Melodic Expectation in Music: A Review of Theoretical Accounts and Empirical Research

Marcus Pearce

Department of Computing, City University,
Northampton Square, London EC1V OHB
m.t.pearce@city.ac.uk

1 Introduction

The ability to anticipate and prepare appropriate responses to events is highly important in an adaptive sense and conflicts or correspondences between actual and anticipated effects often entail significant psychological and biological effects. The development of systems for expectation may derive from both innately determined responses to universal properties of the environment and responses acquired through long-term exposure to the regularities of any given environment. Expectancy is considered to be especially important in music perception where it may account for the expression of meaning and emotional responses to music (Meyer, 1956) and more sophisticated models of expectancy in music have set out in detail the principles governing the influence of universal bottom-up factors and acquired top-down factors (Narmour, 1990, 1992). Theoretical accounts of expectancy in music have been studied in terms of the influence of temporal structure (Jones, 1987; Jones & Boltz, 1989), melodic intervals and registral direction (Cuddy & Lunny, 1995; Krumhansl, 1995a) and harmonic expectancies (Bharucha, 1987; Schmuckler, 1989). Although expectancies in music perception operate over a number of different parameters and structural levels in music, we are concerned here with expectancies in purely melodic music. A variety of tasks have been used to study expectancy in melodic music including rating completions of melodic contexts (Cuddy & Lunny, 1995; Krumhansl, 1995b; Schellenberg, 1996), generating continuations to melodic contexts (Carlsen, 1981; Schmuckler, 1989; Thompson *et al.*, 1997) and the classification of melodies and memory for melodies (Schmuckler, 1997). We focus on those tasks that have actually elicited expectancies as the dependent variable.

These notes are structured as follows. In §2, we examine some theoretical accounts of expectancy in melody including that of Meyer (1956, 1967, 1973), the implication-realisation model (Narmour, 1990, 1992), theories of dynamic-attending (Jones, 1981, 1982, 1987, 1990) and theories based on sequential pattern induction and extrapolation (Deutsch & Feroe, 1981; Simon & Sumner, 1968). In §3, we summarise the experimental work that has examined the prediction of some of these theoretical accounts. Finally, in §4, we summarise this work and discuss some of its implications.

2 Theoretical Approaches

2.1 Meyer's Theory of Musical Expectancy

In his book, *Emotion and Meaning in Music*, Meyer (1956) discusses the dynamic cognitive processes in operation when we listen to music and how these processes not only underlie the listener's understanding of musical structure but also give rise to the communication of affect and the perception of meaning in music. Meyer begins by making an important distinction between two ways in music can be said to have meaning. The first, *absolute* or *embodied* meaning, describes the manner in which musical objects acquire meaning through reference to or being referred by other musical objects. The second, *referential* or *designative* meaning, describes the manner in which musical objects acquire meaning through reference to extra-musical objects. A second distinction drawn my Meyer concerns approaches to studying meaning in music. In the *formalist* approach, meaning in music is a result of the perception and intellectual understanding of the structure of a musical work. In the *expressionist* approach, on the other hand, the same relationships are capable of inducing emotions in the listener and these (rather than the intellectual comprehension of the relationships alone) are the object of study. Meyer places his discussion firmly in the context of *absolute* meaning in music and, within this context, attempts to provide a conception of musical meaning and experience which accounts for both the formalist and expressionist approaches.

Broadly speaking, Meyer proposes that meaning arises through the manner in which musical structures activate, inhibit and resolve expectations in the listener about other musical structures. Meyer notes that expectations may differ independently in terms of the degree to which they are passive or active, their strength and their specificity. He contends, in particular, that affect is aroused when an passive expectation induced by antecedent musical structures is made active when it is temporarily inhibited or permanently blocked by consequent musical structures. Meyer discusses three ways in which the listener's expectations may be violated. The first occurs when the expected consequent event is delayed, the second when the antecedent context generates ambiguous expectations about consequent events and the third when the consequent is unexpected. While the particular effect of the music is clearly dependent on the strength of the expectation, Meyer argues that it is also conditioned by the specificity of the expectation.

