between the results of these studies and the results of the current series of experiments will be evaluated when an appropriate theoretical framework has been outlined and the current results have been further investigated.

For in spite of the tentative explanations offered above, the question of the basis for the improvement of listeners' ability to detect wrong notes in the context of decreasing scale specificity remains unresolved. Before discussing a further series of experiments designed to investigate this finding, it would seem helpful to review the types of cognitive activity that might underlie listeners' responses in Experimental Series 2 and in Brown and Butler's study within a broader context provided by an account of the types of cognitive activity which could be associable with a representation of diatonic interval structure. The elements of such an account are sketched by Browne (1981).

CHAPTER EIGHT

Rarity and ubiquity - Browne and the Diatonic Set

Browne is concerned to examine several aspects of tonal music which appear paradoxical; for example, the fact that it is extraordinarily complex if analysed, yet we seem to "process it in real-time" with apparent ease, or the fact that our perceptions of it seems to involve a sense of duality or ambiguity, a sense of "some things being more important than others" or of hierarchy and a sense of process. He does this by characterising the diatonic scale in group-theoretic terms as the diatonic set and outlining certain properties of that set (in a manner similar to that of Balzano), and then by focusing on how such properties might function in "the process whereby the various elements of the diatonic set are hierarchized, associated and differentiated on the basis of their external referents, their contextual surroundings and the content of those surroundings" (1981, p 20). He suggests that this process is describable in two ways: position finding, and pattern matching.

Browne describes position finding as involving questions that appear to concern what could be called orientation and reference: he suggests that questions such as "where are we" (in respect of tonal space) and of "how long should I note what happened in terms of x" are characteristic of position finding processes. So the identification of the tonic (or other "structural" note) within a passage, or the tracking of modulations between tonal regions could both be described as position finding processes. Pattern matching, on the other hand, involves questions of similarity and difference, such as "is this the same as that ?", or "could this become more or less like that ?". Pattern matching is interesting only if patterns don't match perfectly. Browne's thesis is that the structure of the diatonic set appears particularly suited to these types of processes.

He characterises the diatonic set within the framework suggested by Forte (1973), that is, within the dihedral group of order 24. This means that two (unordered) sets are regarded as equivalent if they are of the same cardinality (have the same number of members) and if one can be transformed into the other by either the operation of transposition or by the operation of inversion or by both operations together. Thus the pc set (0,1,3) - in note-names, C, C#, D# - is equivalent to the pc set (3,4,6) (or D#, E, F#) by the operation of transposition and is equivalent to the set (0,2,3) (or C, D, D#) and to all of its possible transpositions by the operation of inversion. The diatonic set characterised in this way can be shown to contain only a finite number - 119 - of subsets, that is, distinct pc sets which can be formed between members of the diatonic set; these can be categorised (via transposition and inversion) into 42 mutually irreducible pc sets. Within Forte's formulation, a pc set has an associated level of description in terms of the intervals that can be formed between each and all of the set's elements; this level of description is the set's interval vector.

In this framework, the diatonic set in normal order is the set (0,1,3,5,6,8,10) (set number 7-35 in Forte's catalogue), having the interval vector <254361>. Each entry in the interval vector indicates the number of times that each of the six different intervals available within the Fortean framework can be formed between members of the set. There are only six different intervals available because within this framework complementary intervals are taken as equivalent (under inversion); M7 is thus "read" as m2 and the tritone is thus the "largest" formable interval. The diatonic set's interval vector means that between the set's members the interval of a minor 2nd (m2) can be formed - or occurs - twice, M3 occurs five times, m3 four times, M3 three times, P4 six times and the tritone once. Thus every available interval occurs between members of the diatonic set, and each interval occurs a different and unique number of times. Thus the diatonic set possesses the property of <u>unique and complete interval multiplicity</u>, a property which makes it - within the <u>context</u> of the dihedral group of order 24 - a <u>deep scale</u> (Gamer, 1981). Only one other subset of the dihedral group of order 24, the set 7-1, consisting of the pcs (0,1,2,3,4,5,6) also has this property.

Possession of this property of unique interval multiplicity confers several other properties. It means that there is <u>no</u> interval that is not contained in the set - there is no single interval that would be a "marker for leaving the diatonic set". Also, no transposition of the set uses all new notes; some members must be shared between the original set and its transposition. Moreover, each interval class within the diatonic set occurs a unique number of times, giving a continuum of intervals from diatonically rare to diatonically common. So intervals such as tritones (1 occurrence) and semitones (two occurrences) are rare, while perfect fourths (6 occurrences) and major seconds (5 occurrences) are relatively common. In Browne's view rare intervals are likely to aid position finding, while common intervals will aid pattern matching. This arises by virtue of the fact that each interval, or more generally, each subset of the diatonic set, possesses not only a specific content (in terms of intervals) but also a specific context (i.e. the relationships that may be formed between that set and the other subsets of the diatonic set, that is, the set's surroundings). Browne claims (p 5) that "pitches and subsets have context, and derive their musical functions from their relations with their context as well as from their intrinsic content".

For example, as a minor second may only occur twice within the diatonic set it is a reasonably good cue to position within that set; a minor second may only occur within the context of the diatonic set between two particular pairs of its members, the leading-note and tonic and the mediant and subdominant. The tritone is an even better positional cue; it can only occur between the leading-note and subdominant. The perfect fourth, on the other hand, is a very bad logical cue to position within the set, as it may occur between up to six pairs of set members; it is, however, a useful element in matching possible patterns, into which category of activity Browne puts the process of tracking passages comprising "sequence" or literal/non-literal imitation within the set.

For instance, a musical passage consisting of a run of perfect fourths is possible within the set but at some point has to change into a tritone to stay in the same scale. Each repetition of the P4 matches its predecessor, until - to stay in the same scale - the interval has to change from a perfect fourth to an augmented fourth. It is the change to the tritone which provides the cue as to which diatonic scale set or key underlies the passage (in Browne's words, "rarity controls position finding by controlling pattern matching", and thus pattern matching involves the ability of elements of the diatonic set "to do the same thing without actually doing the same thing").

Although Browne does not dwell on this point, incorporated in this idea of pattern matching is the implication that the run of perfect fourths - i.e., of diatonically common intervals - is in itself a good cue to the fact that underlying the overall pitch organisation of the passage is some level of diatonic structure. That is, the high degree of literal repetition of diatonically common intervals embodied in such a passage might serve as a cue to the diatonic conformance of such a passage for a listener simply because these intervals such common features of the diatonic set. In other words, while rare aspects of the diatonic set may subserve functions involving differentiation and hierarchization of notes within the set (through position finding), common aspects of the diatonic set may subserve functions involving association and integration of notes as members of the set (through pattern matching).

What holds for intervals within the diatonic set might also be expected to hold for other components of the diatonic set, i.e. for its other constituent subsets of cardinality greater than 2 (intervals being, in fact, pc sets of cardinality 2), if these can also be shown to possess properties of different rarity or commonality. The diatonic set contains forty-two different types of pc set; each pc set may be formed a number of times (may be formed with a different multiplicity). As an example, the pc set (0,1,3), or the set 3-2, occurs between the notes E, F and G, between B, C and D and (in its inverted form as 0,2,3) between D, E and F as well as between A, B and C; it has a multiplicity of four. The distribution of pc sets as subsets of the diatonic set is not unique, that is, each pc set does not manifest itself a distinct number of times (e.g. the set (0,1,5), or 3-4, also has four formations within the diatonic set).

This means that the rareness or ubiquity of a pc set within the diatonic set cannot serve by itself as an unambiguous guide to position within the set. However, each subset will also have a complementary set or sets associable with it, comprising the diatonic pitch classes which it itself does not contain; the relation between a pc set and its diatonically complementary set or sets may provide a context of possible formable intervals which serve to differentiate each set, as well as each manifestation of each set. Browne states (p 14) that "Any subset of the set is characterisable by, is differentiated from, derives its uniqueness and therefore its function from, not just its intrinsic content but from its contextual relation with its complement". Moreover, although Browne does not bring out this point, each subset has associated with it an interval vector that either shares or does not share particular characteristics with the interval vector of the diatonic set, i.e., common or rare features of the diatonic set are differently shared by its different subsets.

As an example of the way in which the multiplicity, interval vector content and intervallic context of diatonic subsets might function in position finding, consider the following case. The pc set 3-5, which occurs in only two forms within the diatonic set (comprising E, F and B or comprising B, C, and F) and of which the pc content is (0,1,6), has an interval vector that permits the formation of one minor second, one perfect fourth and one tritone; it contains two of the rarest intervals within the diatonic set's interval vector, and might be expected - on the basis of rarity of formation and possession of rare intervals - to be a relatively good cue to position within the diatonic set (i.e. a good cue for position finding). Moreover, the contextual relation of each manifestation of the set 3-5 to each manifestation of its complementary set 4-23 differentiates each form of the set 3-5 on the basis of the intervals that may be formed between the notes in each manifestation of 3-5 and 4-23. That is, as the intervallic contexts of each form of the set 3-5 are different, its already high value for judging position within the diatonic set should be further enhanced.

Conversely, consider the potential of the set 3-9 as a functional cue in pattern matching; this set occurs in five forms within the diatonic set, has the pc content (0,2,7) and has an interval vector that permits the formation of one major 2nd and two perfect 4ths. It therefore contains three of the more common intervals within the diatonic set's interval vector, and might be expected - on the basis of ubiquity of formation and possession of common intervals - to be a relatively good cue for diatonicism itself. That is, the pc set 3-9 constitutes in itself a common feature of the diatonic set (by virtue of its high multiplicity), and it shares particular interval vector characteristics with the diatonic set to the extent that it may be easily associable with it in perception and memory, thus serving as a good cue in pattern matching processes. It can still, however, be of use in position finding because each of its five possible formations within the diatonic set has a different intervallic relation to the appropriate form of one of the complementary sets (in this case, exemplars of the pc sets 4-23, 4-16 or 4-8).

Experimental Series 3

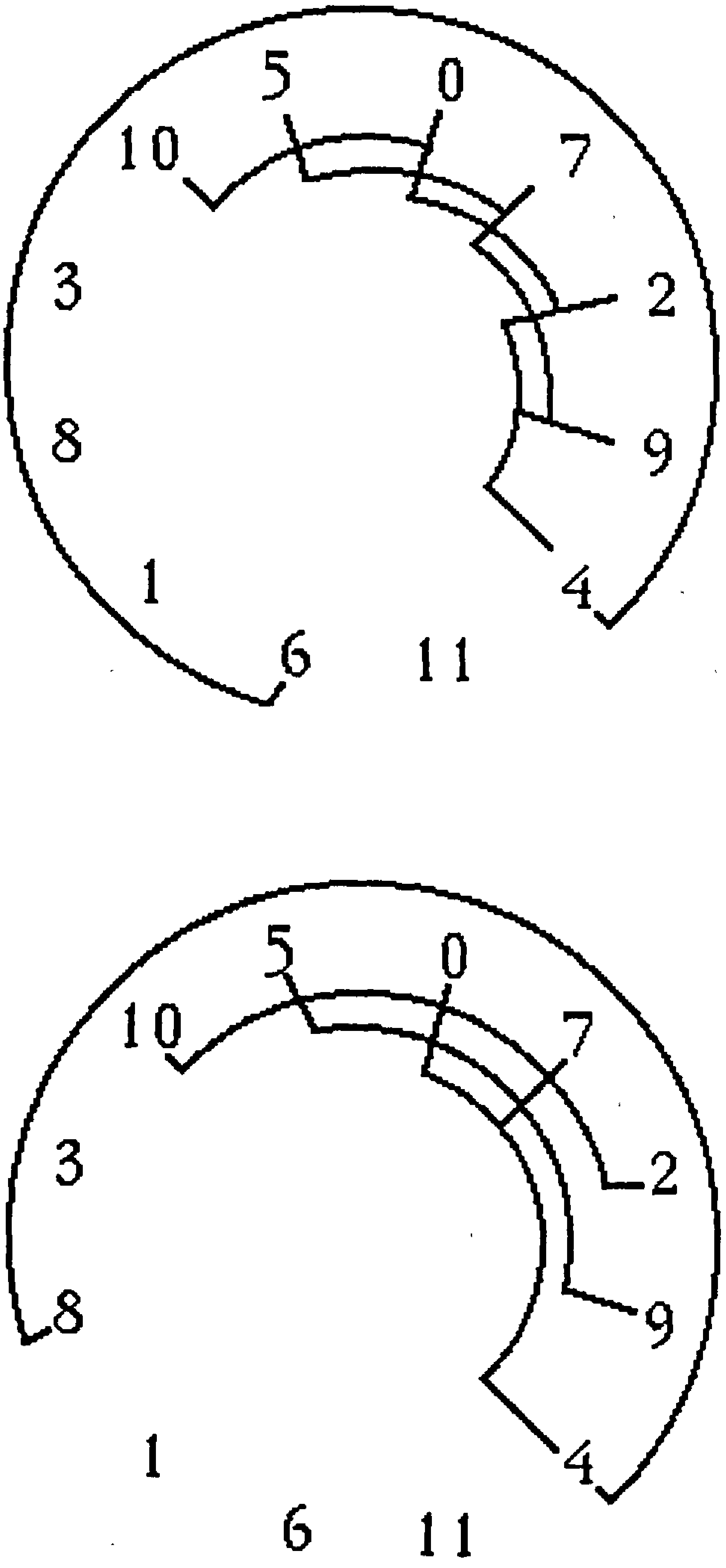
The relevance of the foregoing to the findings of Brown and Butler (1981) is obvious. Their experiments found that pc sets of cardinality three which had a low multiplicity within the diatonic set and of which the interval vectors possessed a diatonically rare interval in the form of a tritone served as the most unambiguous cues in the identification of appropriate tonal centres for musically trained listeners. That is, listeners were most accurately able to carry out a paradigmatic position finding task - which seems likely to have required the identification of a set and a differentiation between notes of the set - in the context of subsets of the diatonic set that possessed rare intervals, were themselves rare within the set and were thus highly specific with respect to diatonic scale.

The relevance of Browne's exposition to the findings of Experimental Series 2 may not be so apparent. To recapitulate, in this series of experiments the lower the diatonic scale specificity of the context set the better subjects were able to detect wrong notes: that is, the lower the specificity, the better subjects' sense of diatonic conformance. Judgments based on a sense of diatonic conformance appear to fall into the domain of pattern matching, involving as they do questions of similarity such as "is this more-or-less like that ?" and - although not explicitly invoked by Browne - questions of association or connectedness, such as "can this be part of the same thing as that ?".

Within the domain of pattern matching Browne implies that diatonically common intervals and a high multiplicity of subset formation within the diatonic set are likely to be functional. As shown in **Figure 8.1**, the lower the diatonic scale specificity of a pc set, the more closely bunched it appears on the circle-of-fifths and the greater the multiplicity of the set within the diatonic set. That is, it would appear that decreasing the spread within the circle-of-fifths of the "prior notes" in Experimental Series 2 not only increased the number of scales from which the sets could derive but also increased the commonality or multiplicity in respect of the diatonic set of the pc sets to which the prior notes belonged. It could be, then, that subjects' detection of wrong notes - or judgments of diatonic conformance - was dependent on the commonality of formation within the diatonic set of the pc set to which the notes prior to the wrong note belonged; however, the earlier hypothesis - that the increase in detection rate of wrong notes was due to a decrease in the number of possible wrong notes with decreasing scale specificity - cannot be ruled out. A further series of experiments was therefore carried out to test these hypotheses.

The experiments in this third series used several different response formats and different stimulus material in order to

Figure 8.1 : The relationship between "bunched-ness" on the circle-of-fifths, scale-specificity and multiplicity of pc sets of the (0,2,7)-type (upper) and (0,4,7)-type (lower)



establish some degree of generality in the phenomena. However, fundamentals of procedure remained the same throughout. This differed from that used in Experimental Series 2 in that subjects performed a rating task rather than a wrong-note detection task after hearing a brief sequence of notes. The object of this series of experiments was to test the effect of using extreme values for the diatonic scalar specificity or diatonic multiplicity of the brief sequence of notes (context sets) on judgments of diatonic conformance. As pointed out in connection with Experiments 2.2 and 2.3, this factor will correlate at the lower limit of scale specificity with the number of notes that are available to form the context sequence. If these contexts are to be minimally scale-specific (with a diatonic multiplicity of five) they will always contain only three notes; a wrong-note detection task would be inappropriate as the "wrong note" would always occupy position four in the sequence, so a ratings task is used.

