CHAPTER 3

INFORMATION PROCESSING IN MUSIC PERCEPTION

Chapter Overview

There are ten sections in Chapter 3. Following a brief introduction, the next two sections consider evidence for successive and simultaneous coding in music information processing. *Section 3.4* reviews evidence for higher order musical constructs arising particularly from simultaneous coding. *Section 3.5* examines the role of attentional control in music information processing, while *Section 3.6* considers the hierarchical interaction of successive, simultaneous and executive syntheses in music perception. *Section 3.7* examines how attentional processes create expectancies and musical enjoyment through informational redundancy. *Section 3.8* considers the contribution that perceptual learning makes to musical abilities. The ninth section presents a review of the literature concerning the veracity of musical aptitude or musical potential as a construct to explain individual differences in music abilities. The final section maps the relationships of terminology concerned with musical abilities.

3.1 Introduction

This chapter reviews some of the literature on music perception and musical abilities from an information processing perspective. The review categorises evidence for simultaneous, successive and attentional processes, following the structure of the Luria model. It should be noted that the terms "simultaneous" and "successive" as used in the music cognition literature to describe features of musical organisation such as harmony and melody, are not necessarily synonymous with their use as processing category labels in the Luria model. Bregman (1994) notes that:

> *Sequential and simultaneous organisation help to create many aspects of musical experience. Sequential grouping is the foundation of rhythm and of many aspects of melodic form, whereas simultaneous grouping is involved with such experiences as chord perception, timbre, consonance, and dissonance* (p. 675).

Nevertheless, it will be argued that successive and simultaneous organisations in music do often require successive and simultaneous dimensions of perceptual processing. Wuthrich and Tunks (1989), for example, note that melodic and harmonic intervals "generate contrasting psychological experiences" (p. 32). Dowling (1971) argues that in melodic contour preservation and inversion tasks "contour and interval size are handled in different and largely independent ways in cognitive processing" (p. 349), including separate encoding into memory (Davies & Yelland, 1977).

Furthermore, it will be argued that performance on musical tasks requires the application of a mix of simultaneous, successive and executive syntheses, usually in a cyclic hierarchy of processing, where successive elements at one level are simultaneously processed into units which become the elements for the next higher level of successive synthesis (Kirby & Das, 1990; Das, Naglieri and Kirby, 1994) [Chapter 2]. For example, in studies into melodic memory Dowling (1991) found that

> *listeners did not ignore interval information in encoding contour, and the contours of melodies that matched tonal scale invariants of the culture were best remembered. This strongly suggests that melodies are heard as integrated wholes in which the various perceptual features interact* (p. 313).

Research over the past twenty years on individual differences in the abstraction of melodic pattern has been extensive, and has been undertaken from both neuropsychological and psychological perspectives. Despite such endeavour, Barsz (1988) notes that "a comprehensive theory has yet to emerge" (p. 293) because of important conflicts in the literature. One important conflict is over the contribution of innate aptitude to individual differences in musical abilities. Another is over whether psychological constructs are isomorphic to musical constructs. An information processing approach attempts to resolve such conflicts.

3.2 Successive synthesis in music perception

There is considerable evidence for successive synthesis in music perception (e.g., Barsz, 1988; Brown, 1987; Brown, 1988; Deutsch, 1991; Monahan, Kendall & Carterette, 1987; Sorkin, 1990; Warren, Gardner, Brubaker, & Bashford, 1991). Auditory information is primarily grouped by temporal dictates (McAdams, 1987; West, Cross & Howell, 1987; Wuthrich & Tunks, 1989). This is the case even when the musical stimulus is multi-streamed (Bregman, 1994; Sorkin, 1989) or genuinely chordal (Allik, Dzhafarov, Houtsma, Ross & Versfeld, 1989; Dai, 1992). Monahan, Kendall and Carterette (1987) note that ability in serial pattern recognition is consistently higher in music experts than in novices.

In the Luria model, successive processing by the second functional unit is required for melodic memory. Luria (1973) notes that modally-specific syndromes can occur from lesions of the second block.

> *In the case of lesions of the middle zones of the temporal region ... the patient could no longer retain lengthy series of sounds or words. ... Acoustico-mnestic aphasia ... was also manifested as difficulty in retaining groups of musical notes and rhythmic structures, although there was no difficulty in retaining a series of visual images or a series of movements* (pp. 297-98).

If the regions responsible for speech and music are independent, or at least separate, then we might expect little correlation between language skills and music abilities. Luria (1973: 132-135) emphasises the role of the linguistic environment in determining culturally relevant and irrelevant phonetic cues. The acquisition of an internalised scale or absolute pitch could be a similar process. If so, the mnestic zones of the second block would be expected to make an important contribution to this process. In Chapter 2 it was noted that within the operationalised model, Kirby and Das (1990) conceptualise successive synthesis as a measure of the capacity of working memory. Hantz *et al* (1992) explained differences in pitch discrimination between musicians with and without absolute pitch by differential usage of working memory. EEG evidence suggested that subjects with absolute pitch did not using working memory to evaluate intervals in contrast to non-absolute pitch subjects who did. Recourse to absolute pitch may rely on long term memory in contrast to updating working memory for each new musical context. Those without absolute pitch require a temporal or sequential context in which to make pitch discriminations.

Bamberger (1991) investigated the concept of a "tune path" (p. 141). She notes that with the task of singing from the middle of a song, most people go back to the start. Here is a clear example of successive synthesis at the mnestic level. Bamberger studied how young children construct rows of bells to make familiar tunes. Many children repeated bells of the same pitch, i.e., made a row which was isomorphic with pitch sequence. Some children, later, used the minimum number of different pitched bells as in a musical instrument. These results demonstrate the use of successive synthesis by young children, and the development in some children from just using successive processing to include the use of simultaneous processing on this task.

Brown (1987, 1988) argues that many of the perceptual models in the literature have little relevance for listening to real music because of the emphasis on psychoacoustical analysis over temporal context. Such analyses, typically using musically sophisticated subjects, do not, Brown suggests, inform our understanding of "the mental representations of music shared by a much broader population of listeners" (Brown, 1988:221). In Brown's studies, subjects' perceptions of tonality were dependent on the serial ordering of tones. "While tonality certainly includes relationships of several tones to a single tone it is critically dependent on listeners' contextual interpretations of relations among all tones, as expressed in musical time" (Brown, 1988: 245). Psychoacoustical reductionism is insufficient to account for perception of tonality. Such perception is "predicated on ... a functional hierarchy formed by the perceptual prominence of contextual relationships of pitches occurring in the dimension of musical time" (p. 246). Allik, Dzhafarov, Houtsma, Ross and Versfeld (1989) argue that expectancies for pitch motion are determined by the succession of all previous intervals, i.e., as the result of ongoing successive processing. While conceding Brown's point of relevance to "real" music, Monahan, Kendall and Carterette (1987) note that "there are usually too many variables in performed music to attribute variance reliably" (p. 599). Results from their "middle course" suggest that perceptual prominence in musical sequences is attained from accented tones, usually at the beginning of musical phrases. Such prominence is interpretatively modified in performance by the use of *rubato*.

Both Janata and Reisberg (1988) and West and Fryer (1990) suggest that perceptual prominence may be explained by recency effect. Recency explains prominence by order of arrival, and emphasises "musical landmarks" such as cadential points. "Recency advantages depend on how the subject understands and organises serial information" (Janata & Reisberg, 1988: 169). This in turn determines listener expectations, particularly for modulation. Response times to tonal probes after scale and chord stimuli showed preference for tonic and dominant, and a general profile which reflected diatonic importance. Janata and Reisberg interpret this as indicating an interaction between recency and "top-down" musical schema.

Warren, Gardner, Brubaker and Bashford (1991) found there was a temporal range of about 150 to 1000 msec in duration beyond which melodic tone sequences could not be recognised. Too slow and "the sequence of notes lost melodic cohesion and the notes were heard as independent sounds" (p. 280). Even musically trained subjects could not recognise common tunes at high speed. All subjects could recognise non-melodic sequences at much briefer durations than 150 msec. This, the authors suggest, indicates that melodic processing has organisational "rules" which do not apply to non-melodic sequences. There is a similar time-rate of detection of syllables in speech.

Deutsch (1988, 1991) and Barsz (1988) report studies into the phenomenon of pitch proximity in perceptual organisation. Melodic lines are perceptually reorganised into sequential groups on the basis of proximity of pitch class. Sequentially and simultaneously presented tones from two different pitch ranges are perceived as two separate melodic lines.

> *The auditory system forms a temporal compound of all sounds presented within some time frame ... and subdivides the compound into streams consisting of relatively similar sounds. Within-stream relationships are postulated to contribute more to recognition than between-stream relationships* (Barsz, 1988:302).

Such local correspondence was shown to operate in the detection of melody from random chord sequences (Allik *et al*, 1989). Such a phenomenon is analogous to perception of motion from serial change of object coordinates. The authors suggest that "in both cases, we are dealing with a general cognitive brain process that extracts information from visual and auditory inputs in a similar manner" (p. 523). Bregman (1994) agrees. He argues that the phenomenon of auditory streaming, which is the basis for the organisation of auditory information, has direct parallels with the visual streaming involved with illusions of apparent motion. It is through streaming that "sequential organisation is responsible for the perceived shape of melodies" (p. 467). Bregman argues that encoding must integrate a sequence of sounds into a distinctive sequential form for memorisation to occur. Similarly Bartlett (1984) argues that sequential stimuli are formulated into a coherent Gestalt. It is such auditory grouping processes which underlie the compositional principles of voice leading and orchestration. Bregman notes that this requirement can be problematic for contemporary composers, particularly those working with computer-generated music where primitive grouping processes do not readily integrate sounds into recognisable musical forms. "Any composer who is too demanding in this respect may produce works which are magnificent to the highly trained ear, but may never be shareable with a community of any size" (p. 468). Bregman states two conditions for sequential organisation to occur: first, that musical lines should be well separated in pitch and not cross; and second, that sequential integration is favoured by small changes in pitch between successive notes.

Pitch proximity might explain the resolution phenomenon investigated by Boltz (1989). Her results confirm that perception of the end of a musical piece is maximised with the ubiquitous temporal ordering of leading note to tonic. Like Monahan, Kendall and Carterette (1987), Boltz found accentuation, such as prolongation of the final note, to play a significant role in enhancing such temporal ordering. She notes parallels with speech processing.

> *At least in declarative statements, a speaker's pitch intonation tends to decline and return to the fundamental frequency at the end of an utterance ... In addition, the duration of the final word is prolonged relative to preceding words ... This suggests that at some level there is a common scheme for using certain structural invariants in the auditory environment for specific functions* (p. 751).

This further suggests that some aspects of speech and music are encoded similarly, not withstanding Luria's position that speech and music stimuli are encoded independently.

3.3 Simultaneous synthesis in music perception

There is also considerable evidence for simultaneous synthesis in music perception (e.g., Deutsch, 1991; Elliott, Platt & Racine, 1987; Jackendoff, 1991; Kendall & Carterette, 1991; Patterson, 1990). Simultaneous synthesis is especially germane to the perception of chords and harmonics (Allik *et al*, 1989; Brown, 1987; Demany & Semal, 1988; DeWitt & Crowder, 1987; Houtsma & Smurzynski, 1990; Hubbard & Stoeckig, 1988; Patterson, 1990; Platt & Racine, 1990; Schmuckler, 1989). Moreover, the perception of some aspects of musical form requires simultaneous synthesis of information originally processed by successive synthesis (Allik *et al*, 1989; Andrews & Dowling, 1991; Boltz, 1991; Cohen, Trehub & Thorpe, 1989; Davies & Yelland, 1977; Dowling, 1991; Monahan, Kendall & Carterette, 1987; Terhardt, 1991).

Luria does not describe amusical syndromes in terms of simultaneous processing. He does note that the symbolic aspects of music perception require the functioning of the second functional unit. "Lesions in the parieto-occipital zones of the dominant left hemisphere can also lead to a disturbance of the perception of ... symbols (including) musical notation" (Luria, 1973: 237). Other researchers, however, have shown positive relationships between visual spatial and musical processing abilities. Karma (1979) undertook a factor study of students' abilities to perceive musical structures. Music structuring scores of students with more than two years music training correlated 0.34 with spatial and 0.04 with verbal abilities, compared with students with less than two years music training, 0.14 with spatial and 0.18 with verbal, and non-trained students, 0.06 with spatial and 0.45 with verbal. Karma suggests that music training, utilising such quasi-spatial concepts as "descending melody", "high notes", and "intervals between tones", "increases the capacity to use this kind of thinking" (p. 51). This suggestion has support from a recent study in which the spatial skills of pre-school children were improved following the intervention of a music program which concentrated on the keyboard (Demorest, 1994). Part of Hassler's (1990) data shows that musicians exhibit enhanced spatial performance over non-musicians. "The outcome is ... in line with [an] hypothesis which postulates that special [musical] abilities are based on a set of skills that involve exact and extensive representation of visual images and the rapid recognition of patterns involving visual and auditory representations" (p. 12).

Godoy (1994) proposes a model for the perception of global properties in music based on the phenomenology of Hurssel:

> *For the comprehension of a sequence of representations ... it is necessary that they be the absolutely simultaneous objects of a referential cognition which embraces them completely and indivisibly in a single unifying act ... Such representations would all be impossible if the act of representation itself were completely merged in temporal succession* (Miller, 1982: 132-133).

In support of this position, Dai and Green (1992) found that discrimination between different auditory stimuli was better when the stimuli were presented simultaneously than when presented successively. The authors suggest that signal noise can be better correlated out when stimuli are heard together. In contrast, Wuthrich and Tunks (1989) present evidence for the perception of intervals through onset asynchronies, typical of performed ensemble music. Subjects found interval recognition easier for successive intervals than for simultaneous. The authors suggest that

> *tones sounding at the same time and tones sounding in succession are two different experiences. ... successive tones require auditory memory to be able to relate to them. [Although] this is not necessary for tones sounding together ... simultaneous sounds must be distinguished to judge interval width* (p. 43).

Patterson (1990) notes that the auditory system uses two modes of analysis: analytic, which changes from sound to sound, and synthetic. "In the synthetic mode, the system focuses on the complete auditory image produced by the set of components as a whole. [This] mode is more common in music perception and in the natural environment generally" (p. 207). This is consistent with Luria's description of simultaneous processing as being based on commonalities between the elements (e.g., Luria, 1978: 149-150). The use of two coding dimensions in music could explain a "puzzling" result of Dowling (1994b). Dowling found that discrimination between target melodies and same-contour lures was better over various time delays than discrimination between target melodies and different-contour lures. The result is not surprising if the two discrimination tasks are dependent on different dimensions of information processing. Discriminating between different contour melodies could require a predominance of successive processing of contour time series (serial elements of "up", "down" or "same"), whereas same contour melodic comparisons could require predominantly simultaneous synthesis of differences in pitch.

Gardner (1982) suggests that Mozart's extraordinary facility for composition may have relied heavily on simultaneous synthesis. Gardner quotes from a letter written by Mozart on his manner of composing.

> *... my subject enlarges itself, becomes methodised and defined, and the whole, though it be long, stands almost complete and finished in my mind, so I can survey it, like a fine picture or a beautiful statue, at a glance. Nor do I hear in my imagination the parts successively, but I hear them, as it were, all at once* (p. 358).

Gardner comments:

> *Mozart ... in his claim to hear a whole piece simultaneously ... did not mean this statement literally. ... Rather, Mozart seems to have employed a metaphor here, but one of some accuracy and power. What Mozart meant ... was that the entire organisation, the distinctive architecture of the piece of music was fully articulated in his mind. The crucial decisions ... could all be grasped at one time* (p. 365).

3.4 Simultaneous synthesis and long term music memory

To explain the phenomenon of musical memory, much of the music cognition literature has sought evidence for internal templates of scales, intervals or pitches as the mental products of music training. The basis of a template is a construct of acoustico-spatial or simultaneous acoustic synthesis stored in long term memory. Davies and Yelland (1977) conclude "that the presence of a stored tonal representation is the critical variable in the memory of tunes" and "the ability to form stable, accurate internal representations is amenable to practice" (p. 8). Tsuzaki (1991) investigated the effect that the context of a preceding scale had on interval judgement. She notes that "the explanation only in terms of short-term memory trace is not sufficient" (p. 66).

In an interval tuning task, Elliott, Platt and Racine (1987) found that musically trained subjects made fewer errors in achieving just tuning than non-trained subjects, and more so on consonant intervals. The results are interpreted to support the development of internal standards for intervals by musically experienced subjects. The strength of the internal standard for each interval reflected the relative frequency of that interval in Western tonal music. This was supported in a study by West and Fryer (1990) in which music students did no better than non-music students at choosing the tonic after a random ordering of diatonic scale notes. Both groups showed preference for major over natural minor resolution with about equal preference for dominant, sub-dominant, mediant and tonic over sub-mediant (minor tonic) and other scale degrees. Cuddy (1991) reports evidence that

> *supports the notion that the degree of tonal structure conveyed by short melodic patterns is dependent on the ease with which a pattern can be mapped on a stable, abstract, internally consistent representation of the hierarchical pitch relationships of Western tonal music* (p. 123).

Musically trained subjects rated major triads over other triads for indication of tonality suggesting, in accord with West and Fryer (1990), "that the major triad is prototypic of tonal structure" (p. 107). Cuddy's experiment was designed to meet criticisms by Butler (1990) who argues that "perceptual activity of key assessment must be simple and intuitive rather than complex and overtly analytical" (p. 16). Butler presents evidence for a model of "intervallic rivalry" in which "the clearest tonal information is conveyed by the clearest musical realisation of the *rare* intervals of the diatonic set" (p. 9) [my emphasis]. Butler claims that real music does not present major triads and scale portions in a tonally unambiguous manner.

Dowling (1991) refutes Butler's claim by suggesting it to be a Gestalten error. Dowling agrees that context is important, but argues that it is the "preponderance of relatively predictable elements" that establishes that context. "A tonally strong pattern is one that clearly represents the predictable, prototypical aspects of tonality" (p. 307). Major tonality is favoured because the intervals of a major scale are more regular than those of minor scales and thus are more predictable (Bartlett & Dowling, 1988). "It follows that octaves should be the most preferred intervals and those intervals absent from the harmonic series or only weakly represented, the least preferred. Thus ... experience can be fully consistent with the acoustical properties of musical tones" (DeWitt & Crowder, 1987: 83). It also follows that the strongest template should be for the octave. Studies of tonal fusion, where the interpenetration of multiple tones creates a fused whole which is different from the sum of its components, confirm such a prediction; octaves were perceived to fuse more than other intervals (Demany & Semal, 1988; DeWitt & Crowder, 1987; Wuthrich & Tunks, 1989). Interval templates, it is suggested, are used by listeners to establish constructs such as keyality. Cohen (1991) reports investigations into perceptions of keyality in which the musical context was the first four notes of each of the twelve Preludes, Book 1, of J. S. Bach. Preludes with major tonality were less problematic and major tonality was preferred by subjects in general.