Meyer provides a further analysis of the experience of embodied meaning, dividing it into three stages. In the first of these, *hypothetical meanings* are attributed to antecedent musical structures as a consequence of the expectations generated by those structures. In the second stage, *evident meanings* are attributed to antecedent musical structures when consequent structures have occurred and the actual relationship between the antecedent and consequent has been apprehended. While it is clear that evident meanings may be conditioned by hypothetical meanings, Meyer also emphasises that evident meanings influence hypothetical meanings in a process of revaluation. Thus in situations where the antecedent arouses uncertainty regarding the consequent, evident meanings may cause confirmation, alteration (due to delay or deviation), clarification (through the resolution of ambiguities) or complete revaluation of hypothetical meaning (in the case where the consequent was unexpected). This revaluation influences the subsequent expectations of the listener. The final stage of development, involves the attribution of *determinate meanings* once the totality of relationships between hypothetical and evident meanings has been apprehended throughout the piece and on several hierarchic levels.

Meyer (1957) discusses the relationship between his theory of musical expectancy and concepts in information theory. He starts with the suggestion that "once a musical style has become part of the habit responses of composers, performers and practiced listeners it may be regarded as a complex system of probabilities" (Meyer, 1957, p.?) and that expectations arise out of these internalised probability systems. In particular, he suggests that a musical style may be conceived as a Markov process and that experienced listeners possess internalised models of that process. The degree to which hypothetical meanings provide ambiguous expectations about consequent structures can be measured by entropy (or uncertainty): "The lower the probability of a particular consequent . . . the greater the uncertainty (and information) involved in the antecedent-consequent relation" (Meyer, 1957, p. ?). An unexpected consequent conveys a maximum of information. The process of revaluation corresponds to the feedback of information such that future behaviour is conditioned by the results of past events.

Meyer notes that uncertainty may arise from different sources. Thus, *systemic* uncertainty decreases throughout the experience of a piece of music as the listener's model develops and the composer may deliberately introduce *designed* uncertainty to combat this effect. Furthermore, the *redundancy* (lack of uncertainty) inherent in a style serves to combat noise be it cultural (resulting from discrepancies between the habit responses of a given listener's and those operating in the style) or acoustical. In making a similar distinction between *perceptual uncertainty* (that which is relative to a particular listener's model) and *stylistic uncertainty* (that which is inherent in the musical style), Witten *et al.* (1994, p. 71) warn that "it is not possible to distinguish between these measures in practice and it is a matter of some debate whether the distinctions hold even in principle." Other difficulties that arise with the statistical modelling of musical styles are discussed by Meyer (1957) and Meyer (1967, ch. 10). First, it is necessary to model different hierarchical levels of structure in music and to use different modelling strategies for different parts of a piece of music (e.g., the development section of a sonata as compared to the exposition). Second, it is likely that not all musical structure is determined by frequency but rather on human perceptual processes and, as a consequence, it is necessary to design representation schemes that have musically and psychologically meaningful interpretations.

In his later work, Meyer (1973) conducted more detailed analyses of the *melodic structures* or *processes* in Western tonal music which give rise to more or less specific expectations in the listener. Meyer describes a number of such melodic processes. A *linear pattern*, for example, consists of a diatonic scale, a chromatic scale or some mixture of the two and creates an expectation for the pattern to continue in stepwise motion. A *gap-fill pattern*, on the other hand, consists of a large melodic interval (the *gap*) which creates an expectation for a succession of notes that fill the gap by presenting all or most of the notes skipped over by the gap. Other melodic processes discussed by Meyer (1973) are more complex. A *changing-note pattern*, for example, is one in which the main structural tones of the pattern consist of the tonic the seventh and second scale degrees (in either order) and a return to the tonic. A *complementary pattern* is one in which a model pattern consisting of the main structural tones of a phrase is followed by a complementary model in which the direction of motion is inverted. Other melodic processes involve *Adeste Fideles patterns*, *triadic patterns* and *axial patterns*.

Rosner & Meyer (1982, 1986) have provided some experimental support for the psychological validity of such melodic processes. Rosner & Meyer (1982) taught listeners to distinguish a number of passages of Western tonal music exemplifying either a gap-fill or a changing-note pattern. The subjects were subsequently able to correctly classify new instances of the two processes. Rosner & Meyer (1986) extended these findings by demonstrating that listeners rated passages of classical and romantic music based on the same melodic process as more similar than passages based on different melodic processes. Although these results provide some evidence for the psychological validity of melodic processes, they do not provide evidence that melodic processes generate expectations for specific continuations (Schmuckler, 1989).[1]