Subjects listened to three-note sequences which were differentiated on the basis of their pitch class content alone (i.e. temporal orderings were random); for each three-note sequence subjects indicated how well they thought a fourth, probe, note fitted in with the preceding three. Experiment 3.1 set out to demonstrate that ratings of fittingness of probes were sensitive to manipulations involving diatonic conformance and could be used to gauge the sense of diatonic conformance produced by the preceding three-note sequences. Experiments 3.2-3.6 dealt with the main issues of possible associations between pitch relations and sense of diatonic conformance.

Experiment 3.1

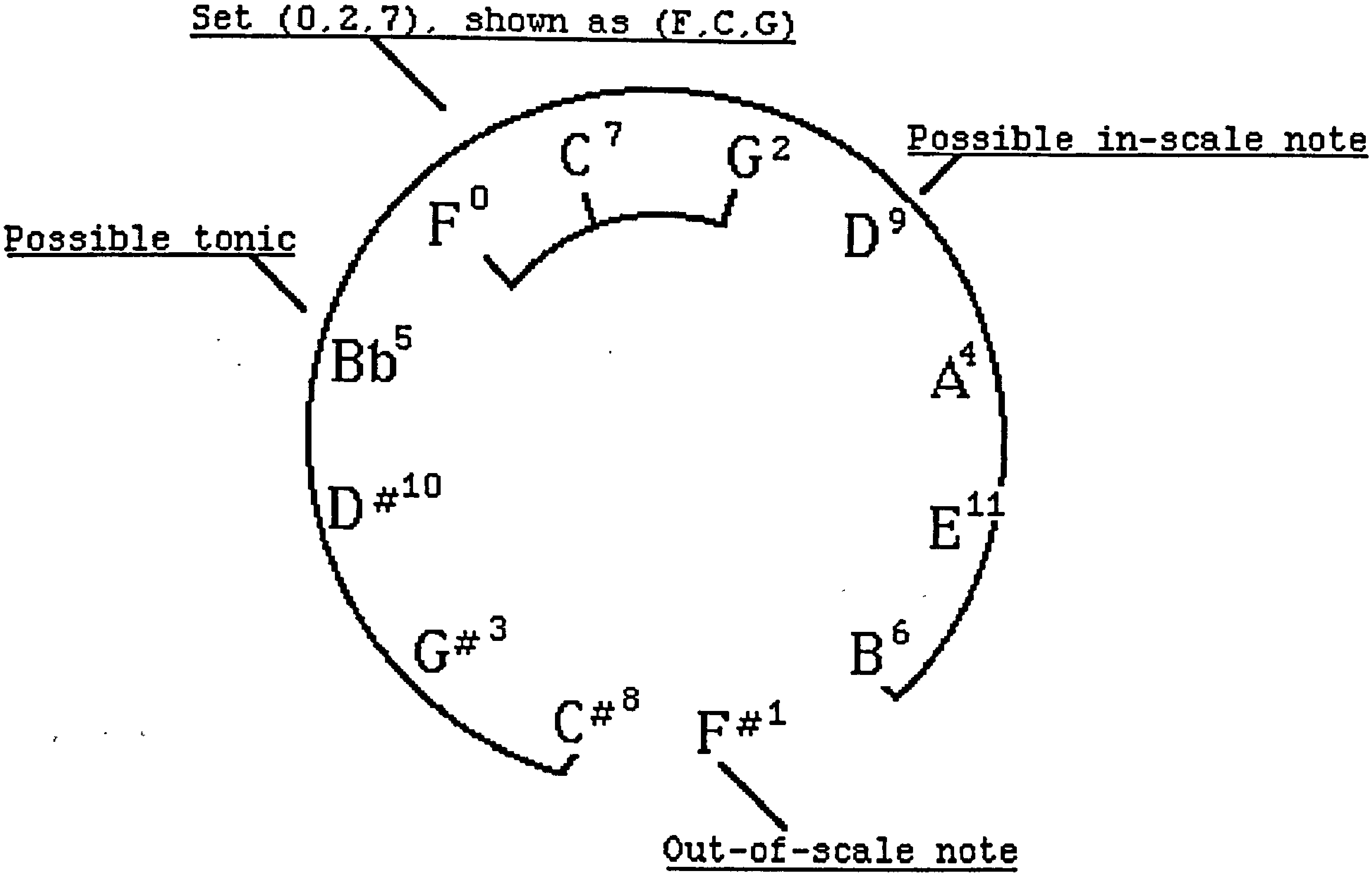
This experiment was intended to test whether ratings of "fittingness" would be sensitive to whether the fourth note (the probe) was in or out of scale with a diatonic three-note sequence, and at the same time establish whether a three-note set which had low scale-specificity because of its minimal spread on the circle-of-fifths (i.e., which could derive from five different scales, and had a high diatonic multiplicity) would be sufficient to evoke a sense of diatonicism for listeners with no formal musical training. Also of interest was whether in-scale probes would be judged as better-fitting if they were potential tonics for the possible scales allowed by the three-note sequences. Although no effect of potential modal status on judgments of diatonic conformance had been noted in the results of Experiment 2.1, it was felt that the possibility of such an effect arising through the more restricted nature of the contexts used in the current experiment should be tested. The relationship between the potential tonics, the three-note set, its spread and multiplicity, and the diatonic set can most easily be grasped by reference to **Figure 8.2**.

Method

Subjects: Subjects were 12 undergraduate and postgraduate students at the University of London. Most had received some tuition on musical instruments at an earlier age, but none had taken their study of music beyond elementary level.

Stimulus material: The three-note sequences were random orderings of the pc set (0,2,7), presented at random transpositions (to give, e.g. F, C, G: D, E, A: C#, B, F#: etc.). Maximum interval

Figure 8.2: The relationship between set-type and possible probe notes in Experiment 3.1.



size between any two consecutive notes was set at a perfect fifth. This type of three-note set was chosen as its members are most closely bunched on the circle of fifths. This means that a given exemplar of the set may derive from any of five diatonic scales, the set-type has a multiplicity of five (can be formed within a single diatonic set at five positions, (see **Figure 8.2**) and there is only one possible pitch class - (1) - that cannot fall into a diatonic set within which (0,2,7) occurs (e.g., if the pc set is read as the notes C, D and G, the only note which cannot occur together with these within a diatonic major scale is the note C#).

Each three-note sequence was followed by a fourth, probe, note. If the first three notes were F,C and G the fourth note was either D, Bb or F#. Similarly, any (re-ordered) transposition of the above three-note sequence would be followed by a similarly transposed probe note (for example, the notes D-A-G could have as a probe either E, C, or Ab depending on the condition). The notes D and Bb were chosen to be scale conformant, while F# was out-of-scale. The Bb was a potential tonic for the three-note set and the D a potential mediant i.e. occupying the inverse position on the circle-of-fifths to the Bb in relation to the three-note set).

The range used was the same as that used for Experiments 2.1-2.3, as was the spectral makeup of the tones used. The tones (generated on Fairlight CMI and digitised into a PDP-11 as previously described) had, in Experiment 3.1, a duration of 0.34 sec; the inter-note interval in the three-note sequences was 0.16 sec. and the interval to the probe was 2 sec. There was a 6 second pause between trials.

Procedure: The subjects sat in a sound-treated cubicle and listened to the sequences through headphones. They were instructed to listen to each sequence and rate on a seven-point scale how well they thought the fourth note fitted with the preceding note (1=very badly, 7=very well). The independent variable was whether the probe was in or out of scale, or if within the diatonic scale, whether it was a potential tonic. There were five trials for each condition, making fifteen trials in all, the order of which was randomised. The mean rating for each of the three conditions was calculated for each subject.

Results

The out-of scale probes were rated as less well fitting with the diatonic three-note sets than were the in-scale probes, $F_{(2,11)}=12.3$, $p<0.01$. The mean ratings were 4.3 and 4.1 for D- and Bb-type probes respectively, compared with 3.4 for the F#-type probes. Effectively, no effect was observed of which of the in-scale probes was used. Thus the potential tonic probe was not rated as fitting better with the preceding three-note set than the other in-scale probe.

Experiment 3.2

The result of Experiment 3.1 appears to indicate that ratings of note-fittingness are influenced by diatonic conformance of the context set; it would also appear that subjects were not sensitive to the (admittedly limited) modal implications of the probe tones in making judgments. However, the result simply indicates that diatonic conformance was operational for subjects' judgments; it does not indicate that the high diatonic multiplicity and low scale specificity of the sequences used was in fact functional. A further experiment was therefore conducted which used two extreme values for the multiplicity/specificity factor of the context sequences, in order to confirm the hypothesis that high multiplicity and low specificity would contribute to a strong sense of diatonic conformance. According to the results of the Experiments 2.1–2.3, the narrower the spread – i.e. the lower the scale-specificity and the higher the multiplicity – the better the sense of diatonic conformance. It was therefore anticipated that the pc set used in Experiment 3.1 would show an advantage in differentiating between in-scale and out-of-scale notes in the ratings task over the tritone-plus-one-type sets such as (0,1,6) (or B, C, F) used by Brown and Butler (1981), which are <u>maximally</u> scale-specific and have low multiplicities.

<u>Method</u>

The subjects were the same as in Experiment 3.1. The stimulus material and procedure were similar except that the three-note context sets were of two general types and a single type of probe was used. The context sets were <u>either</u> the pc set (0,2,7) <u>or</u> one of the pc sets (0,1,6), (0,2,6) or (0,3,6). Thus the context sets consisted either of three adjacent notes in the circle of fifths such as F, C and G (minimal scale specificity, high multiplicity) or of a tritone plus one other note (in the diatonic scale on C, the notes F, B and either C, G, D, A or E), giving maximum scale specificity and low multiplicity. The probe was taken as the note furthest away from the maximally-compact pc set on the circle of fifths (i.e., if the maximally-compact pc set forms the notes F, C, G, the probe note would be F#). As before the order of notes in each three-note sequence was randomised and the sequences and probes presented at random transpositions. The subjects listened to six sequences of each type, randomly ordered, rating the degree of fittingness of the probe as before.

Results

The context set (0,2,7), which had high multiplicity and low scale specificity, gave a greater sense of diatonic conformance than did the low-multiplicity scale-specific types of sets, in the sense that the out-of-scale probe was rated as less well fitting when preceded by an ambiguous than by a scale-specific sequence (means 3.1 versus 3.9, $t(11)=3.3$, $p<0.01$)

These results would seem to indicate that a high multiplicity and low specificity was associated with a greater sense of diatonic conformance; this is in line with the results of Experiments 2.1-2.3, which indicate that detection of out-of-scale notes increased with increasing scale ambiguity. However, the results do not indicate clearly the processes whereby multiplicity and specificity might function in the cognition of diatonic conformance.

Discussion

In the previous two experiments, and in the discussion of the results of Experimental Series 2, it has been taken as read that high diatonic multiplicity and low scale-specificity are directly correlated. It certainly appears that decreasing the spread of a set of notes within the circle-of-fifths both increases the number of scales from which that particular set could derive and also increases the commonality or multiplicity of the pc set-type involved. So subjects' ratings of out-of-scale notes could well depend on a general factor that manifests itself as both multiplicity and specificity. The cognitive activity involved could thus be of the pattern-matching type - wherein "like" (a set-type which is representative of the diatonic set by virtue of its high multiplicity and possession of common diatonic intervals) is cuing "like" (diatonic set), or it <u>could</u> involve something more akin to a form of <u>position-finding</u>, involving questions such as "where are we ?" (in respect of some representation of "tonal space"). In the former case, the high multiplicity of non-scale-specific sets could cue a specific diatonic set or even evoke several closely-related diatonic sets; listeners' judgments of fittingness would be based on which notes were available within the single scale or set of scales. In this case, the representation of diatonic structure underlying subjects' responses need not incorporate any detailed and rigorous depiction of relations between different diatonic sets (other than, perhaps, closely-related sets).

In the latter (position-finding) case, a more complete representation of inter-scale relations is likely to be necessary. Recall that as scale specificity decreases, the number of possible out-of-scale notes also decreases. As suggested earlier, it might be that this decrease in the number of <u>possible</u> out-of-scale notes with decreasing scale-specificity assists subjects in judging when a note "does not fit". If subjects had access to some cognitive representation embodying the precise relations between diatonic sets within the chromatic set, they might be able to monitor the relative "positions" within the representation which non-scale-specific sets and out-of-scale notes occupy rather more easily than would be the case with more highly scale-specific sets which permit of a larger number of out-of-scale notes. This representation need not take an "imagistic" or spatial form; inclusion and exclusion from diatonic sets could be represented equally well in propositional terms.

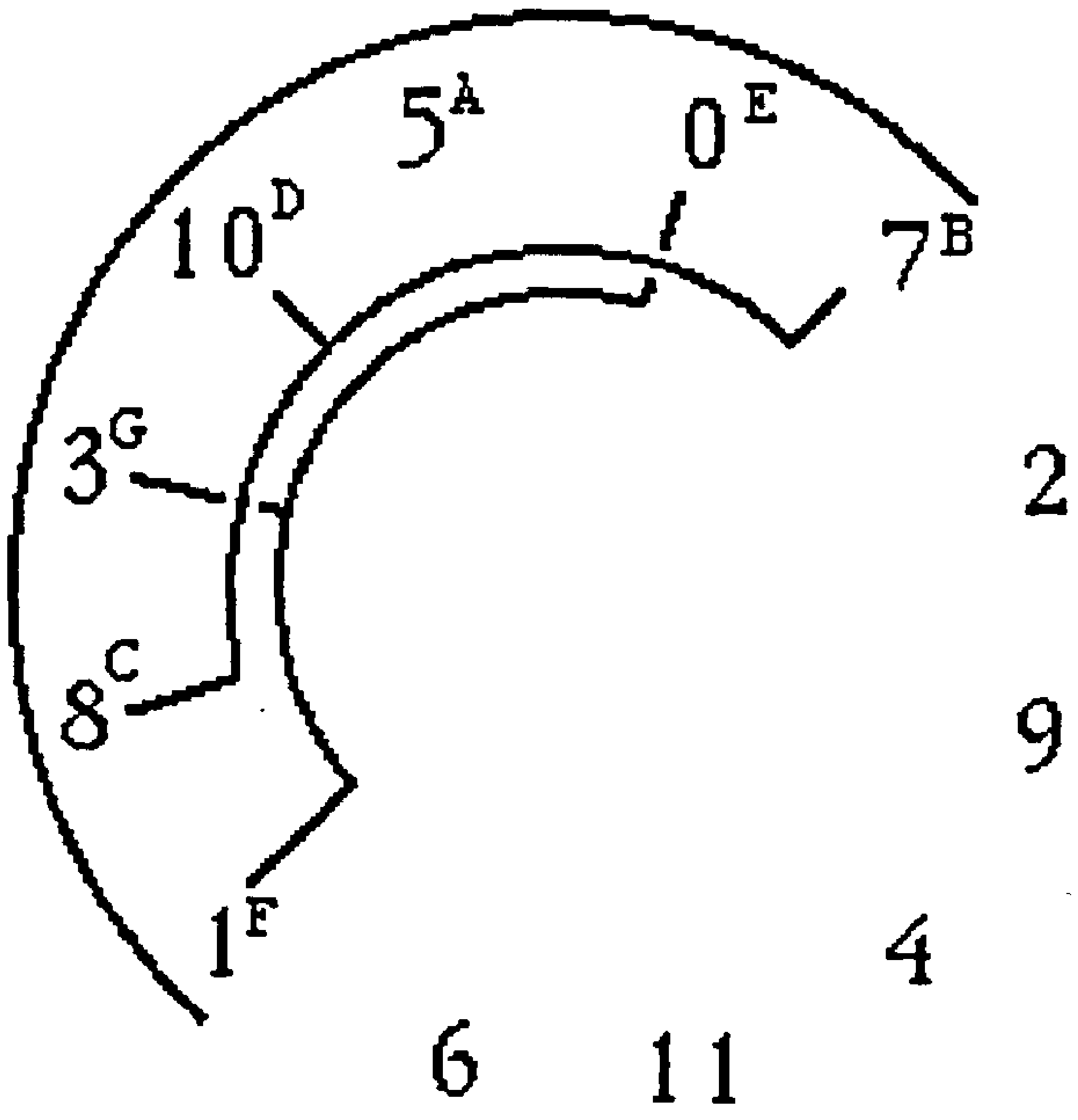
Multiplicity and Equivalence Relations

Thus two possible mechanisms could underlie subjects' judgments of fittingness, as diatonic multiplicity and number of possible out-of-scale notes (scale-specificity) appear to be directly linked. As circle-of-fifths spread decreases, multiplicity appears inevitably to increase, so that, e.g., the set 3-9, with a pc content of (0,2,7), has the most "compact" circle-of-fifths spread of any three-note pc set, has thereby minimal scale-specificity and permits of fewest (one) out-of-scale notes and has a multiplicity of five within the diatonic set. There appears to be no possibility of independently varying specificity and multiplicity, and hence no likelihood that the nature of the process underlying subjects' responses can be clearly understood. However, the necessity of this linkage of spread and multiplicity only holds if equivalence between pc sets is based solely on transposition, as in Balzano's group-theoretic description of pitch relations (1980, 1982). If pc set equivalence is based instead on transposition together with inversion - as in Browne's (1981) characterisation of the diatonic set and its subsets - this link does not hold.