The evidence for a tonal template presented thus far is based on Western music. Pressing (1983) found similar musical structures in musics from African, Asian, Oceanic and Eastern European cultures. Pressing describes these similarity relationships as 'cognitive isomorphisms'. "The observed structural similarities are sufficiently compelling, and their relation to musical perception and training sufficiently direct, to justify the hypothesis that they may result from general cognitive processes" (p. 38). Such isomorphisms include West African rhythm patterns (e.g., crotchet, crotchet, crotchet, quaver, crotchet, crotchet, crotchet, quaver, crotchet, notated as 2212221) with the major scale (tone, tone, semitone, tone, tone, tone, semitone, notated as 2212221 in semitones), and similarly the rhythm pattern 22323 with the pentatonic scale. Isomorphisms can also be found for fundamental harmonic relations such as the tonic-dominant. "Why these few special patterns have been chosen by so many apparently independent musical cultures ... is not an easy question. ... this commonality must tell us something underlying about perception and the mind" (p. 44). Krumhansl (1992) agrees: people are reasonably good "musical tourists". She reports evidence that listeners do not import culture-specific tonal schemas to the experience of listening to music from other cultures. It seems, then, that the possibilities for templates could be constrained by cognitive functioning or neural structure, and that some musical patterns are more suitable for long-term memory storage. A characteristic of the above isomorphic patterns in contrast to others, Pressing notes, is "their property of sampling as equally as possible the various subdivisions of [the octave]" (p. 46). Such sampling may represent the most efficient arrangement for chunking by simultaneous synthesis of successive intervals into a scale template or schema.

On the other hand, Bregman (1994) points out that there are many occasions where the simultaneous organisation of sounds is not integrative, e.g., a soloist playing with an orchestra is usually heard 'above' the accompaniment; a cough heard during a performance is disregarded from the musical input. Simultaneous organisation becomes more demanding when having to determine the concordance of a coincident group of notes with a large range of musical possibilities: a chord, a discord, a 'blue' note against a chord, two chords with deliberate bitonality, etc. Bregman suggests that the auditory system must derive global properties from cues which favour vertical integration. When such cues are absent then the auditory system treats the coincident sounds as belonging to separate streams rather than grouping as a dissonance. The cues, Bregman suggests, include "spectral proximity, harmonic concordance, synchronous changes, freedom from sequential capturing, and spatial proximity" (p. 509). Bregman argues that primitive grouping processes work quite separately, and on occasions, in opposition to higher-order schema-based processes. "Primitive processes *partition* the sensory evidence whereas schema-based attentional ones *select* from the evidence without partitioning it" [his emphasis] (p. 669). In the Luria model, it is one of the roles of executive synthesis to resolve such potential cognitive conflict.

3.5 Executive synthesis in music perception

There is considerable evidence for the importance of executive synthesis in music perception. In the Luria model, the third functional block acts on and synthesises information from the second block. Janata and Petsche (1993) found electroencephalographic evidence that cerebral organisation for expectancy fulfilment and violation changes at sites above the auditory cortices, as well as above the right parietal areas and the frontal lobes. "Presumably, attention is being focussed at salient moments in a perceived rhythmic structure, causing increased processing of events surrounding these moments" (p. 303). This result supports the involvement of Luria's third functional block in the generation of expectancies, as well as the involvement of both successive and simultaneous syntheses in the processing of musical stimuli.

The third block also initiates and integrates affective responses. Webster and Richardson (1994) note that "musical thinking means making sense of the perceived sounds through such activities as grouping, comparing, predicting ... and evaluating" (p. 10). Importantly,

> *musical thinking is special because it can represent a powerful blending of cognitive (factual) content with affective (feelingful) content in ways that few other disciplines can. In Western thought, it is common to think of these as separate entities of the human psyche that have nothing to do with each other. This is far from the truth* (p. 11).

Executive synthesis also includes the resolution of cognitive conflict. Pogonowski (1989) defines "executive control [as] evaluating, planning and regulating the declarative, procedural and conditional information involved in a task" (p. 18), which requires audiation of alternate musical hypotheses in preparation for a musical plan. McAdams (1987) notes the role of executive synthesis in the resolution of cognitive conflict.

> *Even simultaneous and sequential grouping processes can conflict with one another creating situations with multiple perceptual interpretations. The resulting perceived qualities of the sources depend on the way the conflict is resolved ... It is clear that active and passive attentional processes can play a strong role in the resolution of these conflicts,*

> *particularly at higher levels of grouping where functional ambiguities resulting from conflicting "vertical" (harmonic) and "horizontal" (melodic) organisations can be of great musical value* (p. 43).

Thus it is attention that underpins all the functional responsibilities of the third functional block. There is consistent evidence for the importance of attention or focussing in music cognition (Andrews & Dowlling, 1991; Boltz, 1991; Clarke & Krumhansl, 1990; Morgan & Brandt, 1989). The role of attention includes the determination of expectancy (Dowling, 1990; Janata & Petsche, 1993; Jones, 1992) and the formulation of aesthetic responses (Bever, 1988; Madsen, Byrnes, Capperella-Sheldon & Brittin, 1993). Jones (1992) argues that it is through attention that the various perspectives of the composer, performer and listener can become shared.

Luria (1973) defines attention as "the directivity and selectivity of mental processes, the basis on which they are organised" (p. 256). Attention can determine which are the "essential elements for mental activity", whether from external stimuli or memory traces. Attention can also turn inward to monitor "the precise and organised course of mental activity" (p. 256). An essential component of involuntary attention is the orienting reflex which, in contrast to the general arousal reactions of the precortical block, "may be highly directive and selective in character" (p. 259). Luria notes that orienting reflexes are invoked whenever there is a mismatch between a new stimulus and its mental model. This has important implications for expectancy. Expectancy is active and a function of the complexity of the task at hand. The orienting reflex particularly requires change: an increase or decrease or sudden absence of stimulus. Luria (1973) points out that the amplitude of evoked potential increases with "active expectancy or complication of task" (p. 267). Thus music with complete predicability is either uninteresting and can be ignored (e.g., Muzak), or allows attention to focus on other things, such as dancing. This line of argument could explain the "sub-arousal" states induced by repetitive dance rhythms.

Voluntary attention, in contrast, is social in origin. Early childhood socialisation determines how initial involuntary attention to features in the world is given psychological organisation through language. Thus a bridge is constructed "between the elementary forms of involuntary attention and the higher forms of voluntary attention, thus preserving their unity" (p.263). This has important implications for the effects of music experiences in early childhood. West, Cross and

Howell (1987) note the role of language in delineating musical concepts, and its more frequent use among trained musicians. "Musical training provides an ability to put into words our experience of functional relationships between objects in musical experience" (p. 8).

Morgan and Brandt (1989) showed an auditory Stroop effect for pitch using verbal stimuli, "high" and "low" at congruent and non-congruent pitch levels. Dichotic presentation of stimuli did not result in significant differences, supporting the sagittal division in the Luria model of the third functional block. Wolfe and Hom (1993) report further evidence that music can be a positive medium for facilitating the recall and retention of non-musical information, such as words, action sequences, and number strings in young children. They found no evidence that the simultaneous attention to number strings and music pieces caused interference or inhibitory effects on learning; in fact, just the opposite. This evidence supports the independence of executive synthesis from coding processes in the operationalised Luria model.

Evidence for the cognitive development of executive processes in young children [Chapter 2] through the domain of music is presented by Andrews and Dowling (1991). They found the ability of children to attend to target melodies, whose notes were interleaved with distractor notes, improved with age. Whereas 5-6 year olds found all items difficult, 9-10 year olds attained mastery on the instrument. "As children grow older, they become able to control attention by focussing it on specific pitch regions where target events are likely to occur. They also shift from reliance on broad features such as melodic contour to subtler features such as tonality and precise interval size" (p. 367). Success at the most difficult tasks by older children and adults was attributed to their use of expectancy over attention to locate target melodies. Consequently, Dowling (1990) proposes an operationalised model for attentional processing. He reserves the label "attention" for the selection of stimuli for further processing. Expectancy has two forms: procedural expectancies, which are processed "automatically", and "declarative" expectancies, which are explicit. "Expectancy selects elements that match expectancies, and elements that depart markedly from expectancies are lost to processing" (Dowling, 1990:160). However, an alternate interpretation of Andrews and Dowling's results could be that the interleaved notes violated chaining procedures necessary for successive synthesis. The difficulty in melodic recognition in these circumstances may not, therefore, be restricted to shortcomings in attentional processes.

Unyk (1990) outlines an information approach to expectancy in music cognition that highlights the role of selective attention in creating and being guided by expectancies. Much of the information from the musical signal temporarily stored in a sensory memory is not selected for further processing. "Attention is conceptualised as a limited resource; we cannot attend to all stimuli which impinge on our senses" (p. 231). Unyk presents evidence for automatic processing in music perception. Much of our current perception is guided by processes which, she argues, have been established through musical exposure in infancy.

Jones (1992) and Boltz (1991) also present evidence for the generation of expectancy through attentional processes. Jones argues that attention is both "inherently temporal" and "flexibly selective" (pp. 92-93). In converse to Dowling (1990), who argues that "attention ... is difficult to split" (p. 154), Jones posits two mutually exclusive sorts of attending, analytic and future-oriented, based on short and longer-term time intervals respectively. Musical organisation "represents a recursive temporal embedding structure. ... there exist relations among the different embedded time levels that listeners come to rely on and which can be said to control attention" (p. 94). It is analytic attention which determines the perception of pitch streams (Andrews & Dowling, 1991; Deutsch, 1991), whereas future-oriented attention creates expectancy through the dependence of pitch events on their temporal location (Dowling, 1990). Thus the conclusion of Andrews and Dowling (1991) above could be interpreted as older subjects using future-oriented attention over analytic attention in selecting melodies from background or competing information.

Temporal organisation, then, is a determinant of attention (Clarke & Krumhansl, 1990), and consequently, is critical for music perception. Jones (1992) argues that the shared perspectives between listener, performer and composer are temporally dependent and thus open to multiple interpretations. She further suggests that the divergent outcomes of many of the studies reported in the music cognition literature may be attributed to uncontrolled subject attention.

> *Without a corresponding concern for the time structure of the event and the way in which a temporal context guides a listener's attending over one or another time level, the search for the "right" mental representation of a musical creation will be in vain* (p. 105).

For example, Boltz (1991) showed that subjects' estimation of time intervals was dependent on the sort of attention applied to the task. An advantage of using the operationalised Luria model in this research program was that it provided measures of attention, so that other cognitive processes could be investigated in an attention-controlled manner where appropriate.

Boltz (1991) argues that attention can be "down shifted" to the analytic level in particular circumstances, such as when listening to a novice performer who cannot maintain temporal integrity. One's attention is shifted away from musical meaning towards local information such as the correctness or otherwise of individual notes, intonation etc. Novelty can also enforce an attentional down shift. "A lack of familiarity or perceptual learning may initially encourage this mode of attending because the intrinsic organisation of the event has not yet been discerned. This frequently happens, for example, when one hears a new style of music" (Boltz, 1991: 431).

Shifting attention may also contribute to the aesthetic response. Bever (1988) sought to explain the "essential puzzle of music ... that it is both abstract and emotionally powerful" (p. 165) with a model of alternating attentional processing. To make sense of a musical phrase, immediately after it has been heard, attention must switch from the external information of the note sequence to the internal processing which determines a phrase boundary. Such attentional oscillation creates an inner sense of disturbance. "A diligent listener of music ... realises that he is agitated, but does not have direct access to the mechanism causing the agitation: he intuitively attributes the agitation to a particular emotion" (p. 170). That is, emotions evoked while listening to music stem from pre-existing feelings in the listener which are unlocked by the cognitive activity generated by the changing attentional focus. Bever uses this model to explain individual differences in emotional responses to the same piece of music.

Madsen *et al.* (1993) present evidence that some music evokes similar emotional responses across different types of listeners. Specifically, "selections which are among the best from the Western art tradition are [aesthetically] processed by persons with limited formal training in very much the same manner as by those persons who are considered quite musically sophisticated" (p. 189). Bregman (1994) suggests that some compositions are more effective in "unlocking" similar emotions by recourse to principles of auditory grouping which are readily encoded into recognisable musical form.

LeDoux (1994) reports evidence that emotional learning from auditory stimuli does not necessarily require the auditory cortex. Lesions to the sub-cortical area of the amygdala completely inhibited learned fear responses in rats. The amygdala is located in Luria's first functional unit. Within the amygdala inputs are received from sensory organs and from the sensory areas of the cortex. The central nucleus directly transmits information processed by the basal nuclei to the brain stem for physiological responses such as increased blood pressure and cessation of gross movement. Thus, in contrast to much of the above, emotional responses can also be the outcomes of processing of the first functional unit alone.

3.6 Hierarchical information processing in music perception

There is support in the literature for the basic hierarchical organisation of the Luria model applied to music perception. Musical sense data is processed by the first functional block before undergoing a hierarchical cycle of coding into recognisable musical elements suitable for use in short term memory, and storage in the long term memory of the second functional block. Information in a remembered context is then available for processing by the third functional block in order to direct attention and create expectancies as a basis for aesthetic response. Krumhansl (1992) notes that the early processing of musical events is "assumed to be automatic and not under attentional control" (p. 201). Such processing occurs along the psychological dimensions of pitch, duration, loudness and timbre. As pitch and duration "appear to be more discrete and categorical in nature" they "enter more naturally into larger organisational units whose construction is rule governed" (pp. 203-204). Krumhansl's tonal hierarchy has the tonic above the other pitches of the tonic triad, then the remaining scale tones, and last the non-scale tones. The circle of fifths is a demonstrable cognitive construct. Thus pitch perceptions are contextually asymmetrical, which can explain the non-stationarity of pitch within a fluctuating keyality as experienced, for example, by choral singers attempting contemporary SATB pieces. An alternate explanation, however, could be that Krumhansl's hierarchy simply reflects familiarity with consonant intervals since dissonant scale tones, such as the diminished fifth, are rarely played against the tonic.

The notion of hierarchies is not universally supported. Cook (1994) argues against hierarchical grammars as the means by which music is perceivable. A linguistic tree builds from the bottom up, the top is the "datum", viz., an utterance or sentence. In music it is the converse: the bottom is the music, while the top is an imposed abstraction. Serafine (1988) argues that pitches, scales, chords etc., are not the building blocks of music. Cook agrees: "music is not notes." He rejects Krumhansl's tonal hierarchy with two pieces of evidence. The perception of tonal closure only had an effect for works of under a minute's duration - after this time there were no differences between music students' preferences for the composer's choice of tonic closure and non-tonic alternates. This suggests that chunking is limited by the capacity of the working memory. Replication of one of Krumhansl's continuation experiments, where a music class was asked to hum the tonic, resulted in "cacophony". Cook suggests that "musical perception is pluralistic and fluid; listeners make use of multiple cognitive frameworks and shift their strategies from one moment to the next. ... There is no reason why music theory should be like this" (p. 89).

Nevertheless, many models of music perception have been underpinned by an assumed degree of correspondence between cognitive and musical structures. At a basic level, both the elements of human perception and the representational elements of music are quantal or discrete. The granularity of expressions of time is of particular relevance to this relationship (Honing, 1993). Honing (1993) suggests that Serafine's argument for a continuous basis for representation "stands quite alone". There is also correspondence between the hierarchical organisation of musical elements and the hierarchical chunking of informational elements in human cognition (e.g., McAdams 1984, 1987). Lerdahl (1988) proposes two perceptual hierarchies in his Generative Theory of Tonal Music (GTTM). One is a tonal hierarchy which defines pitch class as a location within a quasi-spatial topology where pitches relate along intervallic dimensions; the other is a temporal hierarchy with time-span reduction, which expresses structure from rhythmic units, and prolongational reduction (Bigand, 1990; Jackendoff, 1991; Lerdahl & Jackendoff, 1983), "which expresses recursive patterns of tension and relaxation among events" (Lerdahl, 1988: 316). Lerdahl claims compatibility of GTTM with the earlier theories of Schenkler.

West, Cross and Howell (1987) outline four specific challenges for a hierarchical structure approach. First, the assumption of multi-level perception by listeners is not necessarily justified. Attention seems to vary during listening, so that the information perceived is, most likely, not all that the composer or performer intends. Second, hierarchical structures do not account for musical ambiguity, e.g., the chord ACE could be A minor tonic or C major VI. Jackendoff (1991) addresses this issue and concludes that processing must include multiple analyses in parallel. Third, West, Cross and Howell argue that all of a piece of music needs to be heard before the listener can complete a mental construction of a musical hierarchy. Information at any instant is constrained by a perceptual "temporal window" whose width is determined by short term or working memory capacity. The amount of parallel instantaneous processing required to fully analyse a musical segment under Lerdahl's structure is quite prodigious, and not apparently concomitant with the conscious act of listening. Fourth, the hierarchical model does not account for physiological and psychological differences between individual listeners.

Nevertheless, Terhardt (1991) argues for a hierarchical organisation of processing in which each level, from the sensory to the higher cognitive, extracts meaning from its stimuli. As such, information processing can be regarded as decision making. Importantly, decision making necessitates ambiguity. "This indeed is essential, as the solutions achieved on one level provide the input to the next. If that input did not have any alternatives, there were nothing left to decide" (Terhardt, 1991: 229). Such ambiguity, Terhardt claims, explains chromatic equivalence or pitch class. Furthermore, conflicting decisions at different levels give rise to illusions such as the semitone paradox (Deutsch, 1988), where pitch class interacts with the theoretically independent dimension of pitch height to create the illusion of temporal reversion. Interestingly, such illusions are subject to individual differences in musicians <u>and</u> non-musicians. Terhardt introduces the concept of virtual pitch as an auditory analogy to pre-attentive processing in the visual system to explain such illusions. Virtual pitch involves "subharmonic coincidence detection". The ability to form virtual pitch, Terhardt suggests, is acquired when very young through "the human speech signal that probably provides an important reference for aural evaluation of tonal sounds" (Terhardt, 1991:227).

Minsky (1981) provides an overview of this position which is consistent with Luria and Vygotsky on the social roots of attention. He argues that for young children, music provides a means of discovering time in a manner parallel to the way that playing with blocks is a vehicle for learning about space. Musical structure provides temporal experiences of "besides" "above" "within" that are readily afforded with objects in space. In particular, building larger structures from smaller is a characteristic of musical, and spatial, cognition. Similar processes occur at both experiential and cognitive levels. "Hearing music is like viewing scenery and ... when we hear good music our minds react in very much the same way they do when we see things" (p. 36). Minsky argues that hierarchical characteristics of vision such as feature detection, continuity of scene, and motion perception, can be seen in operation in musical comprehension. Minsky (1988) argues that higher order cognitive activities can only be explained through hierarchical self organisation in which no part knows the big picture, thus avoiding the infinite regress of the homunculus.

With his implication-realisation model Narmour (1990) also presents an extensive theoretical model for the hierarchical resolution of melodic expectancies.

> *the goal of music theory is, or should be: ... to construct a unified perceptual theory of the greatest psychological economy and the most elegant theoretical parsimony - one that will account for the whole created world of musical structure, the whole created world of musical affect, the whole created world of compositional strategy, and the history of music qua music* (p. 423).