2.2 The Implication-realisation Model

Narmour (1990, 1992) has extended Meyer's approach into a complex theory of melodic perception called the *implication-realisation* (IR) model. The model consists of two independent perceptual systems – the bottom-up and top-down systems of melodic implication. While the principles of the former are held to be automatic, unconscious and universal, the principles of the latter are held to be learned and hence culture dependent. Although the model is presented in a highly analytic manner, it has psychological relevance because it presents hypotheses about general perceptual principles which are precisely and quantitatively specified and therefore amenable to empirical investigation (Krumhansl, 1995a; Schellenberg, 1996). The IR model has been the subject of several detailed reviews published in the psychological

[1]See von Hippel (2000) for a re-analysis of the data obtained by Rosner & Meyer (1982, 1986) suggesting that gap-fill plays little or no role in the classification tasks

and musicological literature Cross (1995); Krumhansl (1995a); Thompson (1996) and the present summary is derived largely from these.

In the bottom-up system, intervals vary in the degree of *closure* that they convey which depends on a number of properties of the interval. The following properties contribute to the closure of an interval: 1) a rest following the interval; 2) the second tone has greater duration than the first; 3) the second tone occurs in a stronger metrical position than the first; 4) when the second tone is more stable (less dissonant) in the established key or mode than the first; 5) when three successive tones create a large interval followed by a smaller interval; and 6) when registral direction changes between the two intervals described by three successive tones. Narmour (1990) provides rules for evaluating the influence of each condition of closure. While strong closure signifies the termination of ongoing melodic structure, an interval which is unclosed is said to be an *implicative interval* and generates expectancies for the following interval which is termed the *realised interval*. The expectancies generated by implicative intervals are described by Narmour (1990) in terms of five principles of continuation which are are influenced by the Gestalt principles of proximity, similarity, good continuation and symmetry. In particular, according to the model, small melodic intervals imply a *process* (the realised interval is in the same direction as the implicative interval and will be similar in size) and large melodic intervals imply a *reversal* (the realised interval is in a different direction to the implicative interval and is smaller in size).

In an influential summary of the IR model, Krumhansl (1995a) describes the five principles as follows. Some of these principles operate differently for small and large intervals which are defined by Narmour (1990) to be those of five semitones or less and seven semitones or more respectively. The tritone is considered as a threshold interval which may function as small or large depending on the context.

Registral direction states that small intervals imply continuations in the same registral direction whereas large intervals imply a change in registral direction (cf. the gap-fill process of Meyer 1973). The application of the principle to small intervals is related to the Gestalt principle of good continuation while its application to large intervals is related to the principle of symmetry.

Intervallic difference states that small intervals imply a subsequent interval that is similar in size (± 2 semitones if registral direction changes and (± 3 semitones otherwise), while large intervals imply a consequent interval that is smaller in size (at least three semitones smaller if registral direction changes and four semitones smaller otherwise). This principle can be taken to be an application of the Gestalt principles of similarity and proximity for small and large intervals respectively.

Registral return is a general implication for a return to the pitch region (± 2 semitones) of the first tone of an implicative interval in cases where the realised interval reverses the registral direction of the implicative interval. The strength of the implication is graded according to the distance of the realised tone from the first tone of the implicative interval. This principle can be viewed as an application of the Gestalt principle of proximity.

Proximity describes a general implication for small intervals (five semitones or less) between any two tones. The implication is graded according to the absolute size of the interval. This principle can be viewed as an application of the Gestalt principle of proximity.

Closure occurs when there is a change in registral direction, movement to a smaller sized interval or both. The first condition is related to the Gestalt principle of symmetry. Degrees of closure exits corresponding to the satisfaction of both, one or neither of the conditions.[2]

[2]Note that the principle of closure specifies the combinations of implicative and realised intervals that generate closure (defined above) which determines when expectancies are weak.

Basic melodic structure	Implicative interval size	Direction of realised cf. implicative interval	Size of realised cf. implicative interval	Satisfies Registral Direction	Satisfies Intervallic Difference
Process, P	Small	Same	Similar	✓	✓
Intervallic Process, IP	Small	Different	Similar	X	✓
Registral Process, VP	Small	Same	Larger	✓	X
Retrospective Reversal, (R)	Small	Different	Smaller	X	X
Retrospective Intervallic Reversal, (IR)	Small	Same	Smaller	✓	X
Retrospective Registral Reversal, (VR)	Small	Different	Larger	X	X
Reversal, R	Large	Same	Smaller	✓	✓
Intervallic Reversal, IR	Large	Different	Smaller	X	✓
Registral Reversal, VR	Large	Same	Larger	✓	X
Retrospective Process, (P)	Large	Different	Similar	X	X
Retrospective Intervallic Process, (IP)	Large	Same	Similar	✓	X
Retrospective Registral Process, (VP)	Large	Different	Larger	X	X

Table 1: Definition of the basic melodic structures

In this encoding, the first three principles (registral direction, intervallic difference and registral return) assume binary values while the final two (proximity and closure) are graded (Krumhansl, 1995a). The model also predicts that larger intervals carry stronger implications than smaller intervals. Although the principles can be related to the Gestalt principles, they are parameterised and quantified in a manner specific to music.