As an example of how changing the basis for set equivalence changes the relation between spread and multiplicity, consider a three-note set such as (0,1,3), which could represent the notes E, F and G (see **Figure 8.3**). This has a spread of six notes on the circle-of-fifths, can belong to two different scales, and allows of four possible out-of-scale notes. If equivalence between pc sets is based on transposition alone, the only other exemplar of this set-type within the diatonic set containing the notes C, D, E, F, G, A, and B would be formed by the notes B, C and D; the multiplicity of the set-type (0, 1, 3) would be two. If equivalence between pc sets is based on transposition together with inversion, sets having the content (0,2,3) - such as D, E and F - are no longer considered distinct from (0,1,3)-type sets; exemplars of the (0,2,3)-type sets are therefore "absorbed" into the class of (0,1,3)-type sets, increasing the number of exemplars of that class which are to be found within the diatonic set from two to four. The multiplicity of the pc set (0,1,3) is thus two under transpositional equivalence but is four under transpositional and inversional equivalence, yet its spread within the circle-of-fifths - and hence its scale-specificity and number of possible out-of-scale notes - remains unchanged.

Consider a further three-note set, with the content (0,4,7), - the notes C, E, and G (see **Figure 8.4**). This has a spread of five notes on the circle-of-fifths, can belong to three different scales, and allows of three possible out-of-scale notes. Under transpositional equivalence, there are two other exemplars in the above diatonic set, consisting of the notes F, A and C and the

Figure 8.3: The multiplicity of sets of type (0,1,3) under transpositional equivalence (upper circle) and transpositional <u>and</u> inversional equivalence (lower circle).



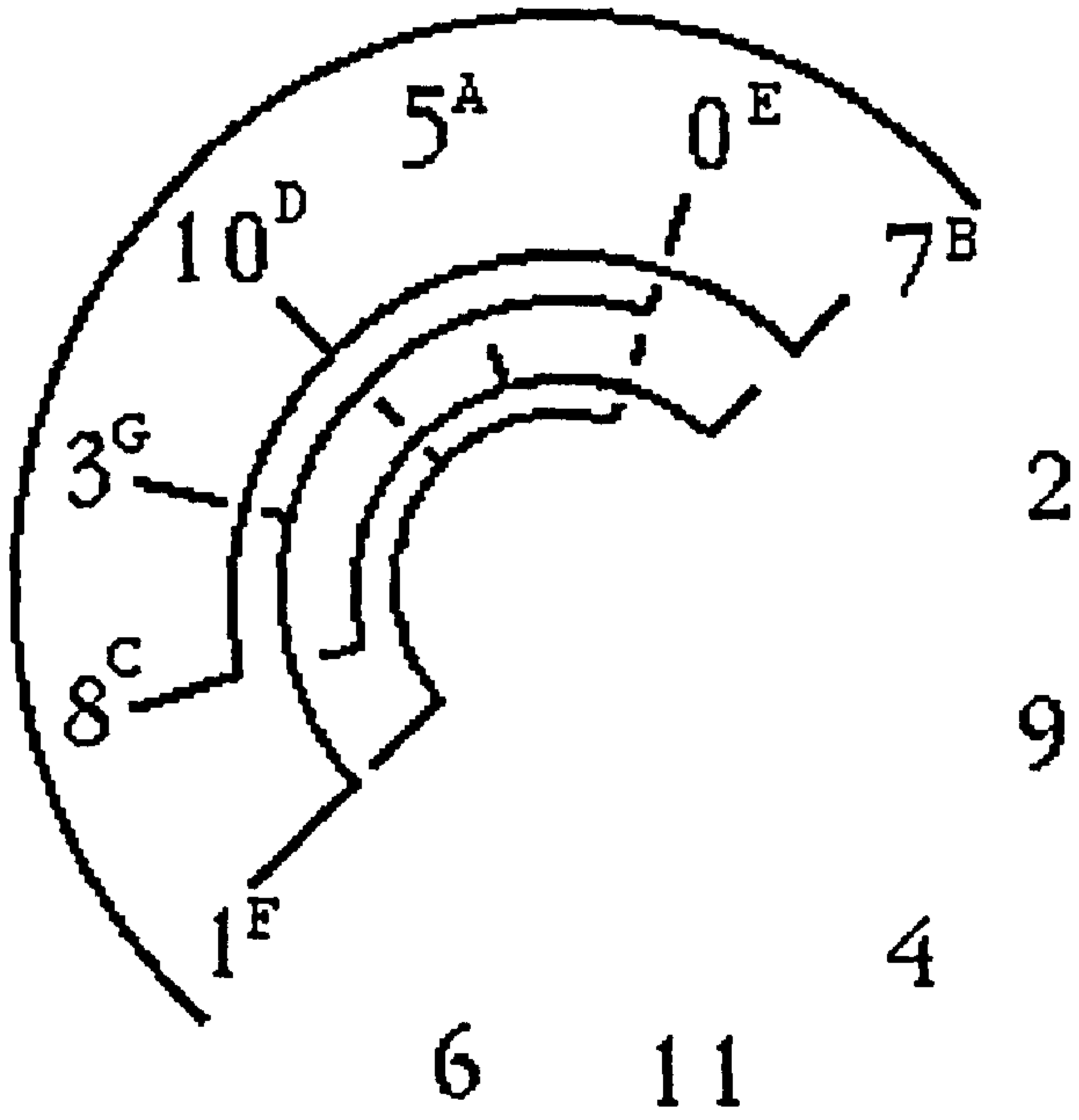
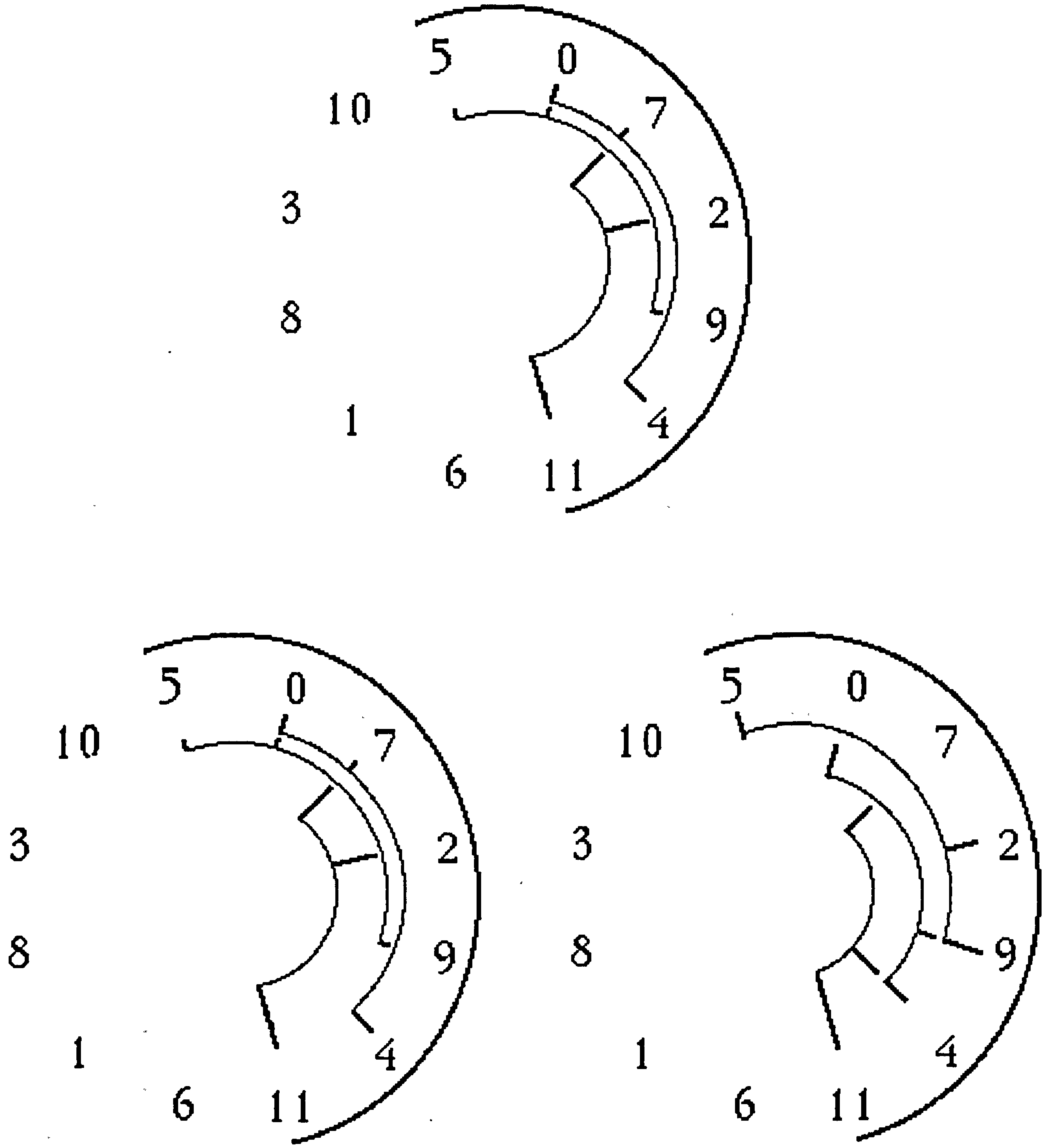


Figure 8.4:The multiplicity of sets of type (0,4,7) under transpositional equivalence (upper circle) and transpositional and inversional equivalence (both lower circles).



notes G, B, and D; the multiplicity of the set-type would be three. If transpositional and inversional equivalence is used, sets having the content (0,4,7) are considered to be exemplars of the pc set (0,3,7). The multiplicity of the pc set (0,4,7) is thus three under transpositional equivalence but is six under transpositional and inversional equivalence, with (again) unchanged spread.

It may seem ludicrous to equate the set of notes C, E, and G, the major triad with the set A, C and E, the minor triad, as is implicit in the (0,3,7)-(0,4,7) equation. After all, these two types of note-formations are generally quite distinct in the musical functions that they subserve. However, if it is admitted that apparent inversional equivalence of intervals may at times play some role in musical cognition - as it appears to do in the "fourths/fifths" confusion which music students often display in listening tasks - then the possibility cannot be ruled out that equivalence relations between pc sets may be best considered to be founded on transpositional and inversional equivalence for some musical situations, or rather, for some types of musical-cognitive activity. The idea that equivalence between pc sets is governed only by transposition has been implicit in the design and analysis of all of the experimental material used in Experimental Series 2 and Experiments 3.1 and 3.2, and has not been explicitly controlled for or tested. As adopting this equivalence basis seems to offer a way of disentangling scale-specificity and multiplicity factors, it seems appropriate to explore the issue experimentally.

Experiments 3.3 and 3.4

Browne's paper shows that under transpositional and inversional equivalence the set with the highest multiplicity of any diatonic subset is the set 3-7, with the content (0,2,5); this set - of which exemplars could consist of the notes A, C and D or the notes A, B and D - has a multiplicity of eight. It has a spread of four on the circle-of-fifths, can thus belong to only four different diatonic scales and permits of two out-of-scale notes. Under transpositional and inversional equivalence, then, the set (0,2,5) has a wider spread than the set (0,2,7) and permits of more out-of-scale notes, yet has a higher multiplicity. On the basis of earlier results, it would be anticipated that subjects would give lower ratings to out-of-scale notes following a (0,2,7) context set than after a (0,2,5) context set, as the (0,2,7) set is maximally compact, with high multiplicity and fewest out-of-scale notes; if this were indeed the case, it could be assumed that set multiplicity under transposition and inversion was not a significant factor. However, if subjects rated out-of-scale notes as fitting worse after a (0,2,5) context than after a (0,2,7) context, either transpositional and inversional equivalence could be presumed to hold for this task and diatonic multiplicity would be shown to be the principal factor in rating diatonic conformance or some hitherto unconsidered factor must be operational. Accordingly, a further pair of experiments was carried out in which the context sets were differentiated on the basis of transpositional and inversional equivalence.

Method

The subjects for Experiment 3.3 were the same as those used in the previous two experiments. The stimulus material was constructed as in earlier experiments. The subjects for Experiment 3.4 were 9 undergraduate and postgraduate students at The City University, London who were undergoing or had undergone formal musical training at University level. The procedure for both experiments was similar to that of Experiment 3.2, except that context sets used were of the form (0,2,7) or (0,2,5): in note names, either (F, C, G)/(D, C, G) or (F, C, D)/(F, G, D). As before, the probe was taken as the note furthest away from the maximally-compact pc set on the circle of fifths (i.e., if the maximally-compact pc set appears as the notes F, C, G the probe note would be F#), and the sequences were presented at random transpositions. The ordering of notes in each set was also randomised as before. Each subject listened to 12 sequences of each type. In Experiment 3.3 subjects rated fittingness on a seven-point scale, while in Experiment 3.4 a five-point scale was used. The change was simply to introduce some variability into the response measure.

Results

The results of both experiments indicated that the out-of-scale probes were judged less well-fitting after the (0,2,5) - (F, C, D)/(F, G, D) - type of sequence than after the (0,2,7) - (F, C, G) - type. In Experiment 3.3 the means were 3.8 versus 4.4 respectively ($t(11)=3.6$, $p<0.01$). In Experiment 3.4 the corresponding means were 2.3 and 2.8 ($t(8)=3.2$, $p<0.05$). It appeared therefore that the (0,2,5) pc sets created a stronger sense of diatonic conformance than did the (0,2,7) sets.

Accordingly, it would seem that either set multiplicity under transpositional and inversional equivalence was acting as a strong determinant of subjects' sense of diatonic conformance, or that some other factor which had not been explicitly accounted for was operational.

Discussion

In the light of the results of Experiments 3.3 and 3.4, it is necessary to review the possible influence on diatonic conformance ratings of factors other than multiplicity and specificity. There appear to be at least three plausible factors which could be involved: (i) a set's ability to evoke a stable reference note or tonic against which to judge note fittingness: (ii) the effect of the contour of context sets on the "expectedness" of a probe note, and hence its fittingness: (iii) the formation of "non-diatonic" (first order) three note-sets between the last two notes of a context and a subsequent out-of-scale probe, or alternatively, the diatonic rarity of the interval between the last note of a context and an out-of-scale probe. The possibility that each of these factors played some role in the experimental results so far reported will be considered.

The likelihood that different set-types have different "tonal implications" appears to be strong. It is possible that subjects could have assumed the first note of a context to be a tonic, reassessing this "tonicity" in the light of each subsequent note as the context set unfolded and altering the attribution of tonic as appropriate. This tonicisation of one of the notes of the context set could then serve to anchor fittingness ratings of probe-notes, which then would be dependent on the degree to which a set-type evoked a possible tonic, out-of-scale notes being rated as fitting worse after sets which successfully or strongly evoked a tonic. If this were the case, one might intuit that (0,2,7)-type sets would provide a better cue to tonicity than (0,2,5) sets. That they "seem" to be more "tonal" than (0,2,5) sets, is backed up by the fact that they are twice as likely to incorporate the "strongly tonal" interval of a fourth or fifth as are the latter sets; indeed, Brown and Butler (1981) found that (0,2,7) sets were better cues to possible tonics than were sets which were more widely spread on the circle-of-fifths, such as (0,2,5).

However, the results of Experiments 3.3 and 3.4, where (0,2,5) pc sets created a stronger sense of diatonic conformance than did (0,2,7) sets, appear to run counter to the argument that subjects' ratings of note-fittingness was predicated on the identification of tonal centres within context sets. Both intuitively and on the basis of Brown and Butler's results there seems to be no reason why (0,2,5) sets should establish a more secure tonal - or rather, modal - framework for listeners than should (0,2,7) sets. In addition, no effect of potential modal or tonal status of notes within context sets on judgments of diatonic conformance was noted in either the results of Experiment 2.1 or of Experiment 3.1. It would appear that even if subjects identified a tonal centre within - or on the basis of - context sets, this had little or no consistent effect on their judgments of note-fittingness. As suggested earlier, it seems that judgments of tonal centres and judgments of diatonic conformance are quite different things, involving different criteria and perhaps different cognitive processes.