The model defines musical style as the result of top-down structures acting on bottom-up processing of musical signals. "Our perceptual-cognitive mechanisms always operate on a two track expectancy system of parametric style shape and complex style structure" (p. 71). Whereas the top-down structures are hierarchical and complex, the signal parameters such as duration and harmony are perceptually primitive. Narmour argues that memory of a particular melody always involves structured realisations of these parameters. This structuring is the main focus of a listener's attention, and accounts for the facility of a listener to compare a musical input with previously learned melodies. The emerging piece is an implicative isomorph to a remembered piece. "The relation between stored style structures ... and the if-then constants governing style shapes is one of partial interdependence and interaction: during the perception of music, cognitive feedback and feedforward between style shape and style structure constantly take place" (p. 317).

In this model Narmour employs cognitive structures which have some degree of universality rather than musical specificity. The model makes much use of parametric scaling, but "there is strong reason to believe that the very act of scaling ... is an inborn inherent ability in the physiological mechanisms of the brain" (p. 284). The values of primitive parameters are determined through a perceptual Gestalt. "The invocation of Gestalt laws as bottom-up processors here constitutes a set of rules determining the non closure of input, not prior closure in the form of preordained Gestalts of "good" or "best" continuations" (p. 66). The deferring of closure is essential for stylistic learning to occur. Narmour's use of parametric scales is different from the notion of templating described above.

> *By parametric scales, I mean to hypothesise inborn cognitive matrices that both order the elementary materials of a parameter across a spectrum of similarity and differentiation and also fix them in proximity concerning their inherent perceptual, syntactic, implicative-realisational functions. In other words, the theory of parametric scales hypothesises "slots" that are filled by elementary style shapes. These make up the implications and realisations of a given parameter. These ... scales do not just measure distances between slots or lay out areas of functional contrast between the materials scaled. Instead, they assert the idea of syntactic direction, whereby certain sequences of style shapes create closure and other sequences non closure* (p. 285).

Consequently, Narmour argues for certain constraints on musical form.

> *Music is not a "Markov system" where no sense of discontinuity and accumulation takes place and where each syntactic event is subject only to a single transition probability. Nor is music "Lamarkian", where we understand the current syntactic event only in terms of the acquired structural characteristics of the style. Nor is it "Darwinian", where the if-then constants predict musical or stylistic change only on the basis of some innate internal principle. Music in fact has properties of all these systems* (p. 318).

The informational consequences of such a balancing act are considered in the next section.

3.7 Information redundancy in music perception

Hebb (1961) argues that attention or interest is maintained through a balance of the familiar and the unfamiliar.

> *It is of course evident that some degree of novelty, combined with what is predominantly familiar, is stimulating and exciting over a large range of activities - from sexual responsiveness to the appreciation of painting or music, or the pleasures of exploration -* (p. 231),

and,

> *in music ... the dissonances that are harsh and disruptive at first become pleasant as they become more familiar, but finally are dull and boring. The course of events is: first, a dissonance with too much conflict to elicit a cerebral action; secondly, with the establishment of new assemblies, or modification of existing ones (due to repeated stimulation by the new tone combination), an organised activity is aroused, in which, however, some conflict remains; and finally, organisation reaches a point at which the sensory stimulation no longer offers any check to the phase sequence, and pleasure disappears* (p. 233).

That is, psychoacoustic dissonance creates temporary instability in the cognitive experience of the listener (Bregman, 1994: 682). Such a development is sometimes utilised as a compositional device, e.g., the deliberately jarring chords at the beginning of the overture to Mozart's opera *Don Giovanni*, or the opening bars of Elgar's cello concerto.

Berlyne (1970), McMullen (1974) and McMullen and Arnold (1976) report responses to musical stimuli which support an inverted-U, or Wundt curve, model of pleasure as a function of novelty and complexity of stimulus. Maximum pleasure was gained from simple/novel or complex/ familiar stimuli, with pleasure attenuating as simple stimuli became more familiar, or complex stimuli became more novel. Hargreaves (1986) reviews fourteen other studies which confirm the inverted-U model.

> *People prefer music that provides them with information, that is, which reduces their uncertainty about subsequent events. Extremely unfamiliar music does not reduce uncertainty, since the events within it are totally unpredictable to the listener; and very familiar music does not do so because it contains very little new information. ... people will prefer music that is moderately familiar, that is, ... contains an optimum amount of information for them* (p. 116).

Bever (1988) finds such a theory problematical for accounting for the emotional response elicited from familiar, much less favourite, pieces of music. And by extension, why is the "great" music, e.g., Beethoven's *Symphony No. 5*, so culturally enduring?

> *It cannot be the case for a piece that one has memorised, that the ebb and flow of partially fulfilled expectations control one's enjoyment of it: every note is exactly what is expected. There must be some computational mechanism at issue other than mere statistical predicability* (p. 166).

Such a criticism was addressed by Jackendoff (1991) with his model of musical parsing. He offers a parallel multiple-analysis model in which analytic possibilities are held together until, often at cadential points, sufficient unambiguous information is made available for resolution. Jackendoff argues that the parser must operate at all listenings including familiar and imagined music. The parser treats each music experience as novel, and so generates "surprise" even with remembered pieces. A similar suggestion is made by Schmuckler (1989).

> *Familiarity appears to have local and fairly specific influences on expectancy formation. One way of characterising this effect is that expectancies are formed along the basis of ingrained general stylistic regularities (such as tonal structure, melodic processes, and common harmonic progressions) which operate imperviously to one's experience with a particular piece* (p. 144).

It must be noted that there seems to be some degree of contradiction in a model which posits familiarity as being independent of experience. Schmuckler (1989) found that melody and harmony were "perceptually independent, such that they combined additively in expectancy formation for a full musical context" (p. 109), and that "expectancies wax and wane in strength and specificity, at times relying on perceived tonal structure or knowledge of chord progressions for their formation, while at other times dependent on principles such as melodic processes and contour" (p. 144) - a result which can readily be re-stated in a Lurian framework, whereby executive synthesis determines the relative importance of either simultaneous or successive synthesis to the coding demands of the task at hand.

Much of Schmuckler's conceptualising follows the work of Leonard Meyer. Meyer (1970) regards information theory as a framework within which musical meaning can be investigated.

In music, Meyer argues, meaning is self-referential or "embodied". Listeners extract meaning whenever they become aware of "implications of a stimulus in a particular context" (p. 9). That is, meaning arises when expectancies are not met by "deviations" such as delay, ambiguity and improbability. More formally, "musical meaning arises when an antecedent situation, requiring the estimate of the probable modes of pattern continuation, produces uncertainty about the temporal-tonal nature of the expected consequent" (p. 11). Thus information and meaning are probabilistic functions of uncertainty. "Meaning, then, is not a static, invariant attribute of a stimulus, but an evolving discovery of attributes" (p. 10). Meyer describes three stages for the embodiment of meaning: hypothetical, evident and determinate. Hypothetical meaning refers to expectancy, and provides no information while expectancies are met. Evident meaning is attributed retrospectively on the basis of the confirmation or otherwise of hypothetical meaning. "Determinate meanings are those which arise out of the totality of relationships existing on several hierarchic levels between hypothetical meaning, evident meaning, and the later stages of the musical situation" (p. 14). Information redundancy is critical for this process, because, as required for Jackenoff's parser, it "allows for those important places in the experiencing of music ... where a listener can pause, albeit briefly, to evaluate what has taken place in the past and to organise this experience with reference to the future" (p. 16). Furthermore, as this progression through these stages of meaning unfolds, "probability tends to increase" (p. 15). Consequently, the musical experience is, in contrast to Narmour, a Markov process, i.e., a particular case of a stochastic process in which the probabilities of present events depend upon the previous events. "If music is a Markov process, it would appear that as a musical event ... unfolds and the probability of a particular conclusion increases, uncertainty, information, and meaning will necessarily decrease" (p. 15). Implicit in Meyer's theory is some cognitive process to evaluate changes in information redundancy over the course of a piece of music, i.e., to compare information redundancy in a current segment with that in previous segments over various time scales. It is argued in Chapter 4 that just such an autocorrelational process is necessary to make sense of music.

3.8 Information processing in music experience

The issue on which the literature has some agreement is the effect of musical training or experience on measures of musical ability (e.g., Brown, 1988; Cohen, 1991; Monahan *et al*, 1987; Platt & Racine, 1990; Schmuckler, 1989; Shuter, 1968; Sloboda, 1991; Sloboda *et al*, 1994a, 1994b; Storr, 1992; Tsuzaki, 1991). In children, musical training or experience is not an add-on effect: rather, experience interacts with general cognitive development. There is disagreement on the relative contributions to this interaction, from a position that music ability is the outcome only of cognitive growth (Serafine, 1988), to a position that music ability is the outcome only of environmental influences (Sloboda *et al*, 1994a, 1994b). Most commentators seek an interactionist position. Storr (1992), for example, argues that music is a cultural artefact, not of the natural world, and consequently "originates from the human brain" (p. 51).

Creider (1989) argues that the popular Suzuki teaching method, with its mother-tongue approach, is not simply a matter of environmental immersion. Suzuki training is a discipline for the emerging personality of the child, with the aim of spiritual development through music rather than the development of musical ability *per se*. Consequently, Suzuki training begins with ear playing and memorising to enhance the cognitive development of audiation. In a Lurian context, this would provide musically rich stimuli for the development of simultaneous and executive syntheses during a child's first decade. Serafine (1988) found that Suzuki subjects out -performed others on a cadential closure test. Serafine argues that this result was due to the advanced musical experience of Suzuki subjects who began their music education at an earlier age. She suspects that such differences would even out at later ages. Serafine used temporal and non-temporal tasks to provide evidence for music performance improvement with age. Most of these tasks, e.g., motivic chaining, or the bead task, require simultaneous synthesis, so Serafine's results can be interpreted as measuring maturation in simultaneous synthesis in the domain of music. She concludes that cognitive processes are generic but formal elements such as pitch, chords, meter etc., are not necessarily "descriptive of cognitive reality" (p. 223).

Freeman (1974) found, in a comparative study of musically and artistically talented children matched on intelligence, a significant 'Social' factor. "Talented children ... lived in an educationally nutritive environment which provided incentive, materials and encouragement in the appropriate field of interest" (p. 8). A personality test showed that musical children were more emotionally stable and confident than controls, with parental attitude of supreme importance.

Freeman concludes that sufficient opportunity is necessary for the development of talent, and that exceptionally talented children should be regarded as a part of the social milieu.

McPherson (1993) presents a model for instrumental training involving re-creative performance skills, viz., visual sight reading, aural playing by ear, and improvisation in stylistic and free idioms, based on four factors descriptive of early childhood music experiences: "early exposure", "enriching activities", "length of study", and "quality of study". With young players (school Years 5 - 8), McPherson found significant correlations between Australian Music Examination Board (AMEB) exam success and early exposure ($r = .31$), length of study ($r = .47$), and quality of study ($r = .33$), and between ability to improvise and early exposure ($r = .36$), enriching activities ($r = .43$), and quality of study ($r = .29$). McPherson suggests that improvement in performance skills could occur if, in instrumental music education programs, sounds are introduced well before symbols, as with the Suzuki method. Again, this is consistent with cognitive developmental organisation in the Luria model where successive coding is available from birth, but the use of simultaneous synthesis is developmental. Success at re-creative performance skills requires significant mental rehearsal: 'playing' through auditory images is a task dependent on successive synthesis. McPherson found significant correlations between ability to play by ear and improvising skills ($r = .77$) which could indicate a shared dependence on successive processing. He also found a correlation between sight reading ability and improvising ($r = .75$) which might suggest a role for simultaneous coding of musical chunks. Further correlations between ability to play by ear and AMEB performance ($r = .64$), and ability to sight read and AMEB performance ($r = .75$), indicate the need for both successive and simultaneous processing during instrumental performance.

Perceptual learning, then, contributes to music abilities through the development of information processing capacities. E. J. Gibson (1969) defines perceptual learning as "an increase in the ability to extract information from the environment, as a result of experience and practice with stimulation coming from it" (p. 3). She takes J. J. Gibson's position that information from the natural environment contains the form for its own perception and comprehension. "If information specifying the environment exists, ... and if perceptual systems have evolved so as to detect this information, then the very act of detection in and of itself constitutes some kind of awareness of what is specified" (p. 305). She argues that the increase in conceptualisability with maturation does not affect the perceptual task, i.e., perception does not gradually become inference. "Perception is not a process of matching to a representation in the head, but one of extracting the invariants in stimulus information. ... We do not perceive less because we conceive more. If we did it would be maladaptive for getting information about what is going on in the world around us" (p. 449). In fact, perceptual learning involves a steady improvement in specificity of discrimination to stimuli through a process of exploratory and selective attention towards the "capture" of invariance. Furthermore, the perceived invariant is a function of the perceptual task. "An instrument in an orchestra can be picked out and heard as distinct by the expert, though the task would be difficult if not impossible for the untutored amateur. In this case the melodic pattern over time is not invariant, but a bundle of acoustic properties unique to the instrument is" (p. 81).

Gibson wants to distinguish inference from remembered imagery in perception. She suggests three cognitive processes: discrimination, recognition and production. Mental representations are based on "prior discovery of ... invariant features of a pattern. Discrimination is thus prior to recognition. And ... both ... are prior to production. ... perceptual learning is a requisite for the three processes [which] form a kind of cognitive hierarchy" (p. 152). Such a hierarchy is not inconsistent with the Luria model. However, for music perception, Gibson is somewhat more circumspect. "In the case of the melody, we are a long way from being able to give an account in terms of distinctive features. That a pattern is abstracted is evident from the fact that a melody can be recognised when it is transposed in various ways, and it must be differentiated in some sense for recognition and reproduction to be possible" (p. 151). In sum, the division between perceptual models based on the convergence of "information in the music", and models in which "information [is] added during processing", Gibson's analysis suggests, is dialectical rather than dichotomous. Perceptual learning creates expectancy as the outcome of processing information redundancy in afferent stimuli.

Dowling (1994a) notes that perceptual learning in music is a lifelong process.

> *The remarkable achievements of adult cognition in hearing and understanding music are made possible by a lifetime of perceptual learning built on a foundation present early in life.*

> *The basic dimensions of musical sound ... are present to the infant. Those become increasingly organised through experience and specialised learning in accordance with the musical norms of one's culture. Certain features of the infant's musical world remain for the adult ... Other aspects ... develop through the first twenty years of life. Our different experiences in music lead us to different perceptual learning and different current perceptions of the music we hear. And our continued perceptual learning will lead us to continually hear things differently in the future* (p. 253).

The final issue for this chapter is the extent to which individual differences in aptitude for perceptual learning, or abilities in information processing, can account for individual differences in music abilities.

3.9 Individual differences in music abilities

Many commentators argue for the existence of an aptitudinal dimension to music ability, often measured as perceptual sensitivity to music stimuli (Colwell, 1970; Gordon, 1979, 1993; Radocy & Boyle, 1988; Swanwick, 1988; Trehub, 1994; Walters, 1989). This position is summarised by McAdams (1987): the richness of musical form "depends very heavily on one's capacities and experiences as a listener" (p. 23). The evidence for an independent music potential factor, however, is often equivocal (Carroll, 1994; Shuter, 1968). In earlier studies, correlations between musical aptitude and dimensions of general intelligence were lower than expected (Bentley, 1966; Gordon, 1971; Shuter, 1968; Shuter-Dyson & Gabriel, 1981). More recent investigations have shown stronger relationships between musical aptitude and other cognitive abilities (Freeman, 1974; Hassler, Birbaumer & Feil, 1985, 1987; Karma, 1982, 1985; Parker, 1978; Schmidt & Lewis, 1987; Taylor, 1973). Others prefer to regard musical intelligence as a separate manifestation of human cognition (Feldman, 1993; Gardner, 1991, 1993; Hargreaves, 1986). Some of the difficulties lie in the low reliability and construct validity of the instruments used to measure musical abilities (Brooker, 1993; Sergeant & Thatcher, 1974; Shuter-Dyson & Gabriel, 1981).

Beament (1977) argues that, since the cortex is structured through developmental processes, differences in inherent musicality must be attributed to variance in functioning at the subcortical level. At this level, auditory responses are different for non-musical and musical sounds.

Whereas sensitivity to non-musical sounds is evolutionally advantageous, sensitivity to music confers no obvious advantage and so is evolutionally neutral. Evidence for this position is that rhythm ability is more general than pitch discrimination, and it is rhythm rather than pitch that can be heard in general sounds. It follows that high music ability is not genetic, but serendipitous in its distribution in the population. Beament's prediction is supported by Figgs (1980). Despite well known examples of musical lineages, such as the Bach family, the more common observation is that extremely talented musicians rarely have parents equally talented, and musically talented parents seldom have equally talented children. Figgs showed that partitions of parents and children into high, medium and low musical talent made a 'classical' regression to the mean across the generation; the talent of children with highly talented parents was lower and close to the mean, the converse for children of parents with low musical talent.

Shuter (1968), in a detailed review of previous studies, notes that while musical training does improve scores on musical ability tests, the more outstanding prodigies develop far beyond the limits of ordinary people, even with considerable training. What is more, some subjects with no musical training scored higher on musical ability tests than those with years of formal training. Training may enable the attainment of a ceiling perhaps earlier than otherwise, but the evidence suggests that there are individual differences both in ceiling height, and the rate at which it is attained.

Gordon (1993) defines music aptitude as potential to learn. "Whereas music achievement is primarily in the brain, music aptitude is primarily in the body" (p. 2). Music aptitude, like other individual difference measures, is normally distributed in the population. "Just as there is no person without intelligence, so there is no person without musical aptitude" (p. 2). Davidson (1994) agrees. He draws attention to the considerable musical abilities in contour, pitch and rhythm among the musically untrained. But, Gordon argues, a person is born with a particular level of musical aptitude. The level of that person fluctuates in accordance with informal and formal music environment until age nine. Their potential is then stable for life and will never reach a higher level than at birth. "Although innate capacities cannot be increased regardless of the broadness of early informal environmental influences, innate capacities can be attenuated as a result of lack of broadness of early informal environmental influences" (Gordon, 1979: 43).

Experimental evidence in support of this position is reported by Billingsley and Rotenberg (1982) who found age and musical aptitude correlations with interval use up to age 12. Children with higher musical aptitude also made better use of non-adjacent intervals in discrimination tasks, suggesting also that they can more readily perceive non-serial aspects of melodic structure. In this respect, pitch perception is privileged over other perceptual characteristics (Gordon, 1971, 1979, 1993; Sergeant & Boyle, 1980).

The stabilisation of music aptitude can be associated with the neurophysiological development of the frontal lobes, a part of Luria's third functional block, because "the frontal lobes ... are associated largely with the ability to anticipate coming events. The basis of musical aptitude, as well as intelligence, is how well a person can generalise and make judgements that predict and possibly influence future occurrences" (Gordon, 1993: 3). Taylor (1973), and Radocy and Boyle (1988) concur: overall harmonic awareness is developed as a product of enculturation and reaches a ceiling effect by age six to eight years, certainly by nine. Radocy and Boyle explain individual differences in music abilities through information theory: "a central problem in dealing with complexity is an individual's capacity for such. Capacity for perception and cognition of complex stimuli is limited for everyone, although each individual's limits are not the same" (p. 169). For example, Adachi and Carlsen (1994) found that melodic expectancy in musical children as young as four and a half years matched that of adults. These children, recruited for the study by music teacher nomination, could sing melodic continuations in tune without repetition of the melodic beginning. Many notable composers were musically precocious, e.g., Handel, Haydn, Mozart, Beethoven, Mendelssohn, Prokofiev and Britten, who were playing the piano at age 3 or 4, and composing by age 10 years.