Narmour (1990) uses the principles of registral direction and intervallic difference to derive a complete set of 12 basic melodic structures each consisting of an implicative and a realised interval. These structures are described in Table 1 according to the direction of the realised interval relative to the implicative interval (same or different), the size of the realised interval relative to the implicative interval (larger, similar or smaller) and the size of the implicative interval (large or small). The *retrospective* structures are so-called because, although they differ in terms of the size of the implicative interval, they have the same basic shape and are heard in retrospect as variants of the corresponding *prospective structures*. While prospective interpretations of implications occur when the implied realisation actually occurs, retrospective interpretations occur when the implications are denied. The strength of the implications generated by each basic melodic structure depends on the degrees to which it satisfies either or both of the principles of registral direction and intervallic difference. Narmour (1992) presents a detailed analysis of how these melodic structures combine together to form longer and more complex melodic structures in the IR model. Three or more consecutive structures form a *chain*. Structures are combined in three ways depending of the closure of the first structure: if there is sufficient closure the structure will be *separated* from the subsequent structure (they share a tone); if closure is weak or suppressed, the structure will be *combined* with the subsequent structure (they share an interval); and finally, one structure may also be *embedded* in another. Chaining is encouraged by weak closure as measured by one or more of its contributing factors.

Another way in which the IR model addresses more complex melodic structure is through the emergence of higher hierarchical levels of structure when strong closure exists at lower levels. Structural tones (those beginning or ending a melodic structure, combination or chain) which are emphasised by strong closure at one level are said to *transform* to the higher level. According to the model, the same bottom-up principles of implication operate on sequences of (possibly non-contiguous) tones at higher transformational levels and, theoretically, a tone may be transformed to any number of higher levels. According to the model, transformed tones may retain some of the registral implications of the lower level – an example of the primacy of the bottom-up aspects of the model. Krumhansl (1997) has found some empirical support for the psychological validity of higher level implications in a perceptual paradigm with specially constructed melodic sequences. It is also important to note that although Narmour restricts his attention to the *parametric scales* of registral direction and melodic interval size, it is held that the bottom-up principles of implication apply to any parametric scales such as duration, metrical position, timbre and dissonance.[3]

The IR model also stresses the importance of top-down influences on melodic expectancy which are learned and therefore specific to musical cultures and dependent on musical knowledge and experience. These influences include both *extra-opus* knowledge about style specific norms, such as diatonic interpretations, tonal and metrical hierarchies and harmonic progressions, and *intra-opus* knowledge about a aspects of a particular composition such as distinctive motivic and rhythmic patterns. Bharucha (1987) makes a similar distinction between *schematic* and *veridical* expectancies: while the former are influenced by schematic representations of typical musical relationships acquired through extensive exposure to a style, the latter are aroused by the activation of memory traces for specific pieces or prior knowledge for what is to come. Since the top-down system is largely independent from the bottom-up system, it may generate implications which conflict with those generated by the bottom-up system.

2.3 Sequential Pattern Induction and Extrapolation

Other models of musical expectancy based on the Gestalt principles assume that listeners use pattern induction processes to develop expectancies for successive events in melodies (Deutsch & Feroe, 1981; Simon & Sumner, 1968). These models attempt to define formal languages for describing the patterns perceived by humans in temporal sequences (such as music) and use them to explain how these patterns are used for prediction. Simon & Sumner (1968) begin with the notion of an *alphabet* for representing the range of possible values for a particular musical dimension. An alphabet is an ordered set of symbols whose order is circular and operations may be applied to an alphabet to derive a more abstract alphabet consisting of a subset of the former alphabet. Simon & Sumner restrict their attention to the dimensions of melody, harmony, rhythm and form and use alphabets for diatonic notes, triads, duration, stress and formal structure.