The second factor to be considered is the possible effect of the contour of experimental sequences on the "expectedness" of the probe note, and hence on its fittingness. Given that the pitch range used was fairly constrained (the largest pitch interval used being a perfect fifth), the most probable effect of contour would appear to depend on the interval between the last note of the context and the probe being in the direction opposite to that of the interval between the penultimate and the last notes of the context, assuming that the principle of "good continuation" is likely to underlie a listener's expectations as to subsequent pitches. This "contoural reversal" could well lead to the probe being less expected - and thus judged as less well fitting - by a listener than would have been the case if no reversal had occurred. It would appear necessary to examine whether or not different types of pc sets might differentially and consistently give rise to such reversals.

Contours	(0,2,5)	(0,2,7)
UUU	2	0
DDD	3	0
UUD	4	3
UDU	1	2
DUU	0	0
DDU	3	3
DUD	4	4
UDD	3	2
Available Figurations	20	14
Number of end-reversals	12	12

TABLE 8.1

Contours formable within contexts-plus-probes for the set-types (0,2,5) and (0,2,7)

The factors involved in determining the contours formed by the context sets and probe notes used in Experiments 3.3 and 3.4 derive from the set-types used, (0,2,5) and (0,2,7), and from the constraint that the largest pitch interval between any two consecutive notes was a perfect fifth. Under this constraint, the pc set (0,2,5) could occur in twenty different figurations, while (0,2,7) could occur in fourteen different figurations, as shown in **Table 8.1**. These figurations fell, respectively, into seven and five out of the eight contour classes which are available between four (non-repeating) notes. The number of possible "end-reversals" - where the interval between the last note of the context and the probe is in the direction opposite to that of the interval between the penultimate and the last notes of the context - for each set-type was identical, although the probability of an end-reversal occurring was higher for the (0,2,7) sets than for the (0,2,5) sets. This arises from the fact that the number of possible end-reversals for (0,2,7) sets - twelve - represents a higher proportion of the fourteen possible figurations available with this set than is the case for the set (0,2,5), where the analogous proportion is twelve out of twenty. If contoural reversal had been operational in Experiments 3.3 and 3.4, one would have expected probes to be rated as fitting worse after (0,2,7) sets than after (0,2,5) sets; as the converse was the case in both experiments, it would seem that contoural reversals did not play a major or consistent role in subjects' judgments of note-fittingness.

Another possibility is that subjects were sensitive to the diatonic" rarity" of the interval between the last note of a context and an out-of-scale probe, or were responding to the occurrence of "non-diatonic" (first order) three note-sets between the last two notes of a context and a subsequent out-of-scale probe, and based their judgments of fittingness on one of these criteria. It could be anticipated that if a diatonically rare interval is highly likely to arise between the last note of a particular type of context set and an out-of-scale probe by virtue of the set-type, the more likely it is that subjects will judge out-of-scale probes as badly fitting. Similarly, if the probability is high of a "non-diatonic" note combination occurring between the last two notes of particular types of context sets and out-of-scale probes, subjects may be more likely to identify this note-combination as non-diatonic and give out-of-scale notes appropriately low ratings than if this probability is low. Recall that some evidence for the hypothesis that subjects were able to detect non-diatonic note combinations per se was given by the "distance-to-the-contradicted-note" result in Experiment 2.3.

Examination shows that in Experiments 3.3 and 3.4 the interval formed between the last note of a - randomly-ordered - (0,2,7) set and an out-of-scale probe was either a minor second or a tritone, with the minor second twice as likely to occur as the tritone. The interval formed between the last note of a (0,2,5)/(7,2,5) set and an out-of-scale probe was either a minor second, a major third or a tritone, with the minor second twice as likely to occur as either of the other two intervals. The diatonically rare intervals of a tritone and the minor second were therefore marginally more likely to occur between the last note of a (0,2,7) set and an out-of-scale probe than following a

(0,2,5) set. Thus the tendency of subjects in Experiments 3.3 and 3.4 to rate out-of-scale probes as worse fitting after (0,2,5) than (0,2,7) sets cannot simply be attributed to rareness of final interval.

Analysis of the set-types that may be formed between the last two notes of context sequences and the out-of-scale probe shows that diatonic and non-diatonic note-combinations were equally likely to arise if the context set was (0,2,5), but that diatonic note-combinations were twice as likely to arise than were non-diatonic combinations if the context set was (0,2,7). That is, the probability of non-diatonic note-combinations arising - and thus of subjects rating out-of-scale notes as badly fitting - is slightly higher with (0,2,5) sets than with (0,2,7) sets, at least with the particular out-of-scale probe which was used in the last two experiments. Accordingly, the hypothesis that subjects' responses in Experiments 3.3 and 3.4 might be accounted for in terms of sensitivity to non-diatonic note-combinations cannot be entirely excluded.

Experiment 3.5

To sum up, it would appear that the results of Experiments 3.3 and 3.4 are unlikely to be attributable to aspects of pitch contour. These results do not seem to be due to the diatonic rarity of the terminal interval of each sequence, and it appears that potential tonal or modal implications of the context sets were not operational in subjects' judgments. The results also appear to exclude totally the possibility that subjects based their judgments on the available number of out-of-scale notes permitted by each type of context set; in these judgments of diatonic conformance it appears that subjects are not engaging in a "position-finding" process based on some cognitive representation embodying the precise relations between diatonic sets within the chromatic set. Two primary hypotheses remain to be considered. The first is that subjects were sensitive to the multiplicity of subsets of the diatonic set, and were performing some sort of pattern-matching activity in making their judgments of diatonic conformance; sets which had a high multiplicity were acting as better cues to representations of specific diatonic sets, or to representations of several closely-related diatonic sets, than were lower multiplicity subsets. The second possibility is that subjects' responses arise from their identification of non-diatonic note-combinations, and that the higher multiplicity sets were more likely to form such sets with out-of-scale probes than were lower multiplicity sets.

Accordingly, a further experiment was carried out, using a similar probe-note rating procedure to that employed in Experiments 3.1 to 3.4. This experiment used both in-scale and out-of-scale probes, with two different types of note in each category. Two set-types were used: the maximal multiplicity set (0,2,5) and the major/minor triadic set (0,3,7). The latter set-type has a lower multiplicity - six - than the (0,2,5) set, which has a multiplicity of eight; if set-multiplicity is the primary factor in subjects' ratings of fittingness, then the set (0,2,5) would be expected to produce lower ratings for out-of-scale probes than would the set (0,3,7).

On the other hand, if the pcs which are out-of-scale for the set (0,2,5) are used as the out-of-scale probes for both set types, the probability that (0,3,7)-type sets will form non-diatonic note combinations between the probes and the last two context notes is slightly higher at 58% than is the 50% probability that such combinations will occur with the (0,2,5) sets. If subjects are indeed responding on the basis of the occurrence of non-diatonic note combinations then the (0,3,7) sets may well lead to lower ratings for out-of-scale probes than might the (0,2,5) sets. However, if this situation were to result, it would be difficult to avoid the hypothesis that the triadic sets simply generated more powerful tonal or modal implications for subjects than did the (0,2,5) sets; even though the results to date do not seem to show any effect of the tonal implications of context sets, triads - often taken to be the most potent signifiers of tonality - have not been explicitly used in Experimental Series 2 or 3.

Thus in the following experiment, two opposing predictions emerge; <u>either</u> (0,2,5) sets will produce lower ratings for out-of-scale probes, in which case the multiplicity theory appears to hold, <u>or</u> lower ratings will arise following (0,3,7) sets, in which case subjects may be either accepting triads as having stronger tonal implications than the other set-type used or subjects may be responding to the higher rate of occurrence of non-diatonic note-combinations with triadic sets.

As stated, in this experiment, both in-scale and out-of-scale probes were used. This was intended to provide a clear index of sense of diatonic conformance by examining the difference between fittingness ratings for each set-type and between in- and out-of-scale probes. This procedure also controlled for the possibility that <u>any</u> probe note (whether in- or out-of-scale) would be rated as less well-fitting after one kind of sequence.

<u>Method</u>

Subjects: The subjects were 21 undergraduate students at the University of London. None was studying music, but previous musical training varied from none to Associated Board Grade V.

Stimulus material: There were two independent variables - whether the probe was in- or out-of-scale with the preceding three notes, and whether the three-note set had the pc content (0,2,5) or (0,3,7). The two possible out-of-scale pitch classes permitted by (0,2,5) sets - the pcs 1 and 6 - were used as the out-of-scale probes for both set-types, while the in-scale probes for both set-types were those falling immediately adjacent to, but outside of, the region formed by the (0,3,7) set on the circle-of-fifths. That is, if the (0,2,5) set-types are represented as either the notes F, C, D or F, G, D, the (0,3,7) set-types contain the notes F, C, A or F, D, A (major or minor triads). The out-of-scale probes were randomly selected from either F# or C# and the in-scale probes randomly selected from either Bb or E. Each sequence and associated probe was presented at a random transposition. There were 16 sequences in each of the four conditions thus created. As in previous experiments the order of the notes of each context set was randomised, as was the order of the sequences themselves.

The sequences were generated and played on a BBC Model B microcomputer with the sound output diverted from the internal speaker through a low-pass filter to a Leak Model 30 Amplifier and an external speaker. The filter was set at 2.5 kHz (48 dB/octave), thus attenuating the upper harmonics of the square wave output of the BBC and producing a more pleasing sound with a clear and unambiguous pitch. The notes played ranged upwards for two octaves from middle C (ca 261 Hz). The presentation rate, interval between sequence and probe and inter-trial interval were all as for previous experiments.

Procedure: All subjects underwent all conditions. The basic procedure was as in previous experiments with subjects listening to a series of three-note sequences, each followed by a probe note and indicating how well or badly the probe fitted in with the previous three. Subjects made their response on a scale from 1 to 7 as in previous experiments.

Results

For both types of sequence the out-of-scale probes were judged significantly less well-fitting than in-scale probes, $F(1,20)=32.8$, $p<0.001$. The difference between ratings for the out-of-scale probes was greater for the (0,2,5)-type sequences than for the triads, $F(1,20)=6.5$, $p<0.025$. This was due to the out-of-scale probes associated with the (0,2,5) context sets being given lower ratings than with the (0,3,7) context sets, $F(1,20)=6.3$, $p<0.025$. Although ratings of in-scale notes following (0,2,5) sequences was higher than when these followed (0,3,7) sets, there was no significant difference between the ratings of the in-scale probes associated with the two types of context set.

Experiment 3.6

Overall, the results of the previous five experiments indicate that pc sets of the type (0,2,5) created a stronger sense of diatonic conformance than did those of either the type (0,2,7) or the triad set, (0,3,7). These results support the idea that diatonic multiplicity or representativeness is highly influential in determining the sense of diatonic conformance, or diatonic belongingness, obtained from three-note sequences. As a further test of this notion an experiment differing from its predecessors in this series was carried out. Firstly, instead of obtaining comparisons between two three-note sets, comparisons between three three-note sets were used, and specific predictions made as to their ordering in how well they would create a sense of diatonic conformance. Secondly, the task was modified; instead of merely rating the degree of fittingness of probe notes subjects had to identify out-of-scale notes. The purpose of this was to confirm that the previous pattern of results obtained using the fittingness measure actually reflected judgments concerning diatonic conformance by means of a task analogous to that involved in a two-alternative forced choice paradigm.

Method

Subjects: The subjects were 10 undergraduate and postgraduate students at the University of London. None had received formal musical training beyond an elementary level, although some still played an instrument occasionally for pleasure.

Stimulus material: There were three conditions. In the first condition the pc set used in the sequences was the (0,2,5) set. In the second condition the major/minor triad set (0,3,7) was used, while in the third condition it was the (0,2,7) set. For all set-types the out-of-scale pc was taken as that furthest from the most-closely-adjacent set - in the case of the set transposition F, C, G it was thus F# (the only available out-of-scale note), while the in-scale note (for the same set transposition) was either E or Bb. There were four trials in each condition. As in previous experiments the order of the notes of the three-note sequence was randomised, as was the order of the sequences themselves and the transpositions at which they were presented.

Procedure: The basic experimental task involved playing pairs of four-note sequences to subjects. Every pair consisted of a three-note sequence followed by an in-scale note and the same sequence followed by an out-of-scale note. Whether the out-of-scale or in-scale note occurred in the first or second sequence was balanced across trials. After each pair, the subjects were asked to indicate which sequence they thought had contained the out-of scale note. Their responses were made on a seven-point scale as follows: 1=definitely the first sequence, 2=almost certainly the first sequence, 3=probably the first sequence, 4=don't know, 5=probably the second sequence, 6=almost certainly the second sequence, 7=definitely the second sequence. Before the experiment the concept of "in-scale" and "out-of-scale" was described and demonstrated to the subjects by means of a piano keyboard. All subjects were able to intuit that having played all the white notes within an octave span, the black notes would be "out-of-scale" in the sense in which the term was being used in the experiment.

The notes were played using a step-time sequencer linked via a MIDI interface to a Prophet 2000 digital sampling keyboard. The sound used was that of a grand piano sampled at 41.2 kHz with 12 bit resolution in a two octave range upwards from middle C (ca 261 Hz). The notes were played at a rate of two per second, with a 2 second pause between the sequences of each pair and a 6 second pause between each pair of the sequences for subjects to make their ratings.

In trials where the out-of-scale note was in the first sequence of the pair, the subjects' ratings were subtracted from 8. The means were then calculated of the adjusted ratings for each subject in each of the three conditions. This represented the accuracy of subjects' identification of the out-of-scale note weighted according to their confidence in their judgment.

It was hypothesised that if diatonic multiplicity, or representativeness of diatonic scale structure, was important in determining sense of diatonic conformance, the (0,2,5) type contexts should give the highest ratings, followed by the (0,3,7) triadic contexts and then the (0,2,7) contexts.

Results

As predicted there was a monotonic trend in the subjects' scores going from the (0,2,5) sets to the (0,3,7) triad sets then the (0,2,7) sets. The respective means were 5.4, 4.8 and 4.2 out of 7. The difference between the conditions was significant, $F_{(2,18)}=6.1$, $p<0.05$ as was the linear trend, $F_{(1,18)}=11.0$, $p<0.01$. Thus the idea that that the multiplicity of pc sets within the diatonic set is highly influential in determining the sense of diatonic conformance or diatonic belongingness of a sequence of notes appears to be supported; diatonic multiplicity is shown to be a major factor underlying subjects' judgments of diatonic conformance in Experimental Series 2 and 3.

CHAPTER NINE

Conclusions

To sum up, the results of Experimental Series 1, 2 and 3 appear to show that diatonic interval structure has a privileged representation in cognition for Western listeners, and that this representation is best cued by sets of note which have a high commonality of formation within diatonic scale structure. More generally, these experiments also show that the responses of listeners who have had no formal training exhibit a consistent sensitivity to diatonic structure, from which it can be inferred that the cognitive representation of diatonic interval structure may arise through a process of acculturation. Taking the current results together with those of Brown and Butler's (1981) study, it appears that listeners may make use of complex aspects of that representation in widely divergent musical situations involving differentiation and hierarchization of notes as well as association and integration of notes within groups. While these findings are consonant with many of the studies cited previously which have shown some primacy for diatonic relations (e.g. Dowling, 1978: Krumhansl and Shepard, 1979), at the same time they imply that the cognition of musical pitch in the melodic domain is both more labile and more influenced by considerations of abstract alphabetic structure than other - particularly cognitive-structuralist - studies would imply.

This concluding Chapter will further explore implications of diatonic set structure in an attempt to assess the general significance of the current findings for pitch cognition and for music theory: it will be suggested that many of these implications should be incorporated into theories or models that seek to represent perceived musical structure. Initially, possible bases for the findings of Experimental Series 2 and 3 will be further explored. The results of a study by Brown (1988) will then be related to the current findings and it will be proposed that these imply a group-theoretic basis for diatonic structure. Possible relations between diatonic and other scale forms will be outlined, the music-theoretic status of diatonicism will be discussed, and a possible basis for treating different diatonic notes as being differentially stable will be advanced. Finally, some observations will be made on the significance of diatonic relations for aspects of pitch structure in post-tonal music.