These observations suggest that musical aptitude is multidimensional. Shuter (1968) undertook a factor study of Wing tests with students at the Eastman School of Music in which three factors emerged: 1. memory, chords, pitch; 2. phrasing, intensity; and, 3. harmony, rhythm. A similar meta-analysis by Whellams (1973) found against any unitary factor of 'musicality'. Shuter-Dyson and Gabriel (1981) note several early factorial studies which showed that all such tests loaded on to one factor, generally interpreted as 'discrimination among sound patterns'. However, a re-analysis of these studies revealed other factors, including: 1. 'auditory memory' on to which loaded verbal and number span tests; 2. 'auditory cognition of relationships', which, the authors suggest, "may well require a more advanced form of processing"; and, 3. 'temporal tracking' which required a reverse-ordering of tone sequences. The factors 'auditory memory' and 'temporal tracking' seem to bear a close relationship to the Luria model factor of successive synthesis, whereas the 'auditory cognition of relationships' factor is consistent with simultaneous synthesis. For Gordon (1979), music aptitude is multidimensional with three factors: melody and harmony; tempo and meter; and, phrasing, balance and style. These factors could be interpreted as musical equivalents of factors for simultaneous, successive and executive syntheses respectively. Carroll's (1994) comprehensive meta-analysis of factor studies reveals four factor types: first, factors which discriminate two or more tonal qualities such as pitch, timbre, duration, rhythm or intensity; second, factors which discriminate on frequency attributes such as pitch and timbre; third, factors which discriminate on temporal and sequential patterns, e.g., duration or intensity; and, fourth, factors which discriminate on musical taste. Again, Carroll's analysis can be understood within the framework of the Luria model. Most factors were in Group 1, which seems to represent a general factor of musicality. Group 2 is of factors representing simultaneous synthesis of pitch relationships. Group 3 has factors of tasks requiring successive synthesis, while Group 4 factors reflect the use of executive processing.

Consistent with a multidimensional conceptualisation of music aptitude, Gordon (1979, 1993), Walters (1989) and McPherson (1993) all conclude that the fundamental difference between individuals lies in their ability to audiate. Walters (1989) defines audiation as "the hearing of sounds that are not before the ear at the moment, through recall, prediction, or conception" (p. 5). Audiation involves comprehension, not just imitation. For example, a musician with notational audiation can "hear with his eyes and see with his ears. ... he can see notation and audiate the sound it represents, and he can audiate sound and visualise the notation needed to represent it" (p. 9). Gordon denotes eight types of audiation, all of which would require high ability on executive synthesis: 1. listening; 2. reading; 3. writing from dictation; 4. recalling or performing from memory; 5. recalling and writing from memory; 6. creating and improvising; 7. creating and improvising while reading; and, 8. creating and improvising while writing.

Each type of audiation goes through six hierarchical stages which are so tightly chained that they overlap:

1. momentary retention;
2. initiating and audiating patterns, and recognising tonal centre and macrobeats;
3. establishing objective or subjective tonality and meter;
4. consciously retaining organised patterns;
5. consciously recalling patterns audiated in other pieces of music; and,
6. conscious prediction of patterns.

Stages 1 and 2 are activities which are dependent upon successive synthesis; Stages 3,4 and 5 need simultaneous coding; and Stage 6 is a responsibility of executive synthesis. Gordon notes the cyclic relations between Stages 1-4 as they interact with each other. Walters suggests that the process of audiation is different for each individual depending on aptitudes, backgrounds, personalities and retention skills.

Relationships between such individual differences in musical ability and general intelligence have received considerable attention. Feldman (1993), in a study of child prodigies, notes that precociousness is usually domain specific, idiot-savants being an extreme example. Shuter (1968) reviewed research into correlations between musical ability and intelligence, and musical ability and academic performance. Musical ability correlates best with intelligence at the low end of the scale. On the other hand, Cattell's factor for speed correlates with Seashore's memory and pitch tests. Shuter suggests that this indicates a basic intellectual skill in transcending time, "temporal integration of the stream of experience" (p. 228), i.e., successive synthesis. Shuter concludes that "musical ability is largely specific. ... Positive correlations are nearly always found between measures of musical ability and other cognitive aptitudes. However ... the coefficients are low" (p. 236). Shuter-Dyson and Gabriel (1981) report correlations from .4 to .6 between measures of musical discrimination and general intelligence scores. On achievement tests, however, music students show a high correlation with foreign language acquisition. Hargreaves (1986) notes the symbolic nature of musical representation. Music is a field in which "highly abstract, impersonal rule systems [are] involved" (p. 173). He argues that this can explain the phenomenon of child prodigies occurring in music and mathematics in contrast to other academic areas in which breakthroughs or exceptional performance are much more dependent on accumulated experience.

This suggests that some positive correlation might exist between musical and mathematical abilities. Such a relationship is frequently cited in common parlance. The argument in this thesis, that musical information processing can be explained by the Luria model dimensions of simultaneous and successive synthesis, could be extended to predict a strong relationship between music and mathematical abilities, given that significant variance in school mathematics performance can be accounted for by abilities on simultaneous and successive synthesis (e.g., Crawford, 1986; Tulloch, 1981; Walton, 1983). On the other hand, Das, Naglieri and Kirby (1994) counsel for caution when seeking relationships across domains of application, such as music and mathematics. Gordon (1971) presents evidence against such a link. Multiple regression analyses showed that intelligence test scores did not significantly improve predictions of musical success of Fifth grade children over that predicted by Gordon's Music Aptitude Profile (MAP). Correlation between MAP scores and academic achievement was also low, but lower for sub-tests of mathematics than for sub-tests of language arts. Shuter-Dyson and Gabriel (1981) note that the differential importance given at school to these subjects, and the considerable time investment required for high performance in music, could preclude development of mathematical achievement even if mathematical aptitude is present.

Mathematical aptitude is commonly measured by visual-spatial abilities. Evidence for relationships between spatial skills and musical aptitude found by Karma (1979) was reported in *Section 3.3* . In further studies using tests for auditive structuring, Karma (1982) found that such relationships were sensitive to gender and subject maturity. With 10 year olds, musical aptitude was correlated with verbal ability, in contrast to college students whose musical aptitude was correlated with spatial abilities. As argued above, such a result demonstrates a change in strategies of information processing music with cognitive development. Ten year olds predominantly use successive synthesis in their cognitive strategies, whereas older students also have simultaneous synthesis available for similar tasks. Hassler, Birbaumer and Feil (1985, 1987) conducted a longitudinal study of musical and visual-spatial abilities. They report two independent factors for visual-spatial ability: 1. retention of configuration under translation or rotation; and, 2. retention of configuration against distractors. Similar tests are used as marker tests for simultaneous synthesis in the Fitzgerald battery used in this research. The authors found a relationship between musical ability and spatial visualisation, which, like Karma, was confounded by effects for gender and maturity.

A contrasting approach is taken by Gardner (1993) in his theory of Multiple Intelligences (MI). His "qualitative factor analysis" of a survey of the evidence produced seven separate intelligences, of which mathematical-logico and music are two. Here, intelligence is regarded as potential. "Individuals may differ in the particular intelligence profiles with which they are born, and ... certainly they differ in the profiles they end up with" (Gardner, 1993: 9). In MI theory, intelligences are used in concert for success in the world, e.g., a professional musician would use musical, bodily-kinaesthetic, inter-personal and intra-personal intelligences. The type of information required is the trigger for the type of intelligence to be employed. "Each intelligence must have an identifiable core operation or set of operations. As a neurally based computational system, each intelligence is activated or "triggered" by certain kinds of internally or externally presented information" (p. 16). Gardner argues that these intelligences are independent. "A particularly high level of ability in one intelligence, say, mathematics, does not require a similarly high level in another intelligence, like language or music" (p. 26). This is not necessarily in conflict with the Luria model. Gardner's seven intelligences could be underpinned by the three information processing dimensions of the Luria model, if, as Das *et al* (1994) emphasise, measures of information processing abilities must be located within specific domains. In any case, Gardner states that "there is certainly no reason why information-processing accounts could not be given for each of these intelligences and their manner of interaction" (p. 41).

A seminal re-examination of previous data on the relationships between musicality and general intelligence was undertaken by Sergeant and Thatcher (1974). The authors noted the uniformly low level of association observed for four decades, in contrast to the common observation of music teachers that high musical achievers have high levels of general intelligence. In fact, some research they reviewed suggested that tests of non-musical intelligence were better predictors of success in music than were some music tests. Sergeant and Thatcher draw attention to the relationship in statistics, apparently overlooked by earlier researchers, between correlation indices, r, and the reliability of the measures being correlated, viz., that r is constrained to be less than the reliability. Since the reliability of music aptitude tests is often "disconcertingly low", typically around 0.4, r cannot rise above about 0.4. Taking this into account, Sergeant and Thatcher show that corrected correlations between musical aptitude and intelligence are typically high, with r exceeding 0.8. That is, some 60 - 70% of the variance (r^2) in measures of musical aptitude can be accounted for by measures of non-musical intelligence. The authors then point out that correlation is also limited by validity, invalid tests for intelligence or musical aptitude adding further to the error variance. Sergeant and Thatcher recommend factor analysis, and cite factor studies which show a definite relationship between musical achievement and intelligence. With regard to the ability underlying musical achievement, the authors note that this is a one-way association: high musical ability implies high intelligence but not vice versa. Musical ability, then, develops from the interaction of intelligence with other factors, especially a musical environment in the home. But, "a favourable musical environment cannot redeem the absence of the level of intelligence necessary for musical cognition, nor can intelligence alone suffice for the development of musicality" (p. 56).

3.10 Definitions of terminology

The literature is not entirely consistent in its use of terminology related to musical abilities. For example, the terms "music ability", "musical talent", and "musicality" are sometimes used interchangeably, and on occasions where perhaps the author is referring to musical aptitude. In this thesis, so far as the literature will allow, an attempt at consistency is made which is based on descriptions by Boyle (1992) of a range of constructs employed in the literature for the evaluation of musical abilities. The relationships between these constructs as used in this thesis are presented in Figure 3.01.

music ability: what a person is able to do musically; a generic term with no implications for the weightings of factors contributing to variance such as innate potential, environment, enculturation or formal instruction. This is consistent with Carroll (1994) that "ability refers to variations in individuals' potentials for present performance on a defined class of tasks" (p. 16), and with Das, Naglieri and Kirby (1994) where "an ability is a trait or characteristic of a person, with respect to some mental task, that has attained a stable level of performance" (p. 8).

music aptitude: potential for learning music, particularly for developing music skills. This term is more related to potential, and may contribute to musical ability. It is what is purported to be measured by music aptitude tests such as Gordon's Music Aptitude Profile (MAP). Gordon (1993) sees music aptitude as a developmental variable depending on musical environment which stabilises by age nine. Carroll describes aptitude as being stable over time and learning situations, and does not include any interest in, or motivation for, a particular activity.

music intelligence: a combination of music ability plus music appreciation as purported to be measured by the earlier tests of Seashore or Wing (1970). Bridges defines musical intelligence as "the ability to generalise music learning in the context of a wide range of unfamiliar music" (in Brooker, 1993: 75).

music capacity: that part of music ability ascribed to genetic endowment. Boyle argues that this is not a useful construct since the separation of innate and environmental influences is impossible to measure. Das, Naglieri and Kirby however, suggest that "abilities are normally thought of as capacities, which can be measured in a 'how much' sense" (p. 8). They argue that performance is more often a function of strategies for mental processing rather than limitations of capacity or ability.

music talent: the capacity for musical performance, and implies biological potential. Again Boyle suggests that this construct is not useful for evaluation purposes. Nonetheless, it is a widely used term. In Gagne's actualisation model of giftedness and talent (Gross, 1993), talent is the outcome of the interaction of giftedness or potential with environmental and personality factors.

music sensitivity: perception and responsiveness to subtle differences in music, reflecting both auditory discriminations and affective responses. This is a generally recognised construct but tests for music sensitivity, suggests Boyle, are not well received as isolated measures because of the lack of correlation with other tests of music aptitude.

musicality: implies both cognitive and affective dimensions, i.e., a strong appreciation for music which includes knowledge of formal elements of musical organisation.

music achievement: musical accomplishments as a result of learning from experience with music and music-related phenomena. This may result from formal instruction or from informal less structured settings. Carroll suggests that it is the combination of achievement and aptitude which provides the best prediction of future performance.

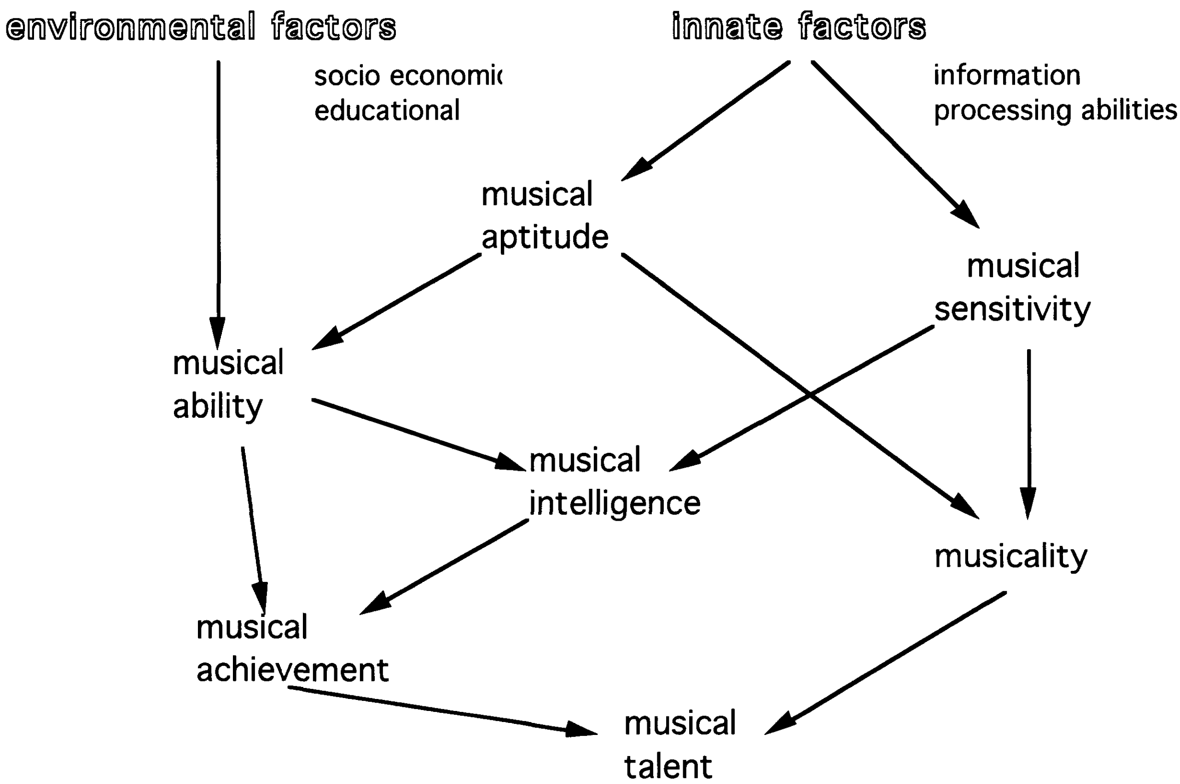

Figure 3.01 Relationships of terminology concerned with music ability

Going down Figure 3.01 the constructs get broader in scope and subsume the constructs above, as indicated by the arrows. For example, musical intelligence requires both musical ability and musical sensitivity, musical intelligence contributes to musical achievement, and musical talent is a product of the interaction of musical achievement and musicality. This analysis suggests that the development of musical talent from the innate factors of musical aptitude and musical sensitivity is multi-staged and arises through interaction with an individual's early musical environment. The map in Figure 3.01 could be used to explain the variety of musical behaviours observed in adults, such as the devoted listener (high musical sensitivity), the professional music critic (high musical intelligence), or the enthusiastic amateur performer (modest musical ability arising from combinations of factors such as a high aptitude but a lack of music education opportunity when young). It should be noted that, following Boyle, the problematic concept of music capacity is not referred to in this study and thus does not appear in Figure 3.01.

Chapter summary

The literature supporting hierarchical models of cognitive processing of musical stimuli can be interpreted within the successive/simultaneous/executive organisation of the Luria model. Music, as a serial presentation of information, is initially coded by successive synthesis into temporally organised units of rhythm and duration, and by simultaneous synthesis into template schemas of pitch, timbre etc. used in long term musical memory. Executive synthesis is employed to focus attention during perception, to resolve cognitive conflicts between the two coding processes, and to create expectancies as a basis for enjoyment, from the evaluation of informational redundancy. Superior musical ability is characterised by high levels of audiation, which are underpinned by an increased use of simultaneous synthesis to employ long-term memory of intervals and scales in perceptual organisation. Musical aptitude is an identifiable trait which interacts positively with early music experiences and musical training to establish music ability. Musical aptitude is underscored by general intelligence, i.e., abilities on successive, simultaneous, and especially executive syntheses.

CHAPTER 4

PERCEPTION OF AUTOCORRELATION STRUCTURE

Chapter Overview

The preceding chapter considered the processing of musical information at a cognitive level. The focus of this chapter is on the question: In what form is information extracted from a musical signal so that the cognitive processes of simultaneous and successive synthesis can generate structural phenomena such as pitch, timbre and resolution to the tonic? It is argued here that the autocorrelation structure of a musical signal is a suitable candidate for encoding such information. There is evidence that autocorrelation structure is a psychologically penetrable signal characteristic. It is argued that autocorrelation structure is a plausible mechanism for the chunking of sequential information, such as a tone series, into simultaneous constructs, such as tonality, as consistent with the Luria model. Recent literature has given some attention to an earlier study by Voss (1975) into the spectral density function of music as a representation of its autocorrelation structure. This approach regards the fluctuations in a musical signal as an example of fractional Brownian motion, as described by Mandelbrot (1983) in his celebrated analysis of fractal geometry.

This chapter has four sections. It begins with a brief overview of fractal geometry. *Section 4.2* describes the properties of fractional Brownian motion, and presents a critical analysis of the study by Voss (1975) into the spectral density characteristics of music. *Section 4.3* presents evidence for and against the auditory perception of fractal sequences by human observers. In *Section 4.4* a model is proposed for the role of autocorrelation structure in music information processing.

4.1 Overview of fractal geometry

Fractal geometry provides a description of the underlying mathematical order of complex dynamical phenomena. Phenomenologically, fractal geometry deals with the rough and irregular forms of the natural world. Fractal form is manifest in spatial and temporal domains, the latter having particular salience for music.