Assuming that structure in sequences arises principally through periodic repetition of structural elements, two operations are invoked for describing sequential patterns:

SAME : the subsequent symbol is identical to the previous one;

NEXT : the subsequent symbol is obtained by taking the next symbol in a specified alphabet a specified number of times.

Given an alphabet a pattern may be described in terms of a sequence of these operations. Such patterns may be further generalised by defining abstract symbols to replace whole subpatterns. Phrase structure is handled in the pattern language by identifying grouping boundaries and forming hierarchies of groups segmented by boundaries of varying strength. The pattern language is also extended to the description of voiced music and multiple dimensions of musical structure. Such multidimensionality is handled by defining several alphabets for the relevant dimensions (e.g., pitch, harmony, duration and so on) and describing patterns independently for each of these alphabets. Variation is handled in two ways in the pattern language: first, through the differential assignment of compound symbols to different attributes of a sequence (e.g., we can vary pitch while maintaining constant rhythm); and second, through the definition of musical relations in relative terms (using the SAME and NEXT operators) the same pattern may be applied to different sequences simply by changing the initial element or the entire alphabet.

Simon & Sumner (1968) suggest that when presented with sequences, we first induce a pattern consistent with it and secondly use that pattern to extrapolate the sequence. They hypothesise that rhythmic and harmonic periodicities are important cues that are used in the segmentation of sequences and the induction of patterns.

[3] The status of these aspects of melody in the model is criticised by some reviewers (Cross, 1995; Thompson, 1996).

Deutsch & Feroe (1981) have extended the model of Simon & Sumner in several ways. First, they extend the set of *elementary operators* to include a *predecessor* operator. Second, they define *structures* as sequences of the elementary operators which contain exactly reference point for the other operators. Finally, a *sequence* is fully specified by a structure and an alphabet. Compound sequences may be formed through the application of a *sequence operator* to two or more sequences. The simplest sequence operator, *prime*, simply recursively expands the sequences in order and two others, *retrograde* and *inversion*, are defined in terms of the prime sequence operator. A third operator defined by Deutsch & Feroe is the *alternation* operator. A string of sequence operators may be applied to a string of operators or compound operators.

Deutsch & Feroe (1981) use this pattern language to define various overlearned schematic collections of notes in terms of alphabets derived from the chromatic scale (defined in absolute terms). The derivation is achieved by the recursive application of different structures, sequences and reference elements. These alphabets include the major and (natural, harmonic and melodic) minor scales, major, minor and diminished triads and seventh chords. These embedded alphabets not held to be hardwired but are described in terms of schema that are acquired through extensive exposure to music in a given style (the examples being taken from Western tonal music). The use of a few overlearned and hierarchically embedded alphabets to define these alphabets is motivated in terms of parsimony of encoding and constraints on memory and processing.[4] In particular, four processing advantages are presented:

1. reduced redundancy of representation;

2. distinct alphabets may be invoked at different levels;

3. embedded sequence structures and their associated alphabets may be encoded as chunks;

4. the chunking of structures allows for the representation of configurations that satisfy proximity and the differentiation of different members of the alphabet in terms of frequency.

Several music examples are discussed in terms of these pattern languages by Simon & Sumner (1968) and Deutsch & Feroe (1981).

According to the model, a melodic sequence is perceived and memorised in terms of a number of sequence structures and overlearned alphabets which run in parallel at a number of hierarchical levels and that the generation of a sequence occurs in a top-down fashion. Since structures are expanded in a top-down manner, the model predicts that notes occurring in the highest level should be most salient since they also occur in all the subsequently processed lower levels and the notes occurring only in the lowest levels should be least salient. The initial groupings of notes which allows hierarchical structure to emerge is taken to rest on simple perceptual mechanisms such as temporal and registral proximity. While large differences in such dimensions may prevent the integration of key elements across the boundary and the formation of higher-order linkages across them, moderate differences function as segmentation points. Finally, multiple representations may be formed by listeners who, according to the model, will tend to choose the most parsimonious. Deutsch & Feroe (1981) propose that the acquisition of a representation is an ongoing process of generation, confirmation and violation of multiple structural representations (cf. Meyer, 1973; Narmour, 1990).