Diatonic subset representativeness and interval vector similarity

Overall, judgments of diatonic conformance in Experimental Series 2 and 3 seem to be dependent on the diatonic multiplicity of the context sets. The higher the multiplicity of a set, the better it seems able to cue a sense of diatonic conformance. That is, the higher the diatonic multiplicity of a set, the more that set appears to be _representative_ of diatonic interval structure in cognition. However, multiplicity in itself might not be the sole basis of a set's diatonic representativeness; as was suggested earlier in the discussion of possible factors involved in pattern-matching, not only the commonality of diatonic subsets but also the commonality of intervals within the diatonic set may play some role. It would seem desirable to take such factors into account in addition to simple subset multiplicity when assessing the degree to which a set may be representative of diatonic structure. For instance, it might be possible for a subset to be formed with a fairly high multiplicity within the diatonic set yet for that subset to allow the formation of few specific intervals that are common within the diatonic set. It is therefore necessary to investigate aspects of the intervals formable within diatonic subsets - i.e., aspects of their interval vectors - in order to obtain a more complete account of the possible bases for a subset's diatonic representativeness.

It may appear appropriate to incorporate some measure of the degree to which the interval vector of a diatonic subset approximates to that of the diatonic set in any assessment of diatonic representativeness. Indeed, viewed in this light, the highest multiplicity three-note set (0,2,5) could be construed as being highly representative of the diatonic set by virtue of its interval vector alone. This takes the form <011010>, or one m3, one M2 and one P4; it thus contains one entry for each of the most commonly formable intervals within the diatonic set, of which the interval vector is <254361>. Unfortunately, it is difficult to differentiate between the diatonic representativeness of subsets on the basis of their interval vectors' approximation to the diatonic set's interval vector; for instance, on this basis the set (0,2,7) appears more-or-less as "representative" as the set (0,2,5), as its interval vector - <010020> - also contains three diatonically common intervals. Moreover, the hypothesis that simple approximation to the diatonic set's interval vector might be a principal determinant of a set's diatonic representativeness in cognition would appear to be undermined by the fact that the higher the cardinality (number of pcs) of a diatonic subset, the higher the likelihood of its interval vector approximating to that of the diatonic set. As was seen in Experiment 2.1, little consistent effect was found on subjects' detection of wrong notes of the number of diatonic notes prior to the wrong note, contrary to what may have been expected if interval vector similarity to the diatonic set had been the primary criterion for diatonic representativeness.

An alternative relation between the intervals formable within a diatonic subset and its diatonic representativeness can be found by examining the degree to which the interval vector of a diatonic subset approximates to that of all other diatonic subsets, i.e., is representative of all of the diatonic set's subsets. This would be equivalent to assessing the degree to which a set could act as a prototype of the category of pc sets which are subsets of the diatonic set. Tversky's (1977) method of assessing prototypicality would appear to offer an appropriate procedure for carrying out this assessment. He proposes a metric for the similarity of members of a category of objects which is based on the numbers of features shared and not-shared by different category members. He suggests that his similarity metric can be extended to give a measure of the prototypicality of category members in respect of the category, by summing the similarity of each category member to all other category members and ranking the category members accordingly such that the member which is most similar to all other members is the "most prototypical". Taking this as a model, the category of objects in question would be the totality of the pc subsets that can be formed within the diatonic set, while the features of each pc set would be the entries in its interval vector. Thus the similarity of each subset to each other subset could be ascertained by means of comparison of their interval vectors: the prototypicality of any given subset in respect of the totality of subsets would depend on the degree to which its interval vector approximated to that of all other diatonic subsets, together with its diatonic multiplicity, as will be seen.

Representativeness and prototypicality

The procedure for deriving Tversky's metric of similarity involves taking a measure of the features shared by two objects within a category, and subtracting from that some measure of the features that differentiate the objects from one another. In formal terms, the similarity between objects **a** and **b**, **s(a,b)** is defined in terms of sets of features here denoted by **A** and **B** respectively:

$$s(a,b)=zf(A\text{\textasciicircum}B)-xf(A-B)-yf(B-A)$$

where **A^B** are the features common to both **a** and **b**, **A-B** and **B-A** are the distinctive features of **a** and **b** respectively, **f** is the salience of the features and **z**, x and y are weighting factors. In other words, the similarity between a and b is equal to some weighting of their common factors - **A^B** - less some weighting of the features that **a** has and **b** does not - **A-B** - less some weighting of the features that **b** has and **a** does not - **B-A**. The prototypicality of an object **a** with respect to a category **L**, **P(a,L)** would be determined similarly:

$$P(a,L)=pn(sum(zf(A\text{\textasciicircum}B))-sum(xf(A-B)+yf(B-A)))$$

where the summations are over all the **not(a)** members of **L**. So **P(a,L)** is a linear combination of the measure of all features of object **a** that are shared with the other members of **L**, less a measure of all the features of **a** which are not shared with elements of **L**. If **P(a,L)** is maximal (i.e. greater than **P(n,L)** where **n** is any member of **L** other than **a**) then **a** is a prototype of **P(a,L)**, that is, can be held to be highly representative of objects in category **L**.

Applied to the totality of the pc subsets that can be formed within the diatonic set, and taking the features of each pc set as the entries in its interval vector, the procedure for estimating the diatonic representativeness of a subset in terms of its interval vector <u>and</u> its multiplicity would be as follows:

1) Each entry in each interval vector of all diatonic subsets of order **n**, 1<n>7, is compared with each analogous entry in all other subsets' interval vectors

2) Each "same" entry in each position is totalled to give a "similarity vector", or more properly, a "shared feature list" for every set.

3) The number of shared features for each set is then totalled, and multiplied by a number which is the "multiplicity" of the set (i.e. is the number of times that it can be formed as a subset of the diatonic set).

4) The sets are then ranked in order of the resulting numbers, which can be taken (after Tversky, 1977) as measures of the "prototypicality" of each set's intervallic structure in respect of the intervallic structures of <u>all</u> of the subsets of the diatonic set.

The non-shared features are not taken into consideration in the prototypicality calculation, as they are directly complementary to the shared features due to the closed and finite nature of the total feature-set. That is, if set A has features (*f*) in common with set B, all **not**(*f*) features will not be shared by the sets. The features that are shared are complementarily related to, and of the same order as, the non-shared features, and are therefore not included in the calculation. This is equivalent to giving the non-shared features a zero weighting, a procedure suggested in any case by Tversky's (1977) statement that common features are likely to weigh more heavily in prototypicality calculations than are non-shared features.

The reasoning behind the multiplication of the total number of shared vector entries by diatonic multiplicity to give prototypicality rankings is as follows. As there are **n** equivalent exemplars of each class of diatonic subset (type of pc set) within the diatonic set, the number of features that an exemplar of a given subset class shares with an exemplar of each other subset class can be summed within the class in question. Thus the "shared feature number" of subset class 3-11 (0,2,7) is the sum of the features that an exemplar of that set shares with an exemplar of all other subset classes; as there are five possible exemplars of 3-11 within a diatonic set, their shared feature numbers are summed to give the putative prototypicality of that class of subset with respect to the diatonic set.

Set	Rankings/zeros counted				Set	Rankings/zeros not counted			
Type	Total	Mult	Proto	Rank	Type	Total	Mult	Proto	Rank
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
3-7	89	8	712	1	3-7	31	8	248	3
2-5	90	6	540	2	2-5	9	6	54	32
3-11	85	6	510	3	3-11	37	6	222	5
4-22	84	6	504	4	4-22	46	6	276	1
2-2	93	5	465	5	2-2	11	5	55	31
3-9	83	5	415	6	3-9	21	5	105	20
3-4	94	4	376	7	3-4	43	4	172	8
2-3	92	4	368	8	2-3	11	4	44	36
3-2	92	4	368	9	3-2	39	4	156	10
4-14	87	4	348	10	4-14	66	4	264	2
4-11	83	4	332	11	4-11	62	4	248	4
5-23	76	4	304	12	5-23	55	4	220	6
4-23	73	4	292	13	4-23	23	4	92	23
2-4	97	3	291	14	2-4	17	3	51	34
3-6	93	3	279	15	3-6	25	3	75	28
5-27	69	4	276	16	5-27	48	4	192	7
4-26	87	3	261	17	4-26	49	3	147	12
5-35	75	3	225	18	5-35	38	3	114	18
2-1	93	2	186	19	2-1	17	2	34	38
3-8	91	2	182	20	3-8	46	2	92	22
4-16	85	2	170	21	4-16	73	2	146	14
3-5	84	2	168	22	3-5	44	2	88	26
4-27	83	2	166	23	4-27	66	2	132	17
4-Z29	83	2	166	24	4-Z29	83	2	166	9
4-10	78	2	156	25	4-10	45	2	90	25
4-13	78	2	156	26	4-13	66	2	132	16
4-20	77	2	154	27	4-20	46	2	92	24
5-29	75	2	150	28	5-29	75	2	150	11
5-25	73	2	146	29	5-25	73	2	146	13
5-24	70	2	140	30	5-24	70	2	140	15
5-20	56	2	112	31	5-20	56	2	112	19
6-33	50	2	100	32	6-33	49	2	98	21
6-32	50	2	100	33	6-32	29	2	58	30
2-6	87	1	87	34	2-6	18	1	18	41
6-Z25	42	2	84	35	6-Z25	42	2	84	27
3-10	77	1	77	36	3-10	29	1	29	40
4-8	71	1	71	37	4-8	49	1	49	35
4-21	70	1	70	38	4-21	32	1	32	39
5-34	70	1	70	39	5-34	53	1	53	33
5-Z12	68	1	68	40	5-Z12	68	1	68	29
6-Z26	42	1	42	41	6-Z26	42	1	42	37

TABLE 9.1

Columns (a) and (f) indicate the Fortean pc set, columns (b) and (g) give the total number of shared interval vector entries for the appropriate set, columns (c) and (h) give the set's diatonic multiplicity, columns (d) and (i) give the prototypicality figure produced, and columns (e) and (j) give the rank orderings of the sets under the two different methods used for prototypicality calculation (see text).

It should be noted that, in the ranking of subsets that arises from adopting this procedure, a zero entry is a real entry and is counted in the total; e.g., if a set A has interval vector A=<101220>, and it is compared with interval vector B=<203111> of set B, the resulting "provisional similarity vector" AB' would be <010000>. This would then be added to the "provisional similarity vectors" resulting from a comparison between interval vectors A and C through to A and N to give the "shared feature list" for the set A construed as a member of the "set of sets" A to N. The rationale behind this is to accurately reflect <u>all</u> of the properties of <u>all</u> of the subsets of the diatonic set (i.e. the fact the a certain interval <u>cannot</u> be formed within either of two sets is construed as being as much of a common feature as is the fact that there exists an interval which <u>is</u> formable within either set).

Columns (a) to (e) of **Table 9.1** show the ordering of diatonic subsets in terms of diatonic prototypicality that results from the application of the above procedure. Note that the most likely prototype of the diatonic set under this procedure is the set 3-7, consisting of the pcs (0,2,5), and that the order of likely prototypicality of the other pc sets that figured in Experimental Series 3 - i.e., the sets 3-11 (0,3,7) and 3-9 (0,2,7) - largely correlates with the degree to which they serve therein as appropriate cues for diatonic interval structure. An alternative procedure could be adopted, in which the fact of two subsets sharing a zero entry is <u>not</u> counted towards a subset's prototypicality rating. However, as can be seen from the rankings in columns (f) to (j) of **Table 9.1**, an ordering which does not count zero entries as significant features still gives high rankings to many of the sets ranked highly in columns (a) to (e). In particular, the three-note set 3-7 (0,2,5) is still the highest-ranked three-note set. Note also that the trichords containing tritones - 3-8, 3-5 and 3-10 - are ranked relatively low as prototypicality candidates in <u>both</u> rankings.

The procedure used to produce **Table 9.1** is predicated on the idea that diatonic interval structure can be thought of as a <u>category</u> - expressed as the <u>diatonic</u> <u>set</u> - which embraces a wide range of pitch relations and interval combinations, expressed as pc sets. The fit of the rankings in **Table 9.1** with the results of Experimental Series 2 and 3 could be taken to imply (after Rosch, 1973, 1975) that diatonic interval structure has indeed some representation <u>in</u> <u>cognition</u> as a category subsuming a range of interval combinations, these interval combinations having a varying diatonic significance for listeners that is dependent on their centrality with respect to the "diatonic category". Notwithstanding this possibility, the high prototypicality rankings of the pc set 3-7 in **Table 9.1** does at least provide a broader context than subset multiplicity alone for the idea that the set 3-7 is highly <u>representative</u> of diatonic interval structure in cognition, as well as for the idea that listeners are differentially sensitive to the different levels of diatonic representativeness. Further evidence for this differential sensitivity is presented by Brown (1988), in the context of a position-finding task (i.e. tonic identification).

Diatonic rarity and the attribution of tonal centres

In a study following-up the results obtained by Brown & Butler (1981), Brown (1988) hypothesised that listeners "tracked" tonal music by constantly assessing the positions of notes as they were heard through a position-finding process made possible by the unique asymmetrical place of each note within the diatonic set (i.e., because of unique interval multiplicity). However, she noted that the order of presentation of notes within scale-specific (i.e. tritone-plus-one) three-note sets had been found to affect the perceived tonal specificity in her earlier study with Butler. Moreover, in a further study requiring identification of tonal centres, Butler (1982) had found - using univalent tetrachords such as (0,1,5,6) - that perceived specificity was affected by whether or not diatonically rare intervals were salient in the presentation of context sets. For example, Butler had found that the ordered set of pcs [0,1,5,6] was a good cue to tonal centre, with rare intervals (m2, tt) being prominent, while the order [0,5,1,6] was not so good, as rare intervals were not prominent. Accordingly, Brown carried out an experiment designed to test whether both set content and the temporal order of notes were important in establishing a tonal framework.

Her study used nine categories of "tonal content" and four categories of "tonal context", and required musically-trained listeners (who were selected on the basis of pre-testing) to sing appropriate tonic notes following brief musical sequences. The variable "tonal content" was more-or-less equatable with the scale-specificity of the nine types of pc set used in the sequences (i.e. was dependent on the degree to which the different types of pc set specified a single scale or were diatonically common or rare), while "tonal context" was largely dependent on whether or not diatonically rare intervals were salient in the presentation of the sequences. She found significant main effects of tonal context, of content, and the interaction of these (all $p<0.001$). She also found that strings of pitches with identical intervallic content which gave a clear tonic indication in particular temporal contexts could be re-ordered so as to give no clear tonic indication, i.e. that temporal contextual ordering could override intervallic content cues as to tonic.

Thus, while Brown confirmed that the diatonic rarity of intervals and pc sets does affect listeners' position-finding abilities, she also showed that the order of occurrence of notes within intervals and sets provides potent cues to tonal centres. She concluded that "subjects...were sensitive to time-dependent functional relationships between pitches and these relationships could be most easily perceived in the presence of rare intervals in optimal temporal orderings", and that "while tonality certainly includes relationships of several tones to a single tone it is critically dependent upon listeners' contextual interpretations of relations among all tones as expressed in musical time".

The fact that Brown found a strong effect of temporal order on her musically-trained subjects' attributions of tonal centre seems to imply that her subjects had formed strong hypotheses as to specific scale or key in the course of listening to the sequences. This specificity of scale-hypothesis need not have been the case in Experimental Series 2 or 3; that is, subjects in these experiments need not have made use of representations of specific scales or keys in making their judgments. Indeed, it may be that listeners in Brown's experiment and in Experimental Series 2 and 3 were accessing different levels or different aspects of their representations of tonal relations. Or it may be that the cognitive representations of tonal relations to which Brown's listeners had access were simply different from those available to subjects in Experimental Series 2 or 3. After all, Brown's subjects were highly musically trained and might be expected to have assimilated the categories and concepts of western music theory as a result of their training, while the subjects in Experimental Series 2 and 3 had, in the main, no formal musical training and were likely to have developed representations of pitch relations largely through passive exposure to or unstructured participation in tonal music. On the other hand, it may be that the types of information relevant to position-finding and to pattern-matching tasks are different in both scope and in kind, as was suggested earlier.