Mandelbrot describes a fractal as "a shape made of parts similar to the whole in some way" (Feder, 1988: 11) (see also Mandelbrot, 1983, 1990). The scaling formula

$$N = s^D \qquad (4.1)$$

where N is the number of the original shapes of dimension D now scaled up to a larger size by a factor s, can be used to derive an expression for a similarity dimension:

$$D = \log N/\log s. \qquad (4.2)$$

As there is no restriction on either s or N being integers there is no such restriction on D (Barcellos, 1984). This can be demonstrated with the task of measuring the length of a convoluted boundary such as a coastline. Regardless of the scale employed, i.e., measuring instrument (e.g., Landsat satellite, light aircraft, theodolite), the measured length will always be less than the actual length because there will always be line segments which are smaller than the measuring scale. The coastline of Australia, for example, has a dimensionality approaching 1.29 (Feder, 1988). Different coastlines and other complex boundaries such as river banks or the edges of oil slicks in the ocean have different fractional dimensions and thereby display variability of fractal form. Fractional dimensionality can be interpreted as meaning that lines with fractal form "begin to occupy area" (Mandelbrot, 1983). More formally, in a fractal set the Hausdorff-Besicovitch dimension exceeds the topological dimension (Bunde & Havlin, 1994). For profiles found in natural scenery, $1.2 < D < 1.3$, while for planar images of natural scenery involving fractal surfaces such as trees, bark, rocks and ocean waves, $D \approx 2.5$ (Field, 1987; Pentland, 1984).

Several characteristics of fractal form are salient. First, fractal forms are generated through iteration.

> *Fractals ... are the end result of physical processes that modify shape through local action. Such processes will, after innumerable repetitions, typically produce a fractal surface shape* (Pentland, 1984:662).

Such physical processes include erosion, aggregation, turbulent flow, and morphogenesis (Bunde & Havlin, 1994; Nonnenmacher, 1993).

Second, fractals, in contrast to lines and surfaces in Euclidean geometry, are continuous everywhere but nowhere differentiable (Mandelbrot, 1983; Devaney & Keen, 1989). This has implications for the perception of fractal form given that perceptual processes, while sensitive to differences, are necessarily coarse-grained. Also, perceptual processes do not necessarily rely on differentiation (Gregson, 1994).

Third, natural fractals are not strictly scale invariant, but exhibit statistical self similarity which is conserved only under affine transformations such as reflections, rotations, re-scalings and translations. Such qualitative similarity is labelled as "random non-standard scaling symmetry" (Pickover & Khorasani, 1985). There are physical restrictions on the range of scaling over which natural objects have fractal properties. The lower limit is determined by the size of the constituent particles; the upper limit is determined by the size of the object (Nonnenmacher, 1993; Pentland, 1984). Multifractals are objects which have different fractal properties over adjacent ranges of scaling. It is argued below that auditory sequences, particularly in music, are more likely to be multifractal rather than strictly self-similar.

Fourth, fractals are phenomenologically complex. This has been well exhibited with non-natural fractals whose plots exhibit random non-standard scaling symmetry. The most celebrated is the Mandelbrot Set, {M}, which is the fractal boundary of the basin of attraction in the complex plane of the iterative mapping

$$z_{n+1} = z_n^2 + c \tag{4.3}$$

where c is a constant. The mapping is conducted for each point c in a region of the plane about the origin. The basin of attraction is the set of c for convergent mappings, typically realised as a computer graphic (Mandelbrot, 1983; Devaney & Keen, 1989). Colour coding the rates of divergence for c in regions outside of {M} creates computer graphic images of "infinite complexity" (e.g, Peitgen & Richter, 1986; Peitgen & Saupe, 1988) which are particularly evocative of natural form, such as the branching of trees, rivers, blood vessels, bronchioli and lightning, and shapes with fuzzy edges such as flames, clouds and mountain-scapes (Gardner, 1978; Gleick, 1987; Mandelbrot, 1983; Peitgen & Richter, 1986). Random self-similarity is apparent from images of successive magnifications (Peitgen & Saupe, 1988).

Fifth, the self similarity of fractals permits large scale compression of information. Barnsley's collage theorem states that a fractal image can always be created from a collage of its affine transforms (Barnsley, 1988; Barnsley & Sloan, 1988; Devaney & Keen, 1989). It is something like remembering the image of a tree having viewed it from many angles. This result has important implications for information compression. An image can be stored in the form of a matrix of its affine transforms (typically 6 x 6), weighted by an estimate of each transform's importance in the collage (an additional column of relative probabilities). That is, an image can be stored by about 40 data compared to the 60000 data required for a pixel by pixel rendering across a full computer screen. Recreating the image as a computer graphic is then simply a matter of employing the transform matrix to plot points. This process, labelled Iterated Function System (IFS) (Barnsley, 1988; Barnsley & Sloan, 1988), can be readily applied to imagery of all types, fractal, Euclidean, or a combination. Similar images are readily created from the IFS matrices of nearest-neighbour approximations. As the storage, recall and creation of mental images must be undertaken by cortical cell assemblies which, though large, are limited in size, Stuki and Pollack (1992) suggest that mental images may be stored and reconstructed through an IFS-type process. The authors report the successful demonstration of a neural net model for reconstructed memory which produced a wide range of imagery having been trained on only a limited set of IFS codes.

Sixth, fractal form is one of the underpinning constructs in the study of dynamical non-linear systems, or chaos theory. The evolution of a non-linear system can be plotted as a trajectory in phase space (Haken, 1983; Schroeder, 1991). The evolution of a non-linear system where the parameters are dynamically stationary forms a "strange" attractor, "a solution shared by multiple trajectories originating from different initial conditions" (Mainzer, 1994: 155). Sensitive dependence on initial conditions, the much heralded "butterfly effect" first described by Lorentz, is thus an unavoidable property of such non-linear dynamical systems (Gleick, 1987; Haken, 1983, 1987; Mainzer, 1994). Shaw (1983a) uses information theory to explain the amplifying effect of iteration and, consequently, how sensitive dependence means that phenomena at a microscopic level can affect a dynamical system at the macroscopic level. It is this deterministic unpredictability which characterises such systems. This is reflected in the fractal dimensionality of the graphic representations of strange attractors (Devaney & Keen, 1989; Schroeder, 1991).

Several parameters, in addition to the fractal dimension D, are available for the quantitative characterisation of strange attractors. The Lyapunov exponent,

$$\lambda = \log (t_{n+1} / t_n) \text{ as } n \to \text{infinity} \qquad (4.4)$$

gives a measure of the divergence of the evolution of an attractor. Trajectories which are close at t_0 will, at some later time t_n, have diverged by a function of e^{λ} along each embedding dimension E. Entropy as information also increases during attractor evolution (Shaw, 1983a).

Entropy measures such as the Bolztmann entropy (log of the state probability distribution) (Wannier, 1966) or the Kolmogorov entropy (sum of the positive Lyapunov exponents) (Gregson & Harvey, 1992) can be used to describe chaotic behaviour. Researchers now recognise a considerable variety of mappings which, within an appropriate parameter range, exhibit chaotic behaviour. The more widely cited include the Lorentz equations from meteorology, the first to be recognised as chaotic, the logistic-difference equation from population dynamics, and the Henon mapping from astronomy (Devaney & Keen, 1989; Schroeder, 1991).

Of interest are those systems which sustain a level of complexity far beyond that of their kernel relationship. Such systems are said to be self-organising. Examples cited extend from insect colonies of ants and bees to human social systems such as the economy (Waldrop, 1992; Mainzer, 1994). The model of Freeman (1990), reviewed in Chapter 2, of how local neural phenomena can create global cognitive activities such as recognition and consciousness, is a pertinent example of self-organised complexity. In contrast, many other biological and psychological non-linear systems are not dynamically stationary, and thus do not always exhibit this sensitive dependence (Gregson, 1994; Ruelle, 1990). A possible example of non-stationarity in music is observed with choral singers who, if not equipped with perfect pitch, can often experience difficulty in pitching notes, even if sung a bar before, in ambiguous harmonic contexts such as the bitonality popular with contemporary composers.

These features of fractal form: iterative generation, non-differentiability, self-affinity, complexity, high informational density and non-linearity, also characterise fractional Brownian motion as described in the following section.

4.2 Temporal fractal form: the study of Voss

The fluctuation of stock exchange cotton prices was one of the first phenomena which Mandelbrot demonstrated to be self similar (Mandelbrot, 1983). The apparently random pattern of highs and lows is repeated over the different time scales of days, weeks, months and years. Voss (1985) suggests that "changes in time ... have many of the same similarities at different scales as changes in space" (p. 811). Such time-dependent fractal phenomena include fluctuations in the height of ocean waves and flooding rivers (Feder, 1988), and fluctuations in the voltages of electronic components, the frequencies of impulses along axons, and even in the flow rates of traffic on freeways (Gardner, 1978; Voss, 1985).

Such time series, unlike fractals in space, are not continuous. Their stochastic behaviour is similar to Brownian motion where displacement, rather than position, is independent of the time interval. The behaviour of a signal V(t) whose fluctuations are time dependent can be described by two related functions:

1. the spectral density (or power spectrum) $S_V(f)$ which is a function of the mean squared variation $\langle V^2 \rangle$ centred on frequency f; and,
2. the autocorrelation function $R_{VV}(T) = \langle V(t)V(t + T) \rangle$ for a stationary signal where T is the period or lag time over which the correlations are measured.

Spectral density, a function of frequency, and autocorrelation, a function of time, are related by Fourier Transform (FT) by the Weiner-Khinchin theorem (Wannier, 1966):

$$S_V(f) = FT\,[R_{VV}(T)]\,. \tag{4.5}$$

Signals with a single correlation time T_c are those labelled as "fractional Brownian motion" (fBm) (Mandelbrot, 1983; Voss, 1988). For fBm the autocorrelation function,

$$R_{VV}(T) = \langle V^2 \rangle \exp(-|T|/T_c)\,. \tag{4.6}$$

Three cases are of interest. The first is where $T_c >>$ all lag times T; here V(t + T) is independent of V(t), and $S_V(f)$ is independent of f. This is known as white noise, and is experienced as the hiss of radio static from an untuned receiver. The second case is where $T_c << T$; here V(t + T) is highly dependent on V(t), and $S_V(f) = 1/f^2$, labelled brown noise. FBm series whose characteristics are intermediate between these two extremes are known severally as pink noise, fractal noise or 1/f noise (Gardner, 1978; Schroeder, 1991; Voss, 1985). 1/f noise has equal power in octaves or any frequency bands on a logarithmic scale (Schroeder, 1991). V(t + T) is somewhat dependent on V(t) over a large range of T, and $S_V(f) \approx 1/f$. The degrees of correlation of V(t) over $T \approx 1/2\pi f$ are seen in plots of log $[S_V(f)]$ vs log[f] [Figure 4.01]. The more negative the slope the higher the correlation.

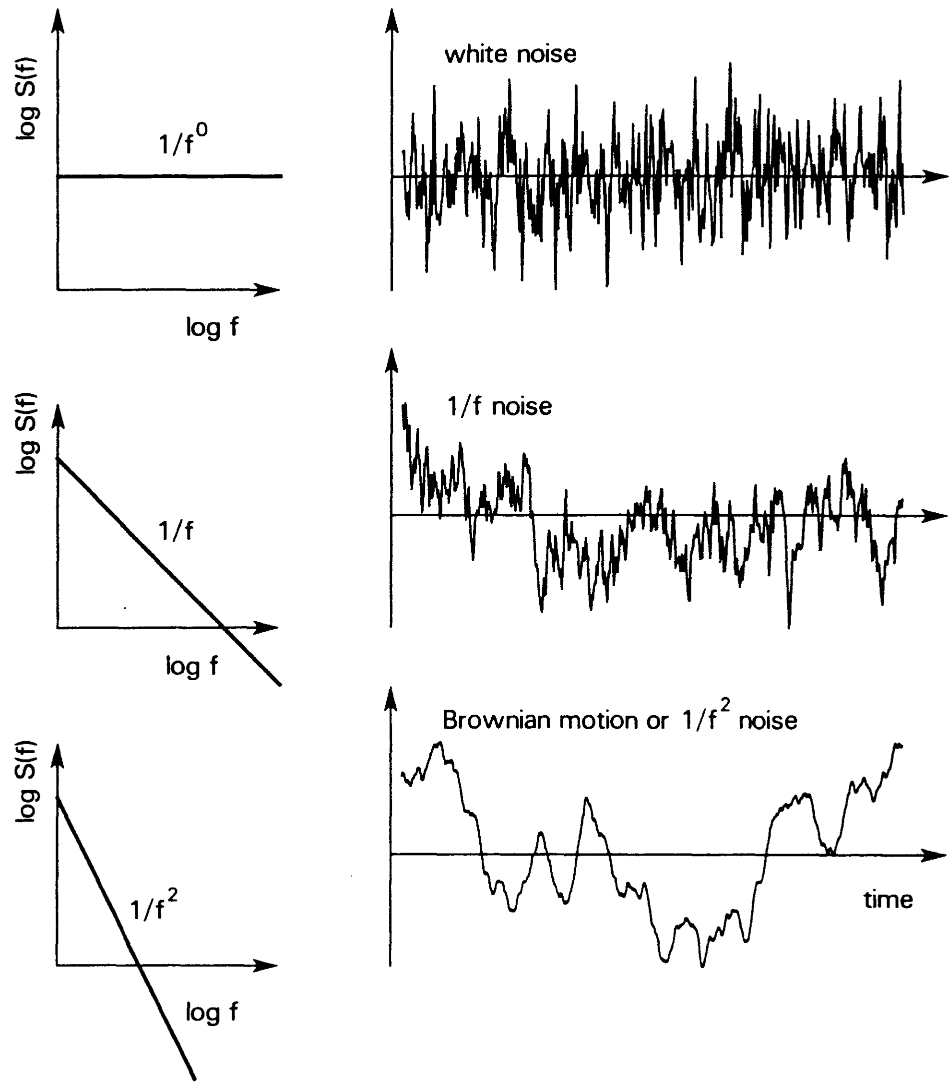

Figure 4.01 Graphical representations of white noise, 1/f noise and brown noise (Voss, 1988: 40).

To the right, graphs of V(t) vs t.

To the left, corresponding graphs of log [S_V(f)] vs log[f].

The degree of self-affinity of an fBm series is given by the Hurst exponent

$$H = \log a / \log b \tag{4.7}$$

where t -> at and V(t) -> bV(t). Graphs of V(t) vs t for H = 0.8, 0.5 and 0.2 are shown in Figure 4.02. It is clear that V(t) becomes more jagged as H decreases; H is also known as the "roughness exponent". For multifractals there is the analogous concept of multiaffinity (Kertesz & Vicsek, 1994). Voss (1988) has employed Equation 4.7 to create realistic computer graphics of landscapes. For example, in creating a mountainscape, if length scales to a factor r, then height must scale to r^H. The absence of such rescaling in satellite imagery renders mountains as seeming to be too flat. Voss suggests H = 0.8 as an optimally useful value for the purpose of landscape 'forgeries'.

The characterisations of an fBm series, fractal dimension D, Hurst exponent H, and exponent of the spectral power curve (negative) β, are related:

$$D = E + 1 - H = E + (\beta - 3)/2 \qquad (4.8)$$

from which

$$H = (2 - \beta)/2 . \qquad (4.9)$$

At H = 1/2, V(t) -> $r^{\sqrt{2}}V(t)$, and the series is a Markov process with β = 1. For H > 1/2, 0 < β < 1, and the correlation over longer time periods increases faster than at H = 1/2; for H < 1/2, 1 < β < 2, the correlation over longer time periods increases more slowly. Feder (1988) describes this as "persistence" and "anti-persistence". Applied to a series of tones, if H = 1/2 then successive intervals are independent of each other. If H > 1/2 then increases (or decreases) in interval size at t_0 would be followed by further increases (or decreases) at a later t_n. Conversely if H < 1/2, increases at t_0 would be followed by decreases at a later t_n. This may have application to a melody line which would be expected to remain within the compass of the performing instrument or vocal part, and conclude in the tonic harmony.

$S_v(f)$ can be measured experimentally from the squared filtered output of the signal characteristic of interest V(t). Voss (1975) (see also Voss & Clarke, 1975, 1978) applied this to musical signals from recordings and radio station broadcasts. Two signal characteristics were analysed: audio power $V^2(t)$, as a measure of total musical output, and the frequency zero crossing rate Z(t), as a measure of perceived melody. In general, for both characteristics, the spectral density was 1/f-like [Figure 4.03]. More specifically, a spectral analysis of Bach's preludes gave a more stable value for β = 1 across the frequency range used than for works of contemporary composers such as Babbit, Carter, Davidovsky and Stockhausen. Subjecting the broadcasts of various FM radio stations to similar spectrographic analysis Voss found that all outputs

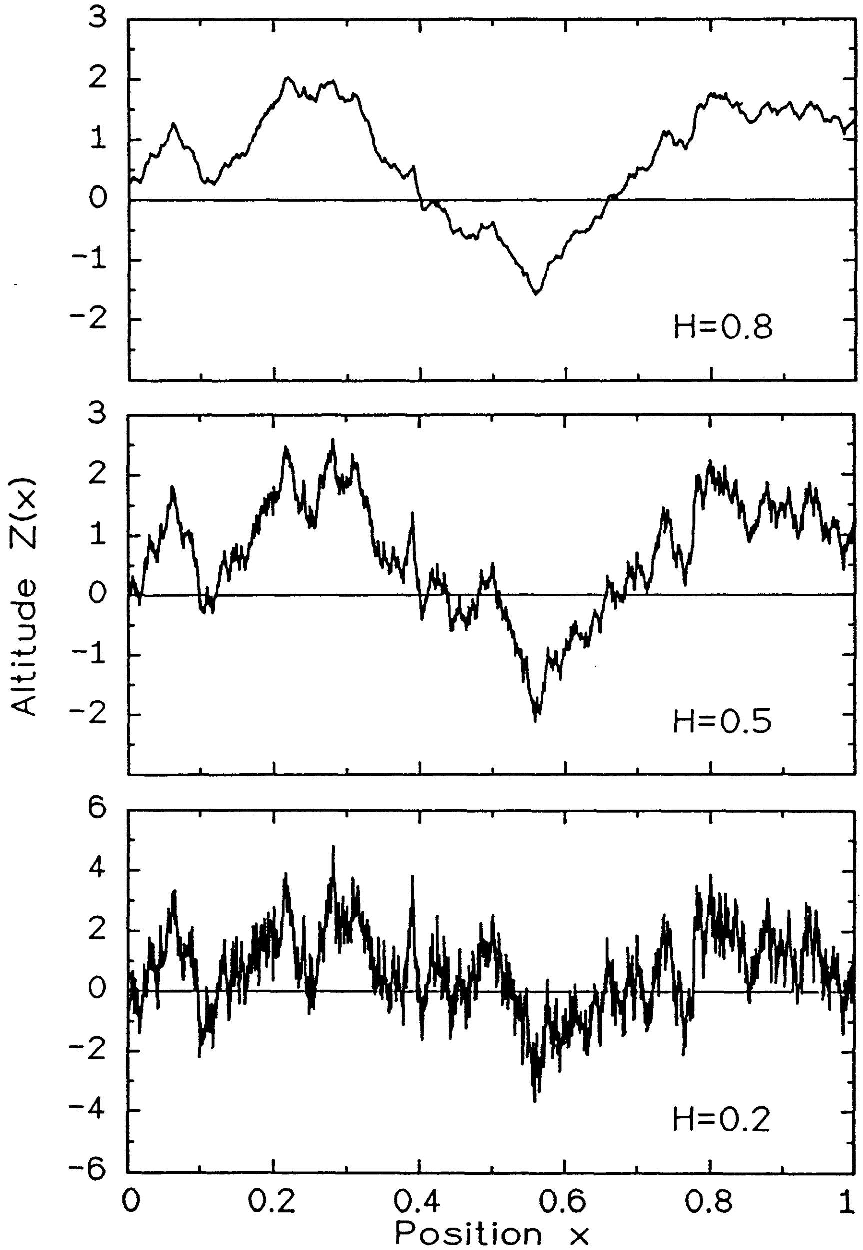

Figure 4.02 Three fBm series with different degrees of correlation

(Bunde & Havlin, 1994: 22)

Hurst exponent: H = 0.2 (low correlation);

H = 0.5 (intermediate correlation); and,

H = 0.8 (high correlation).

were "1/f-like", but that the frequency fluctuations of the classical station were closer to 1/f noise across all frequencies than those of jazz and rock music stations. The frequency fluctuations of a news and talk station showed $1/f^2$ -like humps at low and high frequencies. Voss suggests that the differences between music genres could be due to the higher level of spoken interruptions in radio presentations of jazz and rock music. A more recent and extended analysis has shown that the pitch fluctuations of music from many different cultures and composers displays 1/f-like behaviour (Voss, 1988) [Figure 4.04].