We turn now to experimental support for the model. The psychological relevance of pitch hierarchies has been demonstrated by Krumhansl (Krumhansl, 1979; Krumhansl & Kessler, 1982; Krumhansl & Shepard, 1979). Regarding the importance of such hierarchies in the perception of melodies, Deutsch (1980) describes experiments in which subjects were presented with sequences of 12 notes which they

[4] The same principles are taken to apply to harmonic progressions.

recalled in musical notation and which support this aspect of the model. Half the sequences were structured in accordance with the model such that a higher level subsequence of four elements acted on a lower level subsequence of three or four elements while the remaining sequences were unstructured. Sequences were presented in one of three conditions: first, with no temporal segmentation; second, temporally segmented in accordance with tonal structure; and third, temporally segmented in violation of tonal structure. The results demonstrated that recall was high in the first and second conditions and low in the third condition and for unstructured sequences. However, on a similar task, Boltz & Jones (1986) have demonstrated that rule recursion has only a modest effect on memory for melodies and only in certain conditions.[5]

2.4 The Theory of Dynamic Attending

In contrast to theories such as those of Deutsch & Feroe (1981) in which rules are applied statically and recursively to produce efficient representations of music, the *theory of dynamic attending* (Jones, 1981, 1982, 1987, 1990; Jones & Boltz, 1989) proposes that people use tempo, metre and rhythm to dynamically guide attention to salient events over time. In contrast to the IR model of Narmour (1990, 1992), expectancies suggest not only *what* will happen but also predict *how* and *when* it will occur. Expectancies are conceived as vectors guiding attention towards future pitch-time regions and which are activated by the current musical context. An expectancy scheme consists of the set of expectancies that are activated in a given musical context. The theory contains two aspects: the first describes the abstract representations of musical structure upon which musical expectancies are founded; and the second describes the manner in which listeners employ these abstractions to generate expectations (Jones, 1990).

Abstraction refers to the contextual regularities that generate expectancies and expectancy schemes. In this regard, Jones (1981) draws a distinction between *ideal prototypes* and *ordinary patterns*. The former are culturally determined schema which combine simple rules to summarise the conventions of a particular musical genre and period. They are rarely found in actual music. The alphabets described by Deutsch & Feroe (1981) are good examples of ideal prototypes. Ordinary patterns, by contrast, are found in real-world music and, although they may be related to one or more ideal prototypes, they exhibit interesting deviations from perfect symmetry. Thus ordinary patterns can be objectively specified in terms of the ideal prototypes that they invoke and the artful deviations from these prototypes that they contain.

Jones (1981) discusses three types of symmetry which are present in ideal prototypes. The first is *tonal-harmonic symmetry* which has to do with musical scales and their associated chordal harmonies and which obeys octave equivalence. The second is *melodic symmetry* which has to do with the relations between successive notes in time. These symmetries serve to preserve melodic form and may be based on constraints on pitch contour, melodic pitch interval or tonal-harmonic symmetries where each of this factors may influence the perception of melodic accents (Jones, 1987). For example, reversal of contour, large melodic intervals and tones that are dominant in the prevailing tonal hierarchy may all contribute to the perception of melodic accents. The final kind of symmetry is *time symmetry* which describes metric, rhythmic and tempo regularities in ideal prototypes. Thus metrical hierarchies, relatively long or short notes, rests and ratio based tempo changes may all contribute to the perception of temporal accents (Jones, 1987). Musically, melodic symmetries are defined partly in terms of the stable background provided by tonal symmetries and time symmetries operate in conjunction with melodic symmetries. Jones (1987) discusses the integration of melodic and temporal accents into a *joint accent structure* which arises through the coincidences of melodic and temporal accents in time. Joint accent structure

[5]One criticism of all this research is that the pitch sequences were artificially constructed to conform to or violate the theoretical predictions (Schellenberg, 1996).

results in a range of accent strength differences which serve to emphasise hierarchical structures in ideal prototypes.

We turn now to the generation of expectancies. An expectancy scheme is a temporally based mental set derived from one or more ideal prototypes. These ideal prototypes are activated through synchronisation with corresponding portions of the ordinary patterns present in a piece of music (Jones, 1982). Expectancy schemes serve a preparatory function and their complexity, specificity and appropriateness in a given musical context depend on the musical skill and familiarity with the genre of the listener. Instances of divergence between the ordinary pattern and the ideal prototype that it activates induce "surprise" in the listener and are crucial for the perception and retention of musical sequences. Jones (1982) presents a model of how listeners cope with the integration of expected and observed event in such circumstances. The expected event defines a *perceptual reference frame* within which the observed event is evaluated. A *harmonic* reference frame is closely associated with tonal centre and a *melodic* reference frame arises from the relation of the expected event to previous events.