Group-theoretic rationale for diatonic structure

Notwithstanding these differences, the results of Experimental Series 1, 2 and 3, taken together with those of Brown and Butler (1981) and Brown (1988) all confirm that western listeners exhibit a sensitivity to diatonic set structure and to manipulations of pc set relations within the diatonic set. The nature of this sensitivity as revealed in the experiments would appear to add credibility to the idea that group-theoretic, rather than psychoacoustical, considerations underlie western musical usage of diatonic scale structure. It can be postulated that the special properties of the diatonic scale within the equitempered chromatic system (i.e. of the set 7-35 within the cyclic group of order twelve or the dihedral group of order 24) enable position-finding activities - involving location or identification of local and global reference elements within patterns of pitches - to be carried out efficiently and economically. Equally, these properties enable pattern-matching activities - involving judgments of similarity and belongingness - to be carried out in a domain that offers a rich continuum of matches, near-matches and mismatches. In sum, these properties allow information to be construed as unambiguous in some contexts and highly ambiguous in others. In this, the structural properties of the diatonic set appear well suited for use in the aesthetic domain if theories such as that of Hankiss (1981) - which seeks to relate the nature of aesthetic experience to the fact that most media or symbol systems within which works of art are created allow for a controllable multiplicity of function of their elements - are in any sense correct. As Richmond Browne (1981, p. 20) suggests "tonality may be a maximally well-made construct, if our purposes are to make a game of position-finding and pattern-matching which will *never* come to a reflexive closure."

Nevertheless, the provision of a rationale for diatonic structure does not appear to address several significant aspects of western musical pitch organisation. After all, the diatonic *major* scale is only one of the normative scale-forms of western music theory: what of the minor scale forms ? What phenomenal status should be accorded to the diatonic scale ? How do scale and tonality - or tuning system and mode - interrelate ? Moreover, little has been said of the potential relevance of the current experimental results, or of Brown's results, or even of the group-theoretic representation of pitch itself, for harmonic pitch organisation, which is unquestionably one of the most significant and highly-developed features of western tonal music.

Diatonicism, major and minor

Major diatonic interval structure is of ancient music-theoretic lineage; it is well-documented by the second century A.D., and has persisted as a significant scheme of pitch organisation throughout the ensuing history of western music. The idea of different major and minor scale forms can be traced back to Zarlino in the sixteenth century, who proposed that one should differentiate between modes according to whether in moving upward from their final by two scale-steps they rose by the interval of a major or a minor third (Palisca, 1980). In current conventional music theory, three scale forms exist: the major, and the melodic and harmonic minors. The melodic minor scale differs in its ascending and descending forms; its descending form (T,T,S,T,T,S,T) can be derived rotationally from that of the major scale (S,T,T,T,S,T,T) but its ascending form displays an interval structure (T,S,T,T,T,T,S) that differs from the major. The harmonic minor has the interval structure (T,S,T,T,T,S,A_T,S), and also cannot be derived from diatonic structure.

It thus appears that any theory that concerns the structure of the major scale - such as the theory that the diatonic set has special properties when construed as a subset of the chromatic set - will incidentally encompass the structure of the descending melodic minor. However, the melodic-ascending and harmonic minor forms appear to remain outside the scope of such a theory. Moreover, within set theory, melodic-ascending and harmonic minor scale structures do not share the special properties of uniqueness, simplicity and coherence which inhere in diatonic major scale structure, and cannot be equated with major scale structure in group-theoretic terms. This would seem to limit severely the generality of the current experimental findings, and indeed, of the group-theoretic characterisation of musical pitch. After all, the minor scales appear to play an extremely significant role in tonal music; any theory of pitch cognition which fails to take account of this would be at best incomplete, at worst plain wrong.

Nevertheless, a coherent relation between major diatonic structure and the "incompatible" melodic and harmonic minor forms can be postulated in group-theoretic terms. If the conventional western scale forms - major, melodic minor and harmonic minor - are construed as pc sets enabling the formation of different sets of intervallic relations, then certain pc sets will be available as subsets within diatonic major structure while certain other subsets will be available within minor scale structures. In general, most of the sets that are to be found within one or other of the minor scale structures will be available within diatonic major structure. That is, major diatonic intervallic structure is an efficient and economical way of encompassing most of the pc sets available within the all of the scale forms used in tonal music. However, certain sets, in particular the sets 3-1 (comprising pcs 0,1,2), 3-3 (comprising 0,1,4) and 3-12 (comprising 0,4,8), are unavailable within the diatonic major set. These sets, it may be recalled, were taken as antithetical to diatonic structure in Experimental Series 1; as all individual intervals are formable within the diatonic set, these particular three-note sets were found to be the minimal possible means by which diatonic structure could be abrogated. They constitute the most economical structural "markers" for adiatonicism.

These sets are, however, available within melodic and harmonic minor structure. In the harmonic minor scale on A (comprising the notes A, B, C, D, E, F, G#, A) the sets 3-3 (0,1,4) and 3-12 (0,4,8) can be formed between the notes E, F and G#, and C, E and G# respectively. In the melodic minor scale on A (A, B, C, D, E, F#, G#, A, G, F, E, D, C, B, A), treating the ascending and descending forms as constituting one pc set together, then the set 3-1 (0,1,2) can be formed between the notes A, G and G#. So if the minor scale structures are taken together with major diatonic interval structure, all possible three-note pc sets may be formed. That is, the minor scale forms can be thought of as "extensions" of major diatonic scale structure that enable otherwise unavailable sets of intervals to be accounted for in terms of scale systems.

Scales as theories

This relation between major and minor scale forms may be made more comprehensible if it is stressed that scale forms are in essence abstractions from "real" music rather than themselves constituting that music. Neither the diatonic major scale nor the minor scales themselves constitute the stuff of pitch relations in music; they are theories - both descriptive and prescriptive - about pitch relations in music. To say that the diatonic scale is a descriptive theory about pitch usage in western tonal and pre-tonal (modal) music is to make the claim that most pieces, or separable and coherent segments (phrases) of pieces, or groups of adjacent hierarchically-important (structural) notes within phrases or pieces of tonal music make use of groups or sets of pitches and intervals that can be construed coherently as deriving from diatonic scales. That is, the diatonic scale is a good way of representing typical pitch relations in many western musical pieces and genres within a single coherent framework. The diatonic set is a formalisation of that framework, at least insofar as that framework exists within an equal tempered system. The existence of particular pitch configurations in tonal music which cannot be subsumed into diatonic structure can thus be conceived of as necessitating the postulation of minor scale forms, which may be regarded as having intervallic structures that are less generally applicable to a wide range of pitch relations in tonal music than is that of the diatonic major form. The diatonic scale could be said to function as a prescriptive theory within western formal musical training, acting - in the "quasi-modal" form in which it appears within conventional music theory - as an overlearned framework that constrains melodic motion and (to some extent) melodic tonal function[1].

---

[1] Reese (1941, p 162-163), provides an insightful account of how observation and existing theory might interact in the systematisation and codification of musical practice, and of the role that descriptive and prescriptive theories of musical pitch organisation may play in this process.

The experimental results reported here would appear to confirm that the diatonic scale is a good descriptive theory of pitch usage in tonal music, implying as they do that accultured western musical listeners - who have had little if any formal musical training - build up some cognitive representation of pitch relations that conforms to diatonic constraints. Nevertheless, the fact that these experiments indicate that listeners are sensitive to aspects of diatonic interval structure does not imply that they need have conscious access to some discrete cognitive representations of diatonic scale structure. It would appear to indicate that listeners have internalised the possibilities and typicalities of interval structure within - broadly - tonal music, and as a result have access to some representation that embodies many of the properties of the diatonic set as a subset of the chromatic set.

As was suggested earlier, this representation may take the form of a category subsuming a range of interval combinations, these interval combinations having varying significance for listeners depending on their centrality with respect to the diatonic category. Listeners may have internalised relations within sets of notes or intervals or "musical fragments" (a hypothesis that <u>may</u> be supported by the apparent "a-centric" and fragmentary nature of children's early songs, see Dowling, 1982a: Hargreaves, 1986), resulting in a representation approximating to - or having similar properties to - a coherent model of diatonic interval structure gradually taking shape in their cognition. A sense of the significance or potential function of sets of notes or intervals in respect of diatonic interval structure would arise <u>during the process of formation</u> of the representation of diatonic interval structure; this sense of diatonic function might in fact constitute the cognitive representation of diatonic interval structure. That is, diatonicism in cognition may be construed as an emergent property of the constellation of sets of notes or intervals of which a listener has developed some cognitive representation in the course of exposure to tonal music.

Alternatively, one could postulate that the idea that listeners are sensitive to diatonic properties of pc sets does not necessarily imply that listeners have access to an explicit - group-theoretic - representation of the diatonic set, nor need it imply that the "diatonic set" is represented in terms of its subsets. Such a sensitivity might derive from some distributed representation of tonal relations within which the diatonic set and its subsets do not have any explicit or separable representations, but which has the formal property that it models aspects of diatonic set/subset relations (cf. Bharucha and Olney, 1989). This distributed representation could arise through more-or-less passive exposure to music that exhibits diatonic and/or tonal characteristics.

Tonality and the diatonic set

Whatever the actual nature of the representation, a major issue remains to be discussed - the relation between diatonic interval structure and tonality, or between the "tuning system" and the "modal" levels of pitch representation (after Dowling, 1978). That there is some such relation is unquestionable; after all, both Brown and Butler's (1981) study and Brown's (1988) study imply that there is some cognitive differentiation in terms of referential value between notes of the diatonic set, and that structural aspects of the diatonic set can act as significant cues in identifying referential notes. Admittedly, these studies only demonstrate this effect of diatonic interval structure for musically-trained listeners, but it does seem intuitively likely that such an effect may also hold for accultured listeners. In view of the fact that these position-finding activities are shown to rely on special structural properties of the diatonic set, it would seem desirable to check whether or not such structural properties might imply some referential differentiation between diatonic notes. Unfortunately, it is difficult to postulate a rationale for such differentiation in group-theoretic terms. In fact Balzano goes so far as to state (1982, p 325) that "there is nothing in the theory of sets that provides for individuating a particular element of a set as a reference element".

However, if the idea of referential value is equated with the idea of the _prototypicality_ of a note with respect to the diatonic set (as is in fact suggested by Krumhansl (1979), see Chapter Four), a possible structural basis for differential referentiality emerges. This would result from the application of the same procedure as used at the outset of this Chapter, where the similarity of each member of a category to all other category members is measured and those members that are most similar to all other members are taken to be most "prototypical" in respect of the category. It is possible to regard the diatonic set itself as constituting a category with _individual pitch classes_ as members (rather than, as previously, _pitch class sets_); the features on which the prototypicality metric would be based could then be taken as the intervallic context of each pitch class within the diatonic set (i.e., the intervals that each pitch class can form with all other pitch classes within the diatonic set).

As mentioned previously, Browne (1981) shows that the intervallic context of every pitch class within the diatonic set is different; however, the interval context of a given pc will be more similar to those of certain other pcs than to others. For example, within a diatonic set consisting of the pitch classes C, D, E, F, G, A, B, the interval context of C will differ in only one point from that of G (C's interval context includes a M7, whereas that of G does not, but does include a m7). The interval context of D will again differ in only one point from that of G, whereas it will differ from that of C in two points. On this basis, C can be said to be more similar to G than to D, while G is equally similar to C and to D.

A prototypicality ranking for pcs within the diatonic set can be produced by summing the number of intervals that the context of pc A shares with those of pc B through to pc N, and repeating the process for each diatonic pc. This procedure is directly analogous to that employed in assessing prototypicality rankings for diatonic subsets, and sums all shared features for each category member (here, each pc) while giving non-shared features zero weightings (for the same reasons as adduced earlier). If intervals (or features) are taken as being not equivalent under inversion, the prototypicality ordering for a diatonic set consisting of the pitch classes C, D, E, F, G, A, B shown in columns (b) and (c) of **Table 9.2** results. The note D, the supertonic, has the highest ranking and could be regarded as the set's prototypical pitch class. However, if intervals (or features) are taken as being equivalent under inversion so that a m7 is equated with a M2 (thus reducing the number of possible types of feature to six, the prototypicality ordering shown in columns (d) and (e) of **Table 9.2** results, and the notes C (tonic) and E (mediant) would be co-equally the set's most prototypical pitch classes.

Pitch class (a)	Common features (b)	Ordinal prototypicality (c)	Common features (d)	Ordinal prototypicality (e)
C	20	4	20	1
D	24	1	6	7
E	20	4	20	1
F	16	6	18	3
G	23	2	16	5
A	23	2	16	5
B	16	6	18	3

TABLE 9.2

Column (a) indicates note-name or pitch class within the diatonic set, columns (b) and (d) give the prototypicality figure produced, and columns (c) and (e) give the rank orderings of the pitch class under the two different methods used for prototypicality calculation (see text).

It may appear that these rankings are not particularly informative. After all, the usual reference element for the particular diatonic set instanced above would be the note C, which figures only in the second of the two rankings given, and then only as co-equal with the note E. The note D would not generally be regarded as a particularly stable member of this diatonic collection. However, if it is recalled that diatonic structure played a significant role in pre-tonal (i.e. modal) western music and music-theory of the middle ages and early renaissance, then the emergence of the note D as the set's prototype in one of the possible rankings is of interest; this note served as the final of the first-classified church mode, and would be regarded as highly stable - at least, in theory - within monophonic or early polyphonic western music. Moreover, the notes C and E would be highly likely to figure as stable referential entities in tonal (post-modal) music, and are indeed claimed as prominent members of the "tonal hierarchy" by the cognitive-structuralists.

Thus some rationale predicated on the structure of the diatonic set can be found for referential differentiation between diatonic notes. The adoption of different bases for intervallic equivalence produces different prototypicality results, but these results can be directly related to different historical musical usages. It would appear that one of the reasons for the historical persistence of diatonic scale structure is that it can be adapted to provide a coherent framework for a wide range of musical usages. However, in view of the fact that different prototypicality results can be produced, it is unlikely that this structural rationale alone is sufficient to account for the difference in referential value accorded to different scale-notes in cognition. The mechanisms adduced by the cognitive structuralists for the "acquisition of the tonal hierarchy", based on sampling of the differing syntactical and temporal usages of different scale-notes within tonal music, may well play some role. Moreover, although certain invariant cues as to referential value may be provided by diatonic structure per se, it would appear that the differential stability of notes in cognition is not best expressed as a static hierarchy, but rather as variable according to interval content cues and temporal context, as Brown (1988) implies.

The sensitivity to diatonic relations that is indicated by the experimental evidence has been derived from and applied to the melodic domain only. This evidence can be related to certain aspects of harmonic theory; in particular, Brown and Butler's (1981) and Brown's (1988) evidence that rare (tritonal) relations function as good cues in position-finding seems to provide a feasible rationale - in terms of diatonic interval structure - for the structure of the dominant seventh-to-tonic cadence. However, it is difficult to see how diatonic relations can provide a coherent rationale for harmonic usage, except perhaps in the way that diatonic structure can be shown - through Balzano's (1980) direct product group formulation of the diatonic set - to encapsulate triadic relations in a compact way. It seems unlikely that diatonic relations can account for the structural functions of verticalities, at least when these are regarded as "fused" entities. If the experimental evidence presented here to the effect that tonal relations are labile and are determined by alphabetic and temporal structure is held to discount the static spatial model of tonal relations proposed by the cognitive-structuralists, it may be appropriate to seek a basis for certain aspects of the structural functions of chords and chord successions in the domain of psychoacoustics. It is possible that the theory of diatonic set structure may be integrated with a flexible model of harmonic relations (such as that of Parncutt, 1986) so as to provide a means of articulating the cognitive representation of pitch relations within the "polyphonic homophony" that characterises much of western tonal music. On the other hand, there is a current musicological view that the "rules" of tonality have always been in a constant state of change, and that no one set of principles can be developed which could account for all of the vicissitudes of pitch organisation throughout the so-called "common-practice" period of western tonal music[2]. It may be that to undertake such a quest would be to pursue a chimera.