Gardner (1978) suggests that music sits between sounds which are too predictable, e.g., a drone based on one interval, say, a fifth, and sounds which are too unpredictable, e.g., a young child hitting random notes on a piano. He reports Voss as speculating that this may explain why the experience of music, an essentially non-imitative art, fits so well with the experience of perceiving natural imagery. "Indeed, the sophistication of this 1/f music ... extends far beyond what one might expect from such a simple algorithm, suggesting that 1/f noise ... may have an essential role in the creative process" (Voss & Clarke, 1975:318). These authors further speculate that:

> *measures of "intelligent" behaviour should show a 1/f-like spectral density. Whereas a quantity with a white spectral density is uncorrelated with its past, and a quantity with a* $1/f^2$ *spectral density depends very strongly on its past, a quantity with a 1/f spectral density has an intermediate behaviour, with some correlation over all times, yet not depending too strongly on its past. Human communication is one example where correlations extend over various time scales. In music much of the communication is conveyed by frequency changes that exhibit a 1/f spectral density* (Voss & Clarke, 1978:261).

Voss tested this speculation by investigating the appeal of fBm note sequences generated from sources of white, brown and 1/f noise [Figure 4.05]. For these stochastic compositions, Voss used physical sources such as the voltage fluctuations across various electronic components. Subjects preferred the 1/f noise sequences over the others.

Voss' speculation has been applied to music composition. Nonlinear mappings including {M} have been used to generate synthesised music (Dodge & Bahn, 1986; Pressing, 1989) and other auditory sequences (Gregson & Harvey, 1992). A similar analysis has been made by Pickover

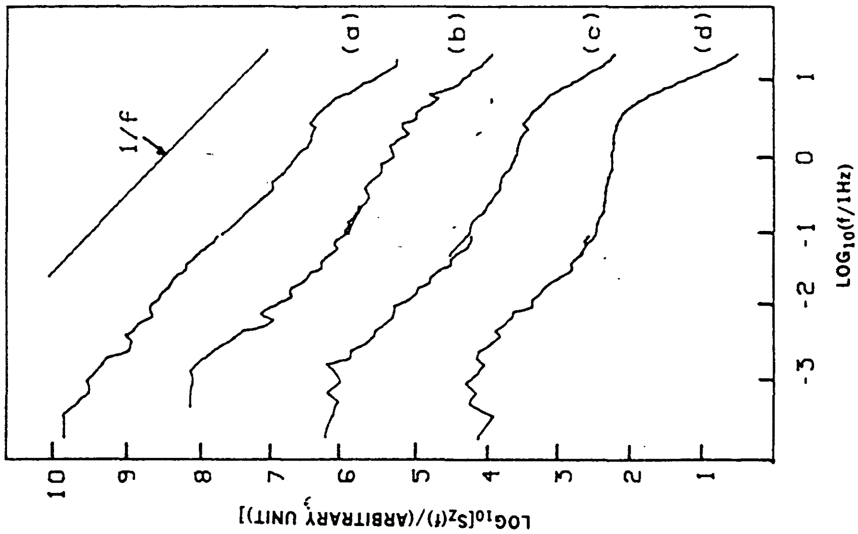

Figure 4.03 Power spectra for pitch fluctuations of four FM radio stations

(Voss & Clarke, 1975: 318)

Radio stations: a - classical, b - jazz, c - rock, d - news and talk

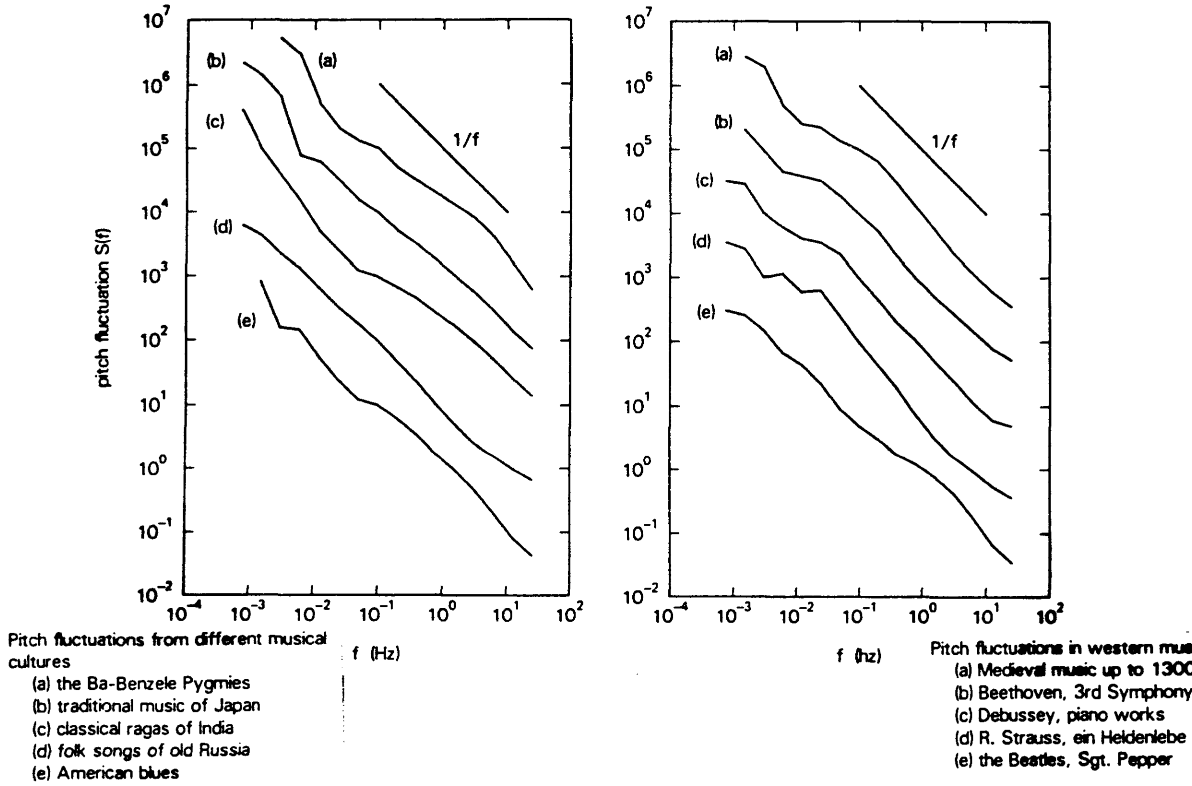

Figure 4.04 Power spectra for pitch fluctuations of various samples of music

(Voss, 1988: 41)

to the left: music from non-Western cultures   to the right: music of various composers

White music

1/f music

Brown music

Figure 4.05 Music scores for examples of white, 1/f and brown music (Gardner, 1978: 24-28) and Khorasani (1986) of the frequency spectra of human speech patterns. They report $D \approx 1.66$. Pickover (1986) also employed the power spectra function to analyse popular melodies. "What is being analysed is something akin to the frequencies of the progressions of hills and valleys of the musical score" (p. 72). Pickover expressed the hope that "the spectrographic methods preliminarily presented here will provide a useful tool for future presentations of melodic sequences" (p. 77).

However, spectral analysis is not without its methodological limitations, particularly in the concatenation of a large number of pieces of music. A thorough analysis of the sampling aspect of the Voss study was made by Nettheim (1992). Nettheim encoded the melody from scores for solo keyboard/piano works by Bach, Mozart, Beethoven, Schubert and Chopin. Pitches were encoded in Hertz, while durations were encoded as integer multiples of the smallest duration such as a sixteenth note. Spectral analyses were undertaken of three sequences, pitch, duration, and melody (combination of pitch and duration), and compared with spectra of brown, 1/f and white music. Nettheim found that for these log $[S_v(f)]$ vs log[f] curves, $\beta \neq 1$; the slopes did not follow the 1/f curve [Figure 4.06].

> *The claim of Voss and Clarke that 1/f processes well represent pitch in music has been found in these preliminary studies of classical music to have only slender support, and the claim for duration must evidently be rejected. Some apparent confusion involving the separation of melodies into pitch sequences and duration sequences has been pointed out, and it is suggested that melody is more appropriately analysed as the single-sequence resultant, particularly if spectra are to be calculated. In the present studies of melodies so defined, the spectrum has been found to tend more towards the 1/f-squared than the 1/f function, for periods of up to four bars of music* (Nettheim, 1992:141).

Nettheim raises several difficulties with the Voss study. The first lies with the definition of melody. Voss operationalised melody as a function of the zero crossing rate Z(t) of an audio signal. Against a slowly changing background, prominent changes in zero crossing are most likely to be from high frequency components. This can be described as:

$$Z(t) = Z_0 + \Delta Z(t) \qquad (4.10)$$

where Z_0 represents background effect. Consequent spectral analysis displays changes in the zero crossing rate as $\log[S_Z(f)]$ vs log f, where Z(t) is now the correlation function. This approach has melody as a distinct component from accompaniment. Presumably Voss saw the human observer as perceiving such dominant pitches as constituting melody. While this is generally true in genres of folk and popular music, Nettheim points out many situations where melody is not so easily defined.

First, two intertwined melodic lines can be created by one instrument, as in the Bach violin partitas. Second, two or more lines may have equal melodic claim, as in a fugue or invention. Third, a melodic line can be created from contributions from two or more instruments or parts, as in piano sonatas of Mozart. Fourth, not all musical passages constitute sustained melody. Finally, rests and trilled notes can create melodic ambivalence. Voss only claimed a rough equivalence of zero crossing rate with perceived melody, and Nettheim's examples impose an upper limit to the generality of the claim. Furthermore, much music is not limited to one slow component. There may well be high rates of zero crossing from the interference effects from two or more slower components.

Nettheim suggests that the assumption of stationarity does not hold for examples of real music. Given the artistic intent of music, a musical phrase is most unlikely to exhibit stationarity. Consequently, the concatenation of musical segments for time series analysis, as performed by Voss, is invalid. None the less, Nettheim concedes that within any particular tonal constraint, "melody might in general not depart too far from stationarity" (p. 139). Given that melody has components of pitch and duration, Nettheim argues that even if pitch were to follow a 1/f spectral function, this does not imply that melody is also 1/f. The combination with a duration function which is 1/f may not necessarily result in a melody function which is 1/f. In fact, Nettheim's duration functions for real melodies were generally flat or white. He concludes that pitch spectra are virtually the same as melodic spectra.

Voss' data came from a concatenation of musical examples taken over a twelve hour period in order that power spectral density $S_V(f)$ could be measured at frequencies below 10^{-2} Hz. The sampling included pauses between movements and pieces, announcements, news breaks and commercials. Gregson (1983) also notes the danger of concatenation of different time series. In the case of the 1975 study it seems that, in order to uncover a general characterisation by

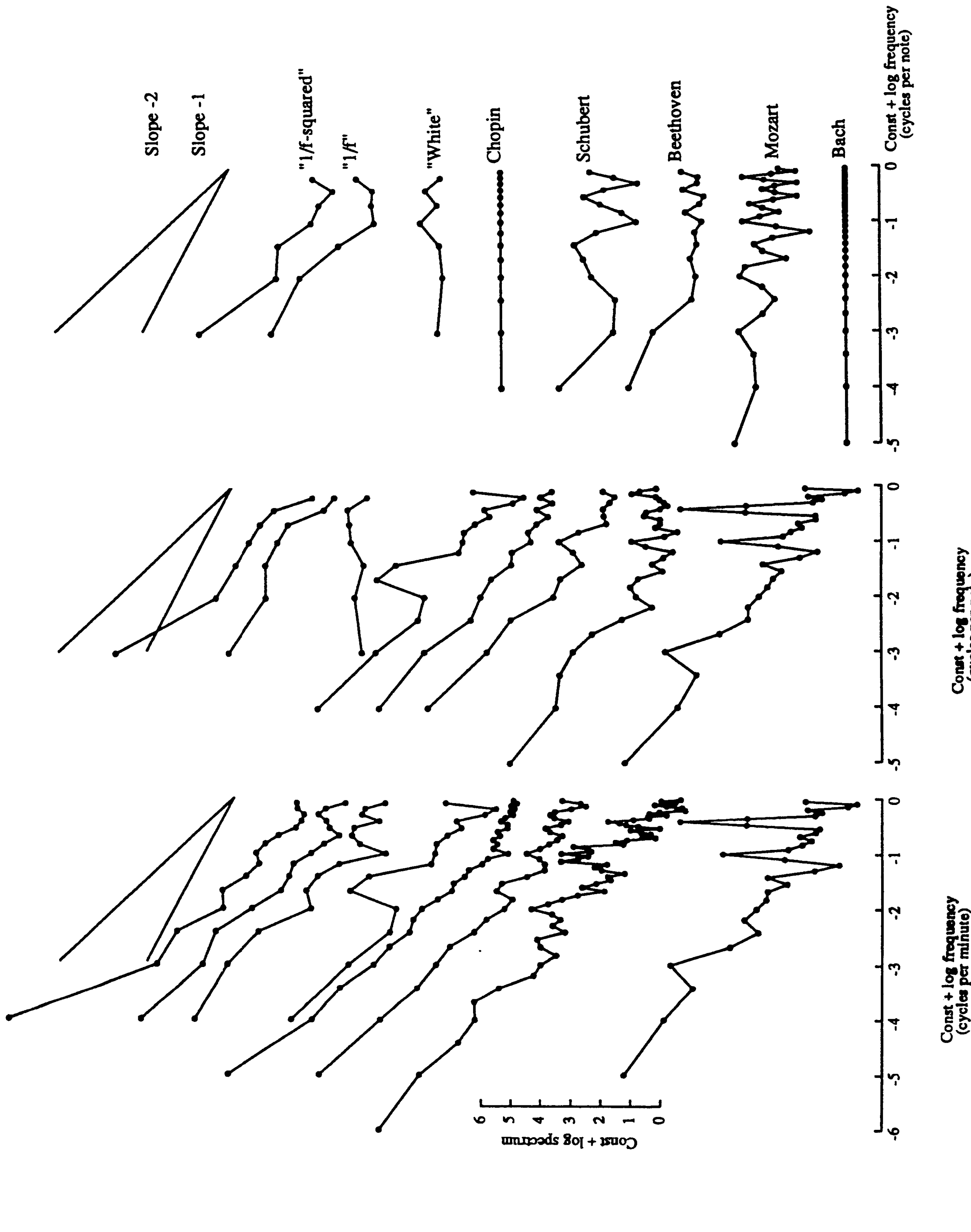

Figure 4.06 Power spectral curves for selections of music

(Nettheim, 1992: 143-145)

graphs: top - duration; middle - pitch; bottom - melody.

averaging over many pieces, a property has been posited which may not apply to any one piece of music in particular. A similar criticism is made of a similar power-law analysis by Voss of coding DNA sequences (Buldyrev *et al*, 1994). Perhaps music should be conceptualised as multifractal, with different sections or parts having multiaffinity. This would create difficulties for the task of using a 1/f generator for stochastic music composition. Nettheim predicts that the likelihood for success for such an enterprise must be low, at least with respect to composition of music acceptable in the current Western tradition.

The contrasting methodologies of Voss and Nettheim perhaps impose some restrictions on the extent to which the latter study can criticise the former. Nettheim studied melodies to four bars in length, a typical musical phrase, compared with the twelve hour durations of Voss. A possible shortcoming of Nettheim's study is that the sample durations were too brief to afford typicality, and the pitch segments analysed may be dominated by characteristic motifs of the composer, that is, the spectral estimates may be subject to large errors. In such dynamic analysis of time series, Ruelle (1992) warns, dimensional estimates for one decade are constrained to $\leq 2 \log N$. However, Nettheim used a high sampling density ($360 \leq N \leq 5000$ data points per segment) which should constrain any such error in the estimated ordinates.

A more successful demonstration of fractal form or self similarity in real music is presented by Hsu (1993) (see also Hsu & Hsu, 1991; Lewin, 1991). While acknowledging that a musical signal is best conceptualised as a time series of intervals, Hsu uses the ratios of pitch frequencies (f) to an arbitrary lowest pitch (f_0) to formalise musical pitch as a power law function

$$f/f_0 = 2^{i/n} \tag{4.11}$$

where n = number of divisions in the octave, usually 12 semitones, and i = the interval expressed in semitones. The ratio is sampled at regular intervals equal to the shortest duration in the piece, say, a sixteenth note or semiquaver. The interval-occurrence data from the resulting digitalised score is obtained after reducing large intervals to their within-octave equivalents. The occurrence of the intervals of the major second, minor and major third, perfect fourth, perfect fifth and major sixth show a log-inverse linear relationship; other intervals, including the octave, do not. This result is perhaps not surprising; the self similar intervals are just those consonant intervals which are used more frequently in the Baroque compositions which Hsu analysed. The diminished fifth, for example, was known as 'the devil's note' and avoided because of its apparent discordance. Hsu has employed this analysis to make self similar reductions of compositions, particularly of Bach, which, he claims, still sound "Bach-like" even with only a sixteenth of their original pitches played. However, this analysis does not escape the criticism that it may not characterise real music. The method of calculating intervals with respect to a bottom note is independent of keyality, i.e., does not denote the intervals with respect to the keyality of that section of the piece unless the bottom note happens to be the tonic for that modulation. This may be the case with simple music, but is much less likely with the music of composers such as Bach who characteristically display a prodigious command of harmonic inversion. Interestingly, such an endeavour as Hsu's was anticipated 35 years earlier by Pinkerton, who calculated measures of entropy as informational redundancy of simple tunes. "A set of tables could be constructed which would compose Mozartian melodies or themes which would out-Shostakovich Shostakovich. We could get as close as desired to the style of any type of music without actually copying the melodies" (Pinkerton, 1956:86).