Jones (1982, 1990) describes several ways in which unexpected events are integrated with expectancy schemes based on ideal prototypes. The first, the *serial integration region hypothesis*, proposes a two stage process: first, the unexpected event is picked-up; and second, it is hierarchically integrated. In the first phase, the unexpected event is evaluated in terms of its deviance from the expected event. In the second phase, the difference between expected and observed events is used to correct the original expectancy scheme. The hypothesis predicts that large differences between expected and observed events will facilitate pickup but impede integration and vice versa. The serial integration hypothesis applies primarily to local melodic changes and single expectancy vectors in which pitch intervals are small and similar. Large or different pitch intervals will lead to the perception of accents and the differentiation of ordinary patterns and the termination of expectancy schemes (Jones, 1990). At intermediate time spans, the guiding principle is the *identity under transformation principle* which states that the speed and accuracy with which a relationship between two melodic segments is abstracted is related to the number of melodic transformations required to change one into the other. Relationships between melodic phrases is influenced by relationships in rhythmic patterning, interval structure and other abstractions and by operations such as identity, transposition, complement, inversion and so on. These relationships have much in common with those described by Meyer (1973) and Narmour (1990, 1992). Finally, over large time spans, expectancy vectors are coordinated by the *melody/rhythm accent hypothesis* which states that the temporal hierarchies derived from joint accent structure guide attending more or less effectively over phrases and groups of phrases. Thus attending will be more effective when simple and consistent temporal, melodic and metrical hierarchies can be established by the listener.

Experimental support for various aspects of the theory of dynamic attending has come from a variety of sources but in particular studies involving melody recognition or reproduction tasks (see Jones, 1982, 1987, 1990, for reviews).

3 Experimental Approaches

3.1 Overview

This section contains a review of experimental studies of human expectancies in melodic music. We focus on experimental paradigms in which subjects are required to provide explicit data on their expectancies either by rating supplied continuations (perceptual paradigms) or by generating continuations themselves (generative paradigms). The majority of studies that have used these paradigms have focused on expectancies regarding pitch. Expectancies concerning temporal structure, including onsets, rests and durations, have typically been studied indirectly in recognition, reproduction or time estimation tasks. Palmer & Krumhansl (1990), however, have investigated the perceptual validity of metrical hierarchies using goodness-of-fit ratings obtained in psychological experiments. Subjects were asked to rate a series of probe tones inserted at various temporal locations into a metrical context based on a variety of different time signatures. The responses demonstrated that listeners represent multiple temporal periodicities which are sensitive to the time signature and which coincide with music-theoretic predictions. Furthermore, the depth of the hierarchy tended to increase with musical training.

3.2 Perceptual Paradigms

Schmuckler (1989, experiment 1) employed an experimental paradigm in which 32 musically trained subjects were asked to listen to melodic contexts with a variable continuation tone and make judgements about how well the continuation fitted with their expectations. The contexts were ten melodic extracts from the first eight bars of Robert Schumann's "Du Ring an meinem Finger" from the song cycle *Frauenliebe und Leben* (Op. 42).[6] The extracts were selected such that the continuations occurred at points of interesting melodic progression and each ran from the beginning of the melody to the selected point. The subjects initially listened to the contexts and provided ratings for how strong and specific their expectations were on seven-point scales. The subjects then listened to each of the contexts followed by a continuation tone and provided judgements of how well the continuation agreed with their expectations on a seven-point scale. The continuation tones ranged from five semitones below to six semitones above the actual completion, were of equal duration to the actual continuation and were presented in a random order. Schmuckler hypothesised that melodic expectations would depend on three factors: first, tonal structure; second, melodic contour; and third, melodic process (Meyer, 1973). In order to test the first of these hypotheses, Schmuckler performed correlations of the expectancy profiles with previously reported data on the perceived stability of chromatic tones in reference to a tonal context (Krumhansl & Kessler, 1982). Data were used for the tonal contexts of E♭ major (the key of the piece), B♭ major (the dominant of E♭ major and the key in which the passage ends) and G minor (the relative minor of B♭ major). The results demonstrated "strong evidence supporting the psychological reality of tonal structure in expectancy formation" (Schmuckler, 1989, p. 123). The results of a contour analysis (based on the two tones occurring immediately prior to the continuation) were ambiguous with some contexts generating stronger expectancies for change in contour and others stronger expectancies for continuing contour. Schmuckler (1989) attributes these limited effects of contour to the local manner in which direction of movement was derived and the strong influence of tonal structure. Perhaps the most interesting finding in this study was that the continuation tones predicted on the basis of linear and gap-fill melodic processes (see §2.1) received significantly higher expectancy ratings than the remaining diatonic tones providing strong evidence for the psychological reality of melodic processes in expectancy formation. Finally, Schmuckler (1989) found that expectancy strength and specificity ratings were highly correlated. In order to examine whether expectancy strength and specificity were related to the expectancy profiles, Schmuckler computed a difference score for each context by subtracting the average ratings for all 12 continuations from the highest rating given to any continuation. These difference scores showed a marginally significant correlation with the average of the strength and specificity ratings.