Nevertheless, it does appear that whatever other factors underlie the cognitive representation of musical pitch for western listeners, the dynamics of diatonic structure can play a considerable role. While it is likely that even an accultured listener will have developed some representation of musical pitch in modal or tonal terms (i.e. in terms of different scale notes having different stabilities), this representation will be embedded in a representation of diatonic structure which may play a functional role in instantiating perceptions of belongingness and relatedness between notes and configurations of notes, and of reference within configurations of notes. Although the music-theoretic definitions of scale presented in Chapter Three treat it as a relatively simple and neutral entity with undifferentiated elements, it would appear that in perception and

---

[2] Indeed, the difficulty of defining the characteristics of western tonality in a "trans-historical" way is evident from the fact that different musicologists find its first, full-blown emergence in the works of different composers. For example, Harman and Milner (1962, p.365) refer to the works of Cesti (fl 1650-1670) as the first to consistently exemplify the principles of tonality, while Blume (1975) refers the emergence of functional tonality to the works of Corelli and Lully.

cognition, the diatonic scale is dynamic and complex; its structure in cognition should not be regarded as a sort of "ladder" or "template", but rather as a flexible network of implications, which is only ever partially disambiguated by its temporal context. Thus, rather than the cognition of tonality simply devolving onto some a priori representation of the differential stabilities of individual notes, the diatonic implications of configurations of notes over time irrespective of their relative stabilities within a conventional tonal hierarchy or (to some extent) their order of occurrence seem capable of constituting a significant component of a listener's awareness of "tonality". The dynamics of diatonic interval structure are subtle, but they appear to exert a potent influence on our musical perceptions and cognition; they must be taken into account in any musical theory or music-analytic method that claims to provide an account of music as it is perceived.

Diatonicism and atonality

As a codetta to the ideas presented in this Chapter, it seems appropriate to conclude by giving a brief outline of how the concept of diatonicism as developed in this thesis may be used to shed some light on aspects of post-tonal music.

The foregoing theoretical and experimental evidence appears to indicate that diatonicism may be a necessary component of western tonality but is not sufficient in itself to characterise tonality. This idea may have interesting implications for understanding the nature of pitch organisation in the post-tonal, pre-serial music of the first two decades of this century. The most salient characteristic of pitch organisation in this music is commonly taken to be its use of *highly* chromatic melodies and non-triadic verticalities in ways that do not seek to integrate these into a framework based in traditional functional tonal relationships, its "emancipation of the dissonance". As Rosen (1976, p 44) says, "The expressiveness of Schoenberg's melodies goes naked, and any attempt to resolve them by [tonal] harmony would only travesty them". That is, the pitch structure of post-tonal music is commonly construed as not being explicable in terms of the implications that the constituents of its melodic or harmonic vocabulary might have within the context of tonal harmony: the music is hence "atonal".

The most influential recent attempt to characterise the pitch organisation of atonal music derives from the work of Babbitt (1961) and later, Forte (1973). As described in Chapter Five, this approach makes use of the group-theoretic representation of the chromatic set, treating it as analogous to the dihedral group of order 24. That is, it partitions the chromatic set into pitch class sets, differentiated by cardinality and pc content. Pc sets that are equivalent under transposition and inversion are regarded as exemplars of the same form of pc set. Pc sets may be related to one another by inclusion (set-subset relation), by complementation, by similarity of pc content or of interval vector, or by membership of the same "set-complex" (a complex formal entity predicated on inclusion and complementation relations). Forte and others have made use of this formulation to analyse the pitch-structure of pieces of atonal music by showing their employment of pc sets in coherent and consistent ways.

Unfortunately, Forte's methodology is not generally employed in such a way as to make clear the relation of pitch structure in atonal music to that within tonal music. However, it may be informally proposed that certain aspects of relations between pitch structure in atonal and in tonal music can be clarified by making use of the group-theoretic formulation of diatonic structure, the diatonic set. If (as suggested earlier) it is accepted that diatonic structure provides a fairly good account of typical pitch relations in tonal music, it might be anticipated that music whose pitch structure cannot be interpreted in terms of tonality will make use of sets of notes that cannot be accounted for in terms of diatonic structure. That is, it is likely that the pitch structure of atonal pre-serial music will consistently abrogate _diatonic_ structure.

Recall that certain sets of notes cannot occur within the diatonic set. In fact, the sets 3-1 (comprising pcs 0,1,2), 3-3 (comprising 0,1,4) and 3-12 (comprising 0,4,8) were instanced earlier in this Chapter as the minimal possible means by which diatonic structure could be abrogated, and constitute the most economical structural "markers" for _adiatonicism_. If the pitch structure of many pieces of atonal pre-serial *works is examined*, in numerous instances these adiatonic pc sets will be seen to play significant roles. For example, the texture of Webern's Op. 11 No. 3 for cello and piano (see **Figure 9.1**) is permeated by such sets, both melodically and harmonically. Similarly, the set 3-3 plays a prominent role in Schoenberg's _Drei Klavierstucke_ Op. 11 at the outset of both No. 1 and No. 3 (see **Figure 9.2**).

Figure 9.1:Webern's Drei Kleine Stücke Op 11, No 3 for cello and piano (entire), showing texture permeated by adiatonic pc sets. The piece gradually employs the total chromatic set as it unfolds.

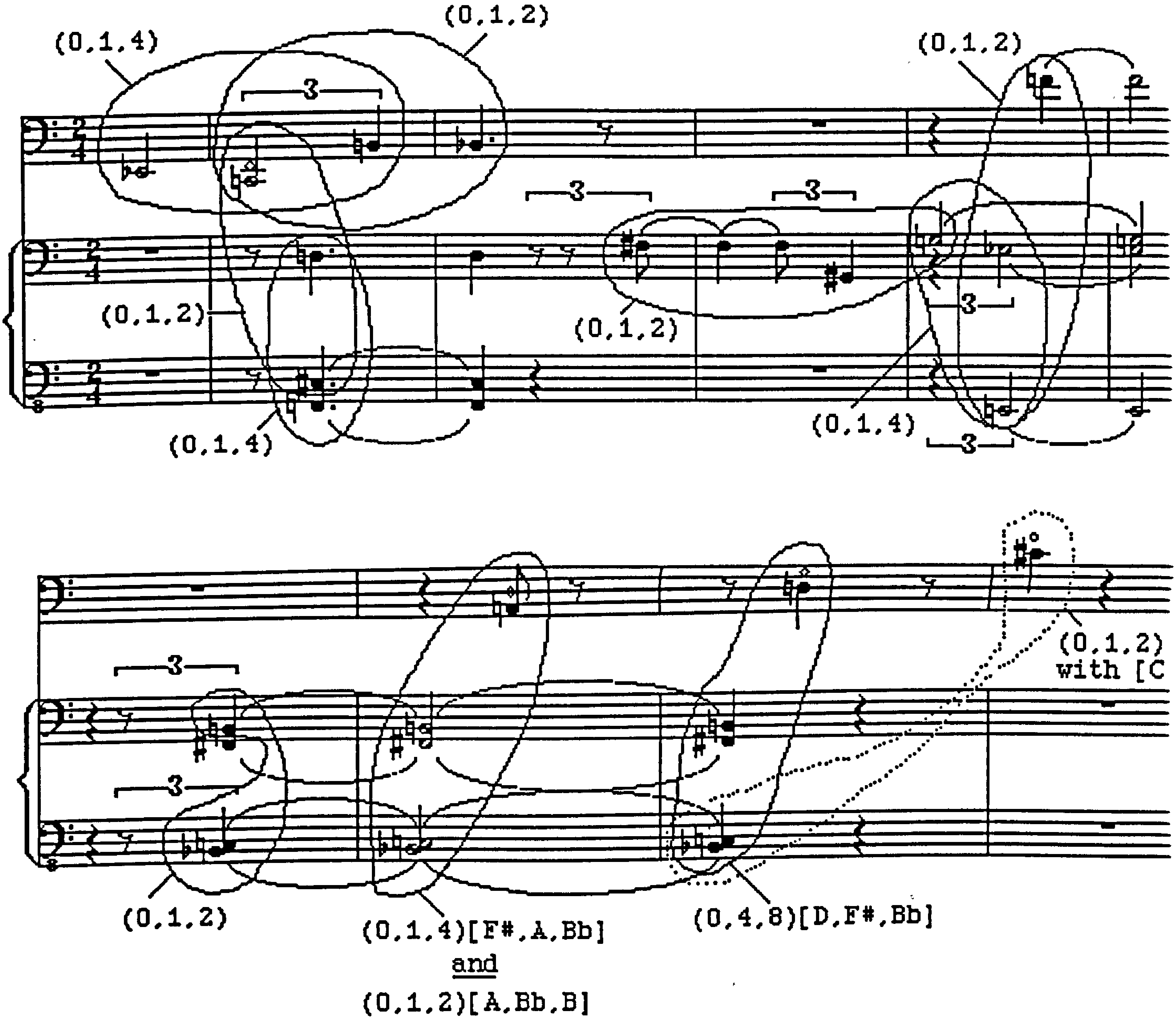


Figure 9.2a:Schoenberg, Drei Klavierstücke Op 11, No 1 first three bars, showing use of the adiatonic set (0,1,4).

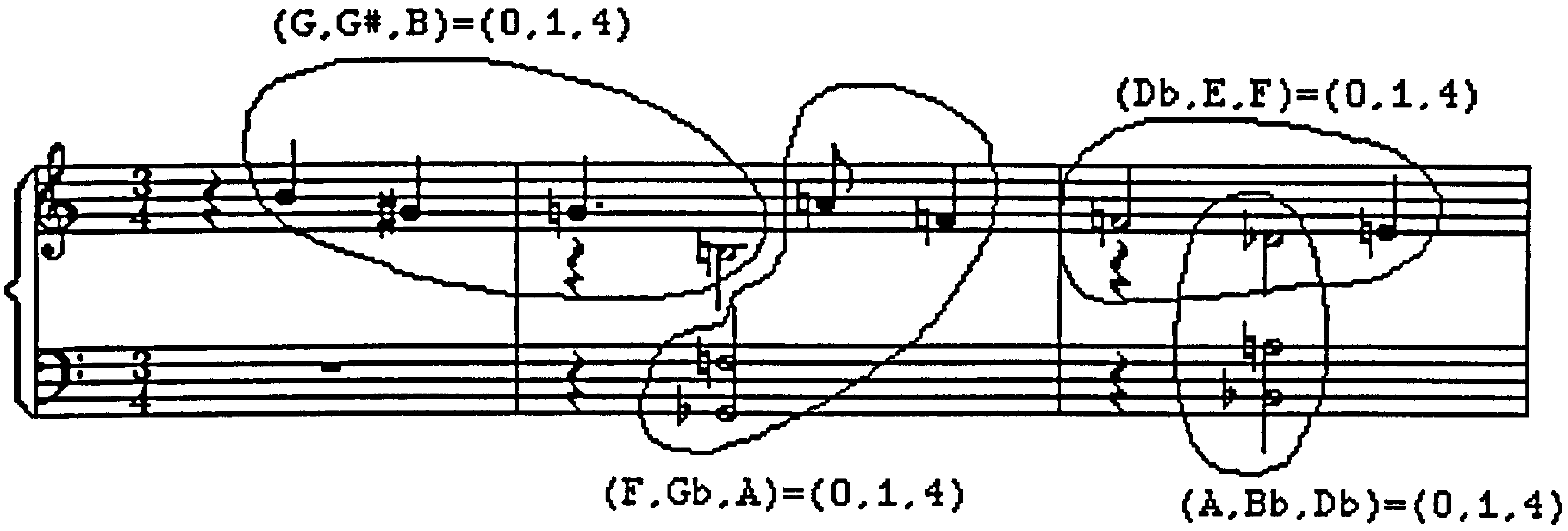
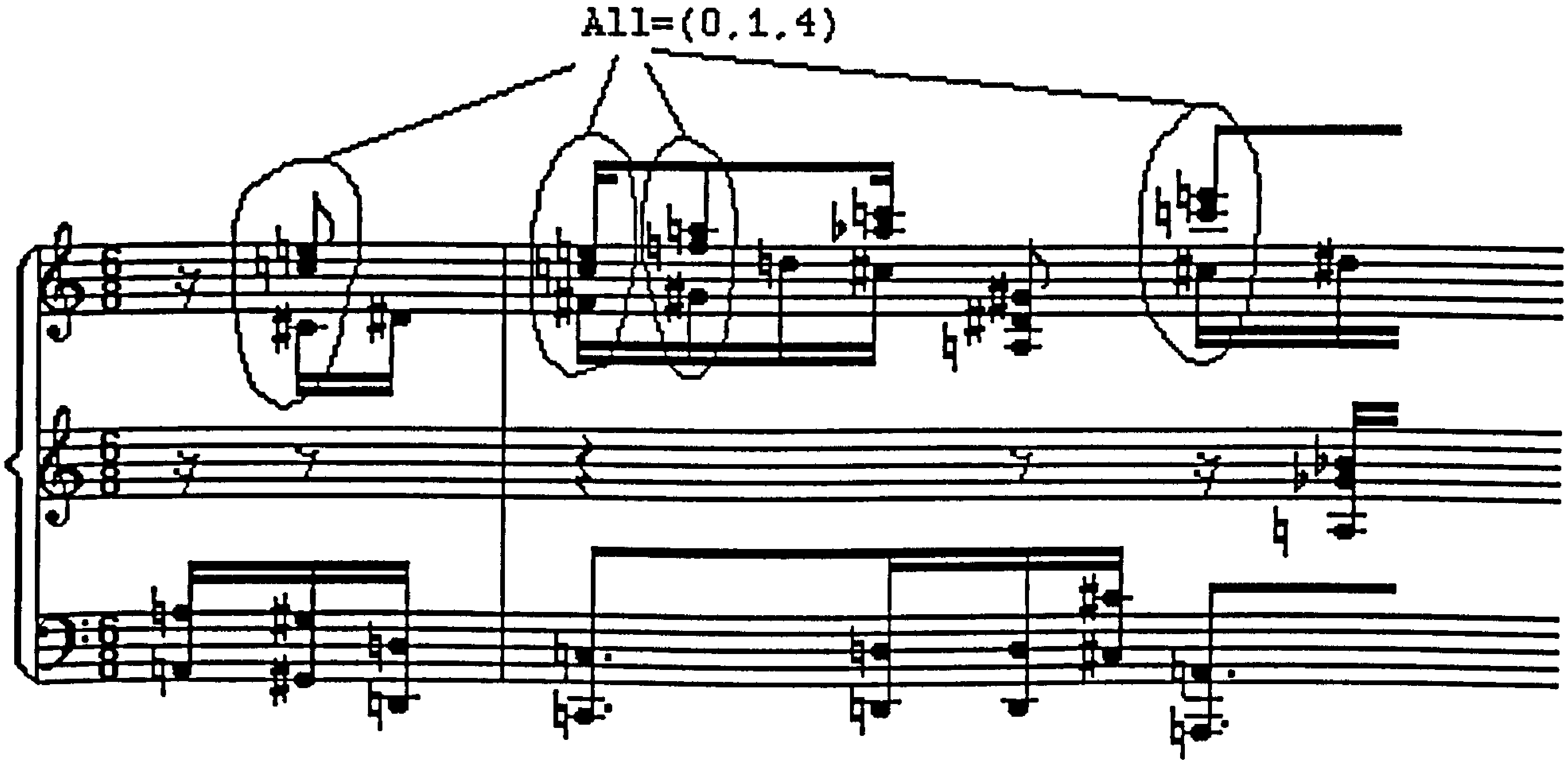


Figure 9.2b:Schoenberg, Drei Klavierstücke Op 11, No 3 opening two bars, showing use of the adiatonic set (0,1,4).



In these examples, and in many others of this period (see, e.g., Pierrot Lunaire), both Schoenberg and Webern consistently make use of pitch structures that cannot derive from the diatonic set. This makes sense if diatonicism is construed as a necessary component of western tonality; in order for composers to transcend the tonal implications of melodic lines or harmonic structures, they must select pitch material that abjures diatonic structure. At the same time, the restricted number of different types of adiatonic sets may assist in imposing some motivic unity. In these ways, adiatonicism may constitute a significant and previously unexplored strand of atonality, and would appear to be deserving of more detailed and exhaustive study.

REFERENCES

Associated Board of the Royal Schools of Music. *Rudiments and theory of music*. Associated Board of the Royal Schools of Music. London, 1958.

Attneave, F. & Olson, R. K. Pitch as a medium: a new approach to psychophysical scaling. *American Journal of Psychology*, 84, 147-166, 1971.

Babbitt, M. Set structure as a compositional determinant. *Journal of Music Theory*, 5, 72-94, 1961.

Bach, C. P. E. *Essay on the true art of playing keyboard instruments* (1787). Trans. W J Mitchell (2nd edition) Cassell & Co.: London, 1951.

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

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

Balzano, G. J. What are musical pitch and timbre ? *Music Perception*, 3, 297-314, 1986.

Bartlett, J. C. & Dowling, W. J. The recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 501-515, 1980.

Becker, J. *Traditional Music in modern Java*. University of Hawaii Press: Honolulu, 1982.

Bharucha, J. J. & Krumhansl, C. L. The representation of harmonic function in music: hierarchies of stability as a function of context. *Cognition*, 13, 63-102, 1983.