Voss also expresses optimism towards the possibility of algorithmic music. Buldyrev *et al*, (1994) suggest that biological rhythms which display non-trivial long-range correlations, such as the heartbeat, may be better candidates for generation of music than artificially generated noises. They report a study where heart rate fluctuations were mapped on to diatonic scale intervals. Healthy (1/f-like, $\beta \approx 1$) and pathological ($\beta \approx 0$) heartbeat time series were compared. The time series from a normal heartbeat produced "a more variable (complex) type of music than that generated by the abnormal time series" (p. 81). Klimontovich and Boon (1987) present a theoretical analysis showing that music, over time spans of the order of the duration of the entire piece, has a 1/f power spectrum. Empirical support is also provided by Boon, Noullez and Mommen (1990). They note that music, over short time scales such as a few bars, is highly correlated. Over intermediate time scales, the authors argue, "one should expect to identify the specific dynamics of a musical sequence, and thereby to characterise dynamically a musical style, a composer, a work" (p. 3). With this approach characterisation of music becomes a problem of characterisation of a dynamical system, where the evolution of the irreversible time series is represented spatially as a trajectory in phase space. Boon *et al* had pieces by Bach, Mozart, and Schumann, together with ascending and descending scales, and 'random' music, played on a keyboard. The pitch contours were digitally encoded and autocorrelation, power spectra, and phase portraits were plotted. The results are interpreted as showing distinctive topological signatures in phase space for each composer. These results would suggest that the analyses of Nettheim are to be expected: music in general displays 1/f-like trends, but individual pieces display characteristic styles. This is in accord with the listening experience, where historical-compositional genres are relatively easy to identify, and the works of individual composers possible to identify after sufficient listening experience.

Voss sought the responses of listeners to music mapped from fBm noise fluctuations. Unfortunately Voss does not report the psychometric instrument he used to measure listener responses; it must be assumed the responses were anecdotal. The listeners were "several hundred people at nine universities and research laboratories. The listeners ranged from those with little technical knowledge of music to professional musicians and composers" (Voss & Clarke, 1978: 262). As the listener selection criteria were not reported it is difficult to know how legitimate are the claims for generalisability. With respect to individual differences in music perception, it would be interesting to know what proportion of respondents did not prefer 1/f to one of the other musics, or for that matter, what proportion of respondents could not discriminate at all between the different tone series.

These criticisms suggest that testing the conjectures of Voss and Clarke along dimensions of individual difference presents a number of conceptual and practical difficulties. The experimental ideal would be to be able to generate music signals with *a priori* correlational characteristics that were isomorphic with the slope of the log[S(f)] vs log f plot, and which could be used as independent variables in studies of perception. Whereas Equation 4.5 suggests that this is possible in principle through an Inverse Fourier Transform (IFT), this ideal is not so easily attainable because S(f) is a measure of average behaviour, and, simply, an infinite number of series can be constructed around any given mean. This point is not of trivial importance here. If 1/f spectral density is indeed a ubiquitous property of all music, it is to some extent, uninteresting - rather like saying all music consists of pitches and rhythms. However, the fact that this property has not, until recently, been described, and that consequent conjectures regarding human perception of fractal structure in music and relationships with natural form have only recently been proposed, suggests that this area could usefully be informed by further research.

4.3 Perception of fractal form

The thrust of many recent investigations into human sensory systems have been predicated on the position of direct perception articulated by Gibson. In this perceptual dialectic the sensori stimuli contain the information that is perceived while the perceptual apparatus has evolved to process information in just that form. As fractal form is ubiquitous in the natural world, it has been argued (e.g., Voss & Clarke, 1978; Gardner, 1978) that human sensory systems should be "tuned" to the perception of fractal form. There is evidence that this is the case for the human visual system, although precisely what is involved in such "tuning" is unclear.

Pentland (1984) generated a set of computer graphic surfaces of various values of $2 < D < 3$ scaled 1 to 10. Subject ratings of the mean roughness of these images, on a 10 point scale, were highly correlated (91%) with D. "The fractal dimension corresponds closely to our intuitive notion of roughness" (Pentland, 1984:662). Pentland argues that fractal functions provide a good model for describing human perception of natural scenes because fractal functions readily enable the matching of the reflectance function with the optimal visual signal. The image of a fractal surface retains the fractal properties of that surface. This is important in explaining how visual information is preserved for observers in various locations with respect to the perceived object.

The statistical self similarity of fractal form also allows for a high degree of information redundancy. The psycho-optic purpose of visual perception is described as being "to represent visual scenes by activity of a sparse selection of reliable and non redundant (i.e., independent) elements" (Field, 1987:2380).

> *Rather than searching for features in an image, the visual system codes a given image with regard to its relation to the statistical properties of the set of natural images. Because the space of possible pictures is so great, it makes good sense to utilise naturally occurring redundancy to recode image information into a less redundant form* (Kersten, 1987:2395).

Kersten concludes that prediction is partly based on the use of prior knowledge, and partly on the Markov statistics of target images. This may be further evident in the positive aesthetic responses to fractal form in baroque architecture (Mandelbrot, 1983), photography of natural scenery (Gleick & Porter, 1991), computer graphics (Peitgen & Richter, 1986; Geake, 1992), and abstract art which captures natural form (Geake & Porter, 1992).

Cohen, Trehub and Thorpe (1989) investigated the perception of pitch contour within an information processing framework, where "processing is facilitated by conditions of low uncertainty and high redundancy" (p. 20). The authors define macrocontour as the overall frequency directional pattern of transposed melodic segments. They predicted that consistency between contour and macrocontour should reduce subjective contour complexity of coding and hence accuracy of representation. While their results generally supported this prediction, high macrocontour complexity did not always produce low performance. The authors suggest that the problem may lie with the definition of contour, and recommend a reconceptualisation of contour which involves "the magnitude of frequency change between successive tones and the overall frequency range of the melody" (p. 27).

Like Cohen *et al*, Terhardt (1991) notes parallels with visual information processing. The concept of contour is critical. "Contourisation provides a key to a unifying concept of sensory information processing on any level" (p. 222). Each contour level is information rich, and human sensory perceptual systems have evolved "to enable an organism to respond most efficiently to external events" by "intelligent" processing which occurs "automatically" at every hierarchical level (p. 219). Thus a higher level contour represents the abstraction of information from an incoming stimulus. In vision, this involves intensity, colour and shape. Terhardt argues that the auditory equivalent is spectral pitch, because in "perception of music spectral pitch plays the role of primary contour on which any auditory Gestalt depends" (p. 223). Harmonisation creates the background - temporal pitch contour is the figure.

McCardell (1993) notes that self similarity over contour levels is prevalent in non-Western musics. In Javanese gamelan music, the rhythmic cycles of various percussive instruments are embedded in a non-overlapping hierarchy. Attention to any part can be enhanced by bringing it into the foreground, and "then through a quickening of tempo signalled by the leading drummer there is a foreground shift to the next higher or lower level. ... Thus this one fractal structure encompasses both the rhythmic system and the system of melodic counterpoint used in Javanese gamelan music" (p. 2).

However, Gilden, Schmuckler and Clayton (1993) report counter evidence that human visual perception of natural contours is not sensitive to fractal structure: rather, perception is organised to view contour in terms of macro-features or signal, modified locally by fine features or noise. Gilden *et al* used "a particular class of algorithms ... that have exactly the same sensitivities to structure as human observers" (p. 461) to model subjective perception of natural contour. Two hypotheses were tested: the first, following Voss and Clarke, that "perception is efficient because perception has been tutored by natural form"; the second, "that perception may have its own protocols ... for reasons unrelated to the occurrence of fractal structure, and it is primarily coincidence that leads to the agreement between environmental frequency and discrimination sensitivity" (p. 467).

The authors show that a bipartite spectral analysis of rough and smooth contours in terms of range, amplitude and difference in bandpass power all yield monotonic but non-linear relationships with the spectral power exponent β. It follows that fractal dimension is not a sufficient statistic for the ranking of fractal structures. Gilden *et al* employed Monte Carlo simulation algorithms on discrimination data to determine how contours are analysed, and what contour features are extracted. Hybrid fractal or multifractal contours were generated for comparisons. The authors note that fractal power spectral curves, log $S_V f$ vs log f, are self-affine, i.e., invariant under vertical and horizontal shifts. Hybrid contours, being non-self-affine or multiaffine, have a different increment frequency from fractal contours. The authors argue that the width of the increment distribution, basically small or large, should be "perceptually penetrable" but that increment correlation may not. Nevertheless, contrast functions were defined for both characterisations within a 2 x 2 design for feature (range or spectral power) by contrast procedure (within or between contours). Subjects were asked for same-different judgements on the hybrid contour pairs. These data were compared with those of simulated observers. Best agreements were found for simulations of bipartite decomposition, i.e., when contours were analysed as signal and noise. "These two forms of structure are not regarded as having a coherent relationship but rather are treated as two separate categories of information. Discrimination is based on range comparisons within components" (p. 475).

Gilden *et al* conclude that "there is no support for the notion that human observers use spectral power in discriminating between fractals. The evidence is clearly that they use the much simpler and more visually salient feature of range" (p. 475). These authors argue that although "the environment is teeming with fractal structure" it does not follow that humans are necessarily sensitive to such structural hierarchies, particularly multifractals. "From an ecological point of view, it makes little difference how information is processed so long as the animal is informed" (p. 465). The authors suggest that the sensitivity findings of earlier researchers such as Field (1987) and Knill, Field, & Kersten (1990) fortuitously peak around D = 2.5 due to the "logic" of bipartite perceptual organisation which "has nothing to do with fractal form" (p. 476). Clearly the generality of such conclusions extends well beyond the specificity of the research. The authors acknowledge this, but argue nevertheless that such general claims can be made given the integrity of the Monte Carlo method with a population of 200 contour pairs for each block in the experimental design. Whereas the reported results may support claims for perception of contours, the perception of other structures could be another matter, although, as the authors note, a separate perceptual organisation for fractal structures is most unlikely. That such a recognition procedure would require a hierarchical emphasis of form and context over particular features has support from earlier literature (Palmer, 1977).

The issue for this research program, however, is whether an analogous argument can be made for musical objects, or at least for fBm tone series. Bregman (1994), for example, argues that visual and auditory scenes can make quite non-analogous demands on human perceptual processing. Schmuckler and Gilden studied perception of auditory fractal contours of white, brown, and 1/f power spectra on dimensions of pitch, duration and loudness. In the first study, listeners could discriminate between signals with different power spectra on the dimensions of loudness and pitch, but not on duration. This is consistent with the findings of Nettheim. The second study employed different ranges to investigate the role of streaming in spectral discrimination. The results were interpreted as evidence against streaming as the perceptual mechanism involved in making such discriminations. The third study investigated perception of auditory fractal contours over a range of power spectra on dimensions of pitch and loudness. An inverted U curve for discrimination (D prime as a measure of discrimination) vs spectral exponent (- β, the negative slope of the log power spectra vs log frequency plot) was reported [Figure 4.07].

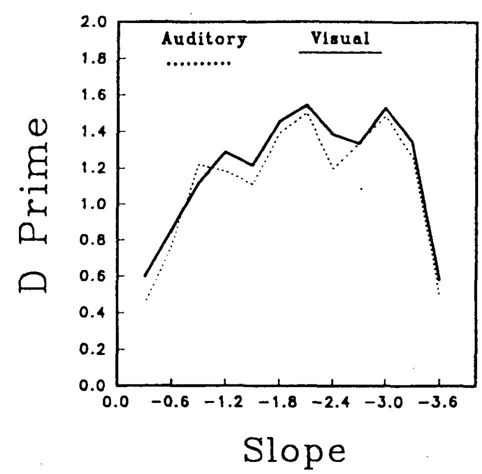

Figure 4.07 Discrimination acuity (D prime) vs power spectral slope of visual and auditory fractal contours (Schmuckler and Gilden, 1993: 657)

The reported maximum is around β = 2.0 to 3.0, in apparent support of Knill *et al.* (1990) and earlier results on the visual perception of natural imagery, e.g., Field (1987) and Kersten (1987). However, there was no maximum at 1.0 as suggested by Voss and Clarke (1975, 1978). Schmuckler and Gilden interpret this result as indicating further support for the conclusion of Gilden *et al*: the evidence that our perceptual systems are tuned to the statistical properties of the environment is fortuitous, and there is "no heightened sensitivity in the range of auditory sources" (p. 657).

However, there seem to be several conceptual inconsistencies with this report. Most seriously, Schmuckler and Gilden seem to have muddled fractal dimension D with β. First they report fractal dimensions with negative values (pp. 656, 658). This is interpretatively challenging. Presumably they mean β, not D, as on p. 657 where they do use "slope" rather than "fractal dimension". Second, they report Voss and Clarke's (1978) result as concerning a fractal dimension of -1 (p. 656). This is inaccurate. Voss and Clarke's maximal value was for slope = -1, not D, which Voss (1988) shows to be around 1.5 in such cases. In their discussion of their inter-modal comparison of maximal β Schmuckler and Gilden do not refer to the analytic relationship between D and β, Equation 4.7, for fBm. To compare results of fractal perception across modalities, E should be adjusted for the embedding dimension of the modality in question: E = 1 auditory; E = 2 visual planar. Their reported result of $2 < \text{slope} < 3$ for an auditory modality, after E is adjusted from 1 to 2 for comparison with visual perception, becomes $2 < D < 2.5$, which is close to Knill *et al* (1990) and Kersten (1987). Schmuckler and Gilden measured perceptual acuity by a comparison task between signals with differences in power spectral slope of 0.6. However, this may not be a perceptual linear scale. A comparison at the white end of the scale, say between $\beta = 0.0$ and $\beta = 0.6$, may not be equivalent to a comparison at the brown end, say between $\beta = 3.0$ and $\beta = 3.6$.

Schmuckler and Gilden employed a visual benchmark for subjects to use to make auditory decisions. This experimental design makes an implicit assumption that subjects can visualise an auditory signal, or audiate a visual signal, or both. To the extent that subjects were reliant on the visual information, there may be additional variance introduced through the demands of cross-modal perception. Handel and Yoder (1975) studied the discrimination of patterns presented for perception by auditory and visual modalities. They concluded that the dimensions on which discriminations are made are modally specific. The reverse process, audiating a visual signal, when applied to pitch, is just sight reading a music score, or in contemporary genre, interpreting an analogue score where scores like the graphs in Figure 4.02 are not unusual. It could be expected, then, that musical training would have an effect here, but this was not the case.

The contour visualisation of other musical dimensions, viz., dynamics and duration, are more problematic. Whereas we perceive pitch to have contour, ups and downs, we do not mean the same when turning the volume up or down. In performing music, dynamics are referred to as "loud" or "soft", and intonational adjustments are often referred to as "higher" and "lower", as well as the older "sharper" and "flatter". Duration is never referred to in terms of contour by listener or performer, and, as the authors note, is strongly arithmetic in its sub-divisions compared to pitch and loudness which are logarithmically scaled. Thus it might be anticipated that subjects would make some sense out of dynamic variations, but little discrimination on duration, as was found.

The argument that human perceptual organisation is independent of the structure of stimuli as applied to music stimuli has been addressed earlier in Chapter 3 where some of the literature reviewed does support this position. Gilden *et al* (1993) argue for a bipartite decomposition as a principle of perceptual organisation. "Most perceptual tasks are linked to distal events that do have a signal-noise structure" (p. 477). Importantly, this is distinguished from the mutuality of other figure-ground protocols. Here the signal "is indeed something essential" while the noise "is treated in perception simply as something extra, added on" (p. 477). In music, then, what is the signal and what the noise? If melody is "added on" to the harmonic development, or accompaniment, then bipartite decomposition flies in the face of the most widely agreed finding in the music cognition literature, the perceptual prominence of melody. It was the prominence of melody on top of an accompaniment that Voss operationalised as the rate of zero crossing. It is the regularities in the zero crossing rate which distinguish melody from background noise. If noise in music is not melody but ornamentation, or less obviously, interpretational techniques such as *rubato*, then new explanations must be contrived for the continuance of a critically acclaimed classical repertoire. Schmuckler (1994) comments that there is no isomorphism between signal-noise and divisions in musical structure; both signal and noise are present in all musical parameters - melody, harmony, etc. This suggests that signal and noise may be processed in parallel for each perceptual characteristic, a view which seems difficult to reconcile with some of the literature reviewed in Chapter 3 concerning the prominence of successive processing in music cognition.

How might bipartite perception be understood in terms of simultaneous and successive processing? An answer may lie in the results of Warren and Bashford (1993) who found that minimal differences in acoustic sequences of brief tones, duration < 100 msec, too brief for tonal perception, could nevertheless be discriminated. They suggest that such rapid sequences form temporal compounds for perceptual discrimination. Stochastic patterns are processed globally as a complex pattern which can be used for comparison with other such patterns. This description seems at odds with Bregman's concept of auditory streaming (Barsz, 1988). The conflict can be resolved with recourse to the Luria model. Without acknowledging it as such, Warren and Bashford's interpretation is consistent with a description of the Luria model dimension of simultaneous synthesis. These authors note "that it is not necessary for auditory sequences of brief items to be processed as perceptual sequences" (p. 121), in effect re-stating Luria's finding that simultaneous and successive processing are independent. Implicit in the Luria model is some requirement for perceptual organisation, as an information processing task, to be guided by the manner in which information is perceived.

The findings of Warren and Bashford provide other grounds to challenge the argument for bipartite decomposition as the overriding organisational mode for all perception. Fine distinctions in noise, and the rapid sequences described by subjects as "wooshing", are discriminable. At presentation rates too fast for identification of melodies, or much less slower moving accompaniments, subject accuracy at correct identification was between 80% and 98%. Warren and Bashford also report that response times to discrimination "were not correlated across listeners, suggesting that perceptual organisation ... may be quite idiosyncratic" (p. 125). In other words, even among four subjects, individual differences were observed in the time it took to extract "whole pattern features" for discrimination, i.e., in abilities on simultaneous synthesis. Although Gilden *et al* (1993) report a large number of data, only five subjects were used. Again, only five subjects were used by Schmuckler and Gilden (1993) in their third study. This could compromise generalisability if individual differences in perceptual abilities account for a significant portion of the variance in the data.

In their conclusion, Schmuckler and Gilden (1993) are somewhat more circumspect. "We find that fractal structure plays a role in discrimination, although it is not an overwhelming source of information", and, "fractal information might simply be another potential source of structure, working in concert (or sometimes in opposition) with other aspects of auditory structure" (p. 658). These limitations notwithstanding, and while acknowledging the authors' note regarding the preliminary nature of their research, the results of Schmuckler and Gilden (1993) and Gilden *et al* (1993), that human perception is not preferentially tuned to fractal structure, do present a serious challenge to the conjecture of Voss and Clarke. Bipartite perception and fractal perception, then, present alternate interpretative frameworks for some of the results of this research where the focus is on the contribution of information processing abilities to individual differences in perception of music signals.

4.4 Sensitivity to autocorrelation structure

Given the conflicting evidence for fractal contour and spectral density function as psychologically penetrable characterisations of a musical signal, other characterisations require consideration. Boon *et al* (1990) consider and reject both information entropy and fractal dimension D of the phase trajectory as suitable quantitative characterisations. They note that D is not a unique measure and that all musics analysed had D around 2, $1.75 < D < 2.25$. Mandelbrot (1993) also argues that the fractal dimension is not a sufficient statistic for the description of an object since any one value of D can result from many different examples of fractal form. Mandelbrot suggests the concept of lacunarity, "the tendency to have holes" (p. 12), as a characterisation worth consideration. Mandelbrot derives a formula for the rate of lacunarity Λ_F based on the upper and lower limits of human perception. Just how lacunarity could be applied to a time series or music is a matter for exploration. Auditory power could possibly be characterised with Λ_F.

Information entropy is a measure of the rate of unpredictability of the music sequence. With respect to tonality, unpredictability is manifest in "the liberty taken with respect to tonality" (Boon *et al*, 1990: 10). As music generally does not show such departures, information entropy will often be zero. Boon *et al* introduce parametric entropy as a measure of deviations from tonality that can account for musical complexity by using weighted probabilities for successive intervals. As values obtained varied by piece across composers, the authors conclude that parametric entropy "appears to be characteristic of composition rather than composer" (p. 12). Although parametric entropy is clearly a function of tonal context, the authors offer no evidence that it is a psychologically perceivable characterisation.

The most appealing characterisation, which has at least conceptual parallels to the requirements for informational redundancy outlined in Chapter 3, is that of the autocorrelation function R_{vv}. Gregson and Harvey (1992) note that, because autocorrelation is related to internal redundancy, it is a property that could be used for discrimination. It is thus conjectured that the processing of information from a musical signal might include an awareness of, or sensitivity towards, its autocorrelation structure. The evidence in Chapter 3 for the perceptual independence of the musical characteristics of pitch, rhythm, dynamics etc., suggests that this sensitivity would arise separately for each characteristic. The interest in this research is on the perception of fluctuating pitch. Thus it is conjectured that the processing of information from a fluctuating pitch signal includes a sensitivity towards its autocorrelation structure. Such awareness, for example, may be evident when a composer seeks a plausible melodic line, or a listener detects a 'wrong' note. This is not to argue that R_{VV} is a sufficient statistic. Different pitch series with embedded subseries can have the same overall R_{VV}. But given the restricted range of any music characteristic, e.g., rhythm or pitch class, the discovery of a wholly sufficient statistic seems unlikely.

The conjecture can be situated within the information processing framework of the Luria model. From Chapter 3 it is clear that musical memory requires some sense of correlation over various time spans. A reliable working memory for music is one of the attributes of high music ability. In Chapter 2 it was noted that the capacity of working memory is also a measure of ability on successive synthesis. In Chapter 3 it was argued that the musical attributes of long term musical memory such as major scale intervals, tonic triads etc. are schema which are constructed by simultaneous synthesis from correlational procedures applied to time coded pitch data. Furthermore, the Luria model posits a hierarchical cyclic relationship between successive and simultaneous processing, coordinated through executive processing. In a similar manner, autocorrelational processes must also be hierarchical, beginning with the correlation of frequencies into a tone or chord, followed by the correlation of tones into a melodic phrase and chords into a modulation, followed by the correlation of phrases into a melody and modulations into an harmonic progression. It is not suggested that this hierarchy is necessarily fractal, although such a hierarchy might be like a multifractal structure (Gregson, 1994). In the second and third experimental studies of this research it is hypothesised that sensitivity to autocorrelation structure will be positively related to abilities on the three Luria model dimensions. That is, sensitivity to autocorrelation structure is an aspect of music information processing which, consequently, is a variable which generates individual differences.

Such a suggestion is consistent with the musical object conceptualisation of Hurssel (Godoy, 1994), where quasi-spatial renderings of musical constants have resolutions over all time scales from single tones to the length of the piece, and with the model of musical attention of Boltz (1991), in which expectancies are generated from the structure of the information.

> *Many events within the natural environment can be considered to be highly coherent in that they transpire over time spans marked by non arbitrary beginnings and ends. Within these time spans, information tends to display a high degree of predicability with multiple levels of structure interrelated through a hierarchical scheme. Such structural arrangements are commonly found within speech utterances and conversations ... and musical compositions* (p. 423).

To begin at the most primitive level, given that the spectrogram of a typical musical signal is made up of a mixture of overlapping frequencies,

> *how can the auditory system know which frequency components to group together to build a description of one of the sounds? It seems that it does so by looking for correlations or correspondences among parts of the spectral content that would be unlikely to have occurred by chance* (Bregman, 1994: 655).

Such an argument was advanced earlier by Licklider (1951). Musical (and non-musical) sounds are typically complex: timbrel fusion, tonal masking, and volume increase with pitch, are common complex auditory phenomena. Unlike pure tones whose frequency can be determined from a mapping of the location of the stimulus on the cochlea, complex sounds cannot be determined by such a simple spatial mechanism. Licklider posits a non-linear process whereby fluctuations of cochlear stimulation resonate with periodicies through increasing bandwidths. Musical elements arise from mappings of these periodicies by autocorrelation processes. Licklider's model is notably robust. It shows that consonance is not a necessary condition for dissonance. It also explains why noise volume grows more rapidly than the linear addition of pure tones would predict. The prevalence of power laws in psychophysics, within narrow stimulus input levels, arise from this tendency of human sensory systems to perceive equal sensation ratios for equal stimulus ratios. In contrast with such considerations, Parmcutt (1989) argues that "the relationship between sound input to the ear and the information conveyed to the brain is essentially the same as the relationship between a sound and its pure tone components.

In this sense, the ear subjects incoming sounds to spectral analysis" (p. 26). Parmcutt suggests that the best mathematical model for such spectral analysis is a Fourier transform from which the acoustic waveforms undergo an exponential decay representing diminishing relevance over time. However, the coarse-grained nature of the human perceptual response suggests that the auditory pathway does not undertake spectral analysis by way of direct Fourier transformations (FT) (Hochberg, 1984). Radocy and Boyle (1988) note that the just-noticeable-differences (jnds) of acoustical responses are non-linear functions of frequency, intensity, duration and rate of change. Also, Gregson and Harvey (1992) present evidence that human observers can discriminate between chaotic sequences and random sequences. Such discrimination is not possible via FT since both chaotic and random sequences have similar flat spectra and the FT of these sequences are indistinguishably similar. Clearly Parmcutt's position in this regard is untenable; Licklider's model seems more plausible.

Other evidence consistent with Licklider's model explains the Pythagorean privilege of consonant intervals such as the octave and the fifth. Beament (1977) notes that the auditory mechanism has a peculiar reaction to sounds which occur in whole number frequency ratios. A transposition of pitch means that frequencies are multiplied by the same ratio, which implies a shift in position of resonance on the basilar membrane. "The mechanism matches experience, and supports a generalisation that the information in sound is obtained from the pattern, rather than the specific location of the pattern" (p. 9). Beament argues that the physiological source of individual differences in abilities in music perception is either in the special regularity of sensor firing in some people, or in their auditory processing machinery which optimally matches sub-harmonic firing rates. Rakowski (1990) found that interval recall in music, and with phonemes in natural speech, varied for a range of frequencies of about ± 50 cents around the correct frequency. Consistent with such a model, Patterson (1990) shows that perception of octaves in multiharmonics can be effected by the removal of certain of the harmonics. He concludes that the human auditory system, like the visual system, works well with natural complexity.

There is some evidence for the role of primitive local autocorrelation structure in the conveyance of musical meaning. Ando, Okano and Takezoe (1989) used autocorrelation to determine listeners' preferences for reverberation times of Japanese music. Listeners preferred a short delay time of about 40 msec which the authors suggest may indicate the temporal width of the "psychological present". This is longer than the 2 to 16 msec as the limits of the perceptual auditory present measured by Woodrow (1951) using jnds. But 40 msec is just the size of Warren and Bashford's temporal compound. This result could suggest that chunking begins at a time scale too brief for conscious experience. This thrust has support from Bregman (1994).

> *Many things we take as self-evident, such as the coherence of a single note or of a single voice, are perceived through a process of grouping. Another non intuitive idea is that many qualities that seem to be the automatic products of simple acoustic properties also depend on grouping. These include pitch, loudness, timbre, location, and dissonance. ... the processes of audition that can accomplish the grouping and use it to derive these experiences must be doing so in time periods that we have to measure in milliseconds* (p. 704-705).

At the cognitive level in the correlation hierarchy, Leman (1994a) (see also Leman, 1992, 1994b) proposes a model of schema-based tone centre recognition which employs autocorrelation of individual tones. Leman's model is 'data driven' rather than top-down, and in this sense, stands in opposition to Narmour (1990). The tonal images from signal information are not the formal tonalities of harmonic analysis, but tone centres - dynamically formed attractors in tonal space. Here, context is all important. A single tone has meaning only through its relationships with surrounding melodic and harmonic tones. For a simple example, in a IV - V - I chord progression in C major, the tone C progresses through the scale degrees of a 5th to a 4th to a tonic. In Leman's model, a piece has many different tone centres simultaneously, but some are stronger attractors than others. Autocorrelation plays a crucial role as one of the perceptual outcomes of information extraction from a musical signal. Whereas musical recognition is a function of following the moving 'time index', "interpretation involves the reconsideration of past interpretations in view of new contextual evidences" (Leman, 1994a: 201). That is, interpretation involves being influenced by different tone centre attractors from those of the listening present, viz., attractors of the time-limited past, and attractors of the integrated past. The attractor power of a tone centre is determined by its (lack of) shared variance (labelled as "distance") with other attractors in tonal space. Such distances are computed from local autocorrelation functions over limited time-lags and the entire piece. It could be noted that computing such limited autocorrelations is analogous to the lacunarity approach in characterising fractal structures.

In its direct response to sense data Leman's model is Gibsonian. In contrast to some other models of music perception, schema are not passive templates but active operators which "process contextual information in order to interpret the images of a near past" (p. 187). That is, perceptual learning of musical stimuli is self-organised. Leman conjectures that "high level brain maps may exist for tone centre recognition" (p. 199). Leman notes that "the autocorrelation analysis comes down to the transformation of temporal aspects of the signal into a spatial organisation" (p. 198). It should be noted that there is no implicit suggestion of topological correspondence between tonal space and its neurological mapping, but that such auditory information can be cortically mapped is consistent with Luria's descriptions of brain function.

Brown and Cooke (1994) present a computer model of perception of timbre which is similar in concept to Leman's model for tonal centering. Timbre is conceptualised within a two-dimensional space of brightness and onset asynchrony. With a Licklider-type autocorrelation analysis performed at each periodicity centre frequency, the Brown and Cooke model combines onset, frequency transition rate and offset to make auditory elements which are then grouped by two grouping processes, simultaneous and successive, into pitch and timbre. At each time frame, possible pitch periods are estimated from periodicity information in the autocorrelation map. A dynamic algorithm is employed to find the most likely pitch profile, as per Leman. Thus timbre is influenced by the degree of non-stationarity of spectral energy. Tonality is established by extracting a global contour from the average of local pitch contours of elements. It should be noted that the models of Leman, and Brown and Cooke, embrace the current paradigm of cognitive self-organisation in which boundaries between perceptual and cognitive processes become diffuse.

There is some earlier evidence that the coherence of tone sequences is perceivable at the cognitive level. McMullen and Arnold (1976) presented subjects with pairs of rhythmic sequences with various levels of redundancy. Subjects were asked to choose the most preferred or the most interesting. The researchers hypothesised that the distribution of preference responses should form a Berlyne 'Wundt' or inverted U curve, while interest responses should monotonically increase with complexity. The results indicated some support for the hypotheses. These findings could suggest that the bipartite perceptual model of Gilden *et al* is not necessarily in contrast with the fractal model of Voss and Clarke. Gilden *et al* argue that perception discriminates only signal and noise, and not intermediate structure. Signal to noise ratio is also a monotonic function of signal complexity; discriminations are easiest where complexity is low and signal is most prominent. Aesthetic responses, on the other hand, are maximal where redundancy, or coherence, is mid-way between low values with low interest, and high values where complexity is impenetrably high. This implies that aesthetic responses may not altogether depend on perceptual acuity or thus, on music ability. For example, maximal aesthetic complexity may vary with experience and exposure. Evidence for such a position was found by Madsen, Byrnes, Capperella-Sheldon and Brittin (1993); in contrast, earlier evidence for differences in aesthetic responses which was independent of music experience is reported by Parker (1978).

Nevertheless, McMullen and Arnold make two observations which suggest that the coherence of rhythmic sequences is directly perceivable: first, redundancy was best computed from rhythmic subunits, anticipating the reductional analysis of Hsu (1994); second, some sequences required longer processing time than others. This is consistent with Shuter-Dyson and Gabriel (1981) who suggest, based on an overview of other evidence, that "when pitches and intervals are sounded in a series then the result may be a new entity with its own rules for being perceived and remembered" (p. 250). Bregman (1994), for example, presents evidence to support the proposition that the perception of streaming is significantly affected by the regularity or irregularity of the sequence of tones.

Such findings are reported by McMullen (1974) with pitch sequences. Complexity was varied on two independent variables: redundancy, and the number of pitches used to generate the sequences, five, seven or twelve. School-aged subjects preferred melodies with intermediate or low redundancy over those with high redundancy, and preferred melodic sequences generated from five or seven pitches over those generated from twelve. That is, children preferred scale-like over twelve-tone generation of pitch sequences. Whereas it could be conjectured that scale-like generation of sequences makes them easier to chunk into a Gestalt for later evaluation, Warren, Gardner, Brubaker and Bashford (1991) argue that temporal organisation of tone sequences is not based on consonance. Sorkin (1990) agrees. His evidence supports a correlational model of discrimination in which listeners use temporal rather than spectral processing. Sorkin suggests that temporal correlation may occur within pitch bands or streams, to be later combined as a composite.

Whereas the term "chunking" postdates Luria's writings, Luria (1966b) describes that concept in the relationship between successive and simultaneous encoding of language: "the organisation of a series to be memorised as a result of which a series of words, repeated several times, is gradually converted into a group and acquires the character of a single simultaneous system" (p. 75). Kirby and Das (1990), and Das *et al* (1994) describe information processing as a cyclic hierarchy of simultaneous and successive coding. Chapter 2 considered this interplay between successive and simultaneous processing in some detail in the task of reading text. The process is spirally hierarchical over several scales from individual letters to whole paragraphs but perceived all at once. Likewise, music perception involves the appreciation of order over a hierarchy of time scales and all at once. Honing (1993) supports a model where discrete musical elements are organised through grouping processes. Simultaneous grouping processes form the acoustic surface into musical events; sequential grouping processes connect these events into streams. Segmentational chunking then organises these streams into musical units in a cyclic hierarchy. As with reading, the levels are available for perception all at once. This is not to say that all processed information necessarily comes into perceptual awareness, much less simultaneously. Which particular level 'dominates' awareness is determined by the attentional processes of executive synthesis.

Hierarchical structure which is available "all-at-once" may be analogous to the self-scaling of multifractals. Feder (1988) emphasises that with the bi-fractal time series of ocean wave heights, $T_1 \approx 2$ days, $T_2 \approx 2$ weeks, both values of D are for local dimensions and not for a crossover effect from local to global conditions. Rather, local conditions change from persistent to independent behaviour over T_2. In music this could describe the phenomenon where, for any particular harmonic modulation, where $T_2 \approx$ a few bars and $\mathrm{T}_1 \approx$ a few notes, a very large number of different phrases have been written or improvised. Bipartite perception, then, could possibly be interpreted as bi-fractal perception.

Such self-similarity across time scales may be the consequence of the "severe constraints on the possibilities of musical form that can successfully be realised within ... a system [of] hierarchical ... serial ordering of tonality " (McAdams, 1987: 49). In this sense human cognition may be regarded as optimally adaptive to the information processing demands of the environment (Anderson, 1991; Edelman, 1994). Eibl-Eibesfeldt (1988) argues that self-similarity in form is the basis for human aesthetics. Self-similarity may facilitate parsimonious functional relationships between perceptual categories and the discrete form-bearing elements of a musical signal. It is important to exercise some caution in the application of such a general notion to particular aspects of individual compositions. Boroda (1993) for example, attempts to demonstrate self-similar distributions of motifs as an indication of the evolution of compositional style. The variety of motifs that are included within each classification, however, render this analysis unconvincing.

Moreover, in an experimental setting, measuring the perception of autocorrelation structure is not unproblematic. Gregson and Harvey (1992) note that, as information processing capacities are obviously not unlimited, an autoregressive moving average function with lag orders of up to 3 is a plausible upper bound for purposes of identification. The sensitivity to autocorrelation structure conjectured above would be consistent with this limitation. Where music is regarded as a chaotic system, following Boon *et al* (1990), it is worth noting again Gregson and Harvey's caveats on the use of the autocorrelation function R_{VV} as a sufficient statistic for the assessment of similarities between non-linear realisations of note sequences. As a linear statistic, R_{VV} fails to detect the non-linear dynamics which distinguish chaotic attractors from random stochastic processes. The autocorrelation of "non-trivial interdependencies" may compute to zero.

However, real music is not directly generated by a chaotic algorithm (leaving aside the role of chaotic attractors in brain activity, and the similarities to short sections of some contemporary minimalist compositions of the outputs of many non-linear algorithms (Pressing, 1990) notwithstanding). Nor is music random. Thus, the limitations on the representational capacity of R_{VV} in real music may not be as severe as noted above. The autocorrelation spectra or autocorrelation structure may be a perceivable characteristic of a piece of music, which, like other characteristics such as scale or rhythm, is a necessary but not sufficient characterisation.

In contrast, Gilden *et al* argue that correlation may not be a perceptual property of contours because it "is completely confounded with the manifest appearance of the contour", and "although it might be possible to develop models for how people derive partial calculations of ... correlation, it is not obvious how to proceed" (p. 471). This research attempted to find a possible route. Following Licklider, it seems more likely that listeners perform complete correlations than partial correlations. In the experimental situations described in the following chapters, the R_{VV} was used to generate stimulus pitch series with greater and lesser degrees of coherence, in order to investigate how individuals differ in their sensitivity to differences in the autocorrelation structure.

Chapter summary

Voss (1975) and Voss and Clarke (1975, 1978) investigated the power-law characteristics of music. They conclude that music follows a 1/f spectral function, like natural fluctuations in time and space. Thus music can be said to have fractal form. Music generated from 1/f processes was preferred over music from other power spectra. It was conjectured that human perception may have a preferential acuity for fractal form. Several studies in visual perception found evidence to support this position. In recent years Voss and Clarke's methodology has been questioned, and hence the validity of their conclusions. Moreover, Gilden *et al* (1993) have produced strong evidence against preferential perception of fractal form in visual and auditory contours. Gilden *et al* propose a model of perception in which fluctuations are perceived independently of context.

For this research a perceivable characterisation of information in a music signal was required. In the light of Gilden *et al* the power-spectral exponent was regarded as problematic. It was conjectured that autocorrelation plays an essential role in the formulation of musical Gestalts for higher order processing. Leman offers a model of music perception in which schema actively provide musical context for self-organised tone centres created from autocorrelational images of present and past. There is evidence which can be interpreted as supporting some degree of awareness of local autocorrelation processes. It was noted that, like other perceptual characterisations of music, the autocorrelation function is a necessary but not sufficient statistic. It was conjectured that the autocorrelation structure may be a perceivable characterisation of music. The following chapters describe the ways in which the plausibility of this conjecture was investigated.