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

Bharucha, J. J. & Stoeckig, K. Reaction time and musical expectancy: priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 403-410, 1986.

Bharucha, J. J. & Stoeckig, K. Priming of chords: spreading activation or overlapping frequency spectra ? *Perception and Psychophysics*, 41, 519-524, 1987.

Bharucha, J. J. & Olney, K. L. Tonal cognition, artificial intelligence and neural nets. *Contemporary Music Review*, special issue on Music and the Cognitive Sciences, S McAdams (Ed.), Vol 3, 1989.

Blackwell, R. & Schlosberg, H. Octave generalisation, pitch discrimination loudness thresholds in the white rat. *Journal of Experimental Psychology*, 33,407-419, 1943.

Blume, F. *Renaissance and Baroque Music*. Faber: London, 1975.

Bregman, A. S. The formation of auditory streams. In Requin, J. (Ed.) *Attention and Performance VII*. Erlbaum: Hillsdale, New Jersey, 1978.

Bregman, A. S. Asking the "what for" question in auditory perception. In Kubovy, M. & Pomerantz, J. R. (Eds.), *Perceptual Organisation*. Erlbaum: Hillsdale, New Jersey, 1981.

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

Bregman, A. S. & Dannenbring, G. L. The effect of continuity on auditory stream segregation. *Perception and Psychophysics*, 13, 308-312, 1973.

Bregman, A. S. & Dannenbring, G. L. Auditory continuity and amplitude edges. *Canadian Journal of Psychology*, 31, 151-159, 1977.

Bregman, A. S. & Pinker, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32, 20-31, 1978.

Bregman, A. S. & Rudnicky, A. I. Auditory segregation: stream or streams ? *Journal of Experimental Psychology: Human Perception and Performance*, 1, 263-267, 1975.

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

Brown, H. & Butler. D. Diatonic trichords as minimal tonal cue-cells. *In Theory Only*, 5, 39-55, 1981.

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

Burns, E. M.	Octave adjustment by non-Western musicians. *Journal of the Acoustical Society of America*, *Supplement*, *Fall*, 56, S25-26(A), 1974.
Burns, E. M.	Circularity in relative pitch judgments of complex tones: the Shepard illusion revisited again. *Perception and Psychophysics*, 30, 467-472, 1982.
Burns, E. M. & Ward, W. D.	Categorical perception - phenomenon or epiphenomenon: evidence from experiments on the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 63, 456-468, 1978.
Burns, E. M. & Ward, W. D.	Intervals, scales and tuning. In Deutsch, D. (Ed.), *The Psychology of Music*. Academic Press: London, 1982.
Butler, D.	A further study of melodic channelling. *Perception and Psychophysics*, 25, 264-268, 1979.
Butler, D.	The initial identification of tonal centers in music. In Rogers, D., and Sloboda, J. (Eds.), *The acquisition of symbolic skills*. Plenum: New York, 1983.
Butler, D.	Review of Sundberg, J. (Ed.), "Harmony and Tonality". Royal Swedish Academy Publication No. 54: Stockholm, 1987. *Psychology of Music*, 16, 164-170, 1988.
Butler, D.	Describing the perception of tonality in music: a critique of tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, 6, 219-241, 1989.
Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L.	Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, 113, 394-412, 1984.
Cuddy, L. L.	On hearing pattern in melody. *Psychology of Music*, 10, 3-10, 1982.
Cuddy, L. L. & Badertscher, B.	Recovery of the tonal hierarchy: some comparisons across age and levels of experience. *Perception and Psychophysics*, 41, 609-620, 1987.
Cuddy, L. L., Cohen, A. J. & Mewhort, D. J. K.	Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 869-882, 1981.

Cuddy, L. L., Cohen, A. J. & Miller, J.	Melody recognition: the experimental application of musical rules. *Canadian Journal of Psychology*, 33, 148-157, 1979.
Dahlhaus, C.	Entry on "Harmony", *Grove's Dictionary of Music*, Macmillans: London, 1980.
Dannenbring, G. L. & Bregman, A. S.	Stream segregation and the illusion of overlap. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 544-555, 1976.
Davies, J. B.	Memory for melodies and tonal sequences: a theoretical note. *British Journal of Psychology*, 70, 205-210, 1979.
de Boer, E.	On the residue and auditory pitch perception. In W. D. Kiedel & W. D. Neff (Eds.), *Handbook of sensory physiology*, Vol V/3. Springer-Verlag: New York, 1976.
Deutsch, D.	Music recognition. *Psychological Review*, 76, 300-307, 1969.
Deutsch, D.	Tones and numbers: specificity of interference in short-term memory. *Science*, 168, 1604-1605, 1970.
Deutsch, D.	Octave-generalisation and tune recognition. *Perception and Psychophysics*, 11, 411-412, 1972.
Deutsch, D.	Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 57, 1156-1160, 1975a.
Deutsch, D.	The organisation of short-term memory for a single acoustic attribute. In Deutsch, D. and Deutsch, J. A. (Eds.), *Short-term Memory*. Academic Press: New York, 1975b.
Deutsch, D.	The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, 28, 381-389, 1981.
Deutsch, D.	The octave illusion and auditory perceptual integration. In Tobias, J. V. & Schubert, E. D. (Eds.), *Hearing research and theory* (Vol 1). Academic Press: New York, 1981.
Deutsch, D.	The processing of pitch combinations. In Deutsch, D. (Ed.), *The Psychology of Music*. Academic Press: London, 1982.

Deutsch, D. Two issues concerning tonal hierarchies: comment on Castellano, Bharucha and Krumhansl. Journal of Experimental Psychology: General, 113, 413-416, 1984.

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

Dewar, K. M., Cuddy, L. L. & Mewhort, D. J. K. Recognition memory for single tones with and without tonal context. Journal of Experimental Psychology: Human Learning and Memory, 3, 60-67, 1977.

Divenyi, P. L. & Hirsh, I. J. Identification of temporal order in three-tone sequences. Journal of the Acoustical Society of America, 56, 144-151, 1974.

Divenyi, P. L. & Hirsh, I. J. Some figural properties of auditory patterns. Journal of the Acoustical Society of America, 64, 1369-1386, 1978.

Dowling, W. J. Rhythmic groups and subjective chunks in memory for melodies. Perception & Psychophysics, 14, 37-40, 1973.

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

Dowling, W. J. Melodic information processing and its development. In Deutsch, D. (Ed.), The Psychology of Music. Academic Press: London, 1982a.

Dowling, W. J. Chroma and melody recognition: effects of acquiring a tonal schema. Journal of the Acoustical Society of America, 72, S11(A), 1982b.

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

Dowling, W. J. & Bartlett, J. C. The importance of interval information in long-term memory for melodies. Psychomusicology, 1, 30-49, 1981.

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

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

Edworthy, J. Melodic contour and musical structure. In Howell, P., Cross, I. & West, R. (Eds.), Musical structure and cognition, Academic Press: London, 1985.

Erickson, R. New Music and Psychology. In Deutsch, D. (Ed.), The Psychology of Music. Academic Press: London, 1982.

Forte, A. The structure of atonal music. Yale University Press: London, 1973.

Forte, A. & Gilbert, S. Introduction to Schenkerian Analysis. W. W. Norton & Co: Toronto, 1982.

Gamer, C. ET Setera: some temperamental speculations. In Browne, R. (Ed.), Music Theory: special topics. Academic Press: London, 1981.

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

Gregory, R.L. Mind in Science. Peregrine: London, 1981.

Halpern, A. R. & Zatorre, R. J. Identification, discrimination and selective adaptation of simultaneous musical intervals. Journal of the Acoustical Society of America, 65, S40(A), 1979.

Hankiss, E. Semantic oscillation: a universal of artistic expression. In The sign in music and literature, W. Steiner (Ed.). University of Texas Press: Austin, Texas, 1981.

Hansen, C., Kessler, E. & Shepard, R. N. Musical perception in Bali. Music Perception, 2, 131-165, 1984.

Hargreaves, D. The developmental psychology of music. Cambridge University Press: Cambridge, 1986.

Harman, A. & Milner, A. Man and his music. Barry and Jenkins: London, 1962.

Helmholtz,. H. On the sensations of tone. A. Ellis (trans.), Dover: New York, 1885/1954.

Hindemith. P. *The Craft of Musical Composition* (trans. A. Mendel). Schott & Co: London, 1945.

Hitch, G. J. Developing the concept of working memory. In Claxton, G. (Ed.), *Cognitive Psychology: New Directions*. Routledge & Keegan Paul: London, 1980.

Hood, M. *Patet in Javanese Music*. Wolters: Groningen, 1954.

Houtsma, A. J. M. Discrimination of frequency ratios. *Journal of the Acoustical Society of America*, 44, S383(A), 1968.

Hughes, M. A quantitative analysis. In Yeston, M (Ed.), *Readings in Schenker analysis and other approaches*. Yale University Press: Yale, 1977.

Johnson-Laird, P. *Mental Models*. Cambridge University Press: Cambridge, 1983.

Jones, M. R. A tutorial on some issues and methods in serial pattern research. *Perception and Psychophysics*, 39, 492-504, 1981.

Jones, M. R., Boltz, M. & Kidd, G. Controlled attending as a function of melodic and temporal context. *Perception and Psychophysics*, 32, 211-218, 1982.

Jones, M. R., Kidd, G. & Wetzel, R. Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*,7, 1059-1073, 1981.

Kallman, H.J. & Massaro, D. W. Tone chroma is functional in melody recognition. *Perception and Psychophysics*, 26, 33-36, 1979.

Katz, A. *Challenge to Musical Tradition: A New Concept of Tonality*. Alfred E Knopf: New York, 1945.

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

Krumhansl, C. L. Perceptual structures for tonal music. *Music Perception*, 1, 28-62, 1983.

Krumhansl, C. L. Perceiving tonal structure in music. *American Scientist*, 73, 371-378, 1985.

Krumhansl, C. L., Bharucha,J. J. & Castellano, M. A. Key-distance effects on perceived harmonic structure in music. *Perception and Psychophysics*, 32, 96-108, 1982.

Krumhansl. C. L., Bharucha J. J. & Kessler E. J. Perceived harmonic structure of chords in three musical keys. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 24-36, 1982.

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

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

Krumhansl, C. L. & Shepard, R. N. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594, 1979.

Lerdahl, F. & Jackendoff, R. *A Generative Theory of Tonal Music*. MIT Press: Cambridge, Mass., 1983.

Lindley, M. *Lutes, viols and temperaments*. C.U.P.: Cambridge, 1984.

Lindsay, P. H. & Norman, D. A. *Human information processing*. Academic Press: London, 1977

Lloyd, Ll. S., & Boyle, H. *Intervals, scales and temperaments*. Macdonald & Co.: London, 1963.

Longuet-Higgins, C. H. The perception of music. *Proceedings of the Royal Society, London*, Series B 205, 307-322, 1979.

Luce, R. D. & Green, D. M. Two tests of a neural attention hypothesis for auditory psychophysics. *Perception and Psychophysics*, 23, 262-271, 1978.

McAdams, S. & Bregman, A. Hearing musical streams. *Computer Music Journal*, 3, 26-44, 1979.

Marr, D. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco, 1982.

Moore, B. C. J. *Introduction to the psychology of hearing*. (2nd edition). Academic Press: London, 1982.

Munsinger, H. & Kessen, W. Stimulus variability and cognitive change. *Psychological Review*, 73, 164-178, 1966.

Neisser, U. — *Cognition and Reality*. W. H. Freeman: San Francisco, 1976.

Nettl, B. — *Folk and traditional music of the Western continents* (2nd ed.). Prentice-Hall: New Jersey, 1977.

van Noorden, L. — *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven, 1975.

Ortmann, O. — On the melodic relativity of tones. *Psychological Monographs*, 35, 1-47, 1926.

Palisca, C. — Entry on "Theory", *Grove's Dictionary of Music*, Macmillans: London, 1980.

Parncutt, R. J. — *Sensory bases of harmony in Western music*. Unpublished Ph.D thesis, University of New England, Australia, 1986.

Parncutt, R. J. — Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception*, 6,65-94, 1988.

Patterson, R. D. — Spiral detection of periodicity and the spiral form of musical scales. *Psychology of Music*. 14, 44-61, 1986.

Posner, M. — *Cognition: an introduction*. Scott, Foresman and Co: Brighton, 1973

Povel, D. J. & Okkerman, H. — Accents in equitone sequences. *Perception and Psychophysics*, 23, 262-271, 1978.

Powers, H. — Entry on "Mode", *Grove's Dictionary of Music*, Macmillans: London, 1980.

Rahn, J. — *Basic atonal theory*. Longman: London, 1980.

Reed, S. K. — Schemes and theories of pattern recognition. In Carterette, E. C. and Friedman, M. P., *Handbook of Perception: Vol 8, Perceptual Coding*. Academic Press: New York, 1978.

Reese, G. — *Music in the Middle Ages*. Dent: London, 1941.

Ritsma, R. J. — Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America*, 42, 191-198, 1967.

Rosch, E.	Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233, 1975.
Rosen, C.	*Schoenberg*. Fontana: Glasgow, 1976.
Sanford, A. J.	*Cognition and cognitive psychology*. Weidenfield and Nicholson: London, 1985.
Schenker, H.	*Free Composition*. (trans E. Oster). Longman: London, 1935/1979
Schenker, H.	*Harmony*. (trans E. M. Borgese, ed. O. Jonas), University of Chicago Press: Chicago, 1954/1980.
Schoenberg, A.	*Structural functions of harmony*. (rev. ed.). W. W. Norton/E. Benn: London, 1969
Scrulowicz, P. & Goldstein, J. L.	A central spectrum model: a synthesis of auditory nerve timing and place cues in monaural communication of frequency spectrum. *Journal of the Acoustical Society of America*, 73, 1266-1276, 1983.
Sergeant, D.	The octave - percept or concept. *Psychology of Music*. 10, 3-18, 1983.
Shaffer, H.	Rhythm and timing in skill. *Psychological Review*, 89, 109-122, 1982.
Shepard, R. N.	Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, 36, 2346-2353, 1964.
Shepard, R. N.	Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305-333, 1982a.
Shepard, R. N.	Structural representations of musical pitch. In Deutsch, D. (Ed.), *The Psychology of Music*. Academic Press: London, 1982b.
Siegel, J. A. & Siegel, W.	Absolute identification of notes and intervals by musicians. *Perception & Psychophysics*, 21, 143-152, 1977.
Simon, H. A.	Complexity and the representation of patterned sequences of symbols. *Psychological Review*, 79, 369-382, 1972.
Simon, H. A. & Sumner, R. K.	Pattern in music. In Kleinmuntz, B. (Ed.) *Formal representation of human judgment*. Wiley: New York, 1968.

Solso, R. L. & McCarthy, J. E.	Prototype formation of faces: a case of pseudomemory. British Journal of Psychology,72, 499-514, 1981.
Terhardt, E.	The two-component theory of musical consonance. In E. F. Evans & J. P. Wilson (Eds.), Psychophysics and physiology of hearing. London Academic Press, 381-390, 1977.
Terhardt, E.	Psychoacoustic evaluation of musical sounds. Perception & Psychophysics, 23, 483-492, 1978.
Terhardt. E., Stoll, G., & Seewann, M.	Algorithm for extraction of pitch and pitch salience from complex tonal signals. Journal of the Acoustical Society of America, 71, 679-688, 1982.
Terhardt, E., Stoll, G. & Seewann, M.	Pitch of complex signals according to virtual-pitch theory: Tests, examples and predictions. Journal of the Acoustical Society of America, 71, 671-678, 1982.
Thomassen, J. M.	Melodic accent: experiments and a tentative model. Journal of the Acoustical Society of America, 71, 1596-1605, 1982.
Tversky, A.	Features of similarity. Psychological Review, 84, 327-352, 1977.
Vitz, P. C. & Todd, R. C.	A coded element model of the perceptual processing of sequential stimuli. Psychological Review, 76, 433-449, 1969.
Werner, H.	Musical "micro-scales" and "micro-melodies". The Journal of Psychology, 10, 149-156, 1940.
West, R. Howell, P. & Cross, I.	Modelling perceived musical structure. In Howell, P., Cross, I. & West, R. (Eds.), Musical structure and cognition, Academic Press: London, 1985.
Wright, J. K. & Bregman, A.S.	Auditory stream segregation and the control of dissonance. Contemporary Music Review, 2(1), 1987.
Zajonc, R. B.	Feeling and thinking: preferences need no inferences. American Psychologist, 35, 151-175, 1980.