Krumhansl (1995a) (see also Schellenberg 1996) reports a similar series of three experiments designed to test the psychological validity of the bottom-up principles of the IR model. The first experiment used eight melodic fragments taken from British folk songs, diatonic continuation tones and twenty subjects of whom 10 were musically trained and 10 untrained. The second experiment used eight extracts from Webern's *Lieder* (Opus 3, 4 and 15), chromatic continuation tones and 26 subjects generally unfamiliar with the atonal style of whom 13 were musically trained and 13 untrained. The third experiment used 12 melodic fragments from Chinese folk songs, pentatonic continuation tones and 16 subjects of

[6] Schmuckler (1989) reports three experiments with the same piece of music which examined respectively melodic, harmonic and full context expectancies. Here we are concerned only with the experiment on melodic expectancy.

whom 8 were Chinese and 8 American. All melodic fragments were chosen such that they ended on an implicative interval and all continuation tones were within a two octave range centred on the final tone of the melodic context. There was high inter-subject correlation for all experiments (except for four untrained subjects in the second experiment) warranting the averaging of the data for further analysis. The degree of consistency across subjects suggests that patterns of musical expectancy do not depend on specialised training, experience with a style or knowledge of technical concepts or vocabulary. A multiple regression analysis of the data was conducted using as predictors the five IR principles as well as a tonality predictor based on the data of Krumhansl & Kessler (1982). The overall model correlated significantly with the data with each predictor making a significant contribution for all three experiments (with the exception of intervallic difference for the second experiment). Overall, the weakest contribution was made by intervallic difference and the strongest by proximity. Once again, there was no difference in the regression coefficients computed for listeners with moderate and limited training (Schellenberg, 1996). Finally, Krumhansl (1995a) examined whether the results provided any support for the differential degrees of surprise predicted by the IR model for different basic melodic structures. High average ratings of goodness-of-fit would be expected to correspond to low levels of surprise and vice versa. The results of this analysis conformed quite well to the predictions of the IR model. There were, however, several notable exceptions which lead Krumhansl (1995a) to suggest that expectancies depend not only on registral direction and intervallic difference (which define the basic melodic structures) but also the principles of registral return and closure.

Cuddy & Lunny (1995) have provided convergent support for the bottom-up principles of the IR model in a similar task using only two-tone intervals as the melodic contexts. Twelve musically trained and 12 musically untrained subjects were asked to rate how well a test tone continued a two-tone context on a seven-point scale. Eight implicative contexts were used consisting of ascending and descending intervals of a major second, a minor third, a major sixth and a minor seventh for which the second tones of each interval were either C4 or F♯4. Continuation tones consisted of all 25 tones from one octave below to one octave above the second tone of the implicative context. An analysis of variance with the factors musical training, implicative interval and continuation tone failed to reveal a significant effect of training but did reveal a significant interaction between implicative interval and continuation tone as hypothesised. A multiple regression analysis using the five implicative principles as predictors, accounted for 29% of the variance of the mean ratings and registral return and proximity emerged as significant predictors while the other three principles approached significance. A second multiple regression analysis considered intervallic difference, proximity, registral return, a revised version of registral direction which only pertained to large intervals and a predictor for pitch height, based on the observation that ratings tended to increase as as the pitch height of the continuation tone increased. Also included were two tonality predictors. The first, *tonal strength* was based on the assumption that the rating of a continuation tone would be influenced by the degree to which the pattern of three tones suggested a tonality.[7] The second tonality predictor, *tonal region*, was derived by listing all possible major and minor keys in which each implicative interval could occur and coding each continuation tone according to whether it represented a possible tonic for the preceding interval. These tonality predictors replaced tonal hierarchy without loss of predictive power and the resulting model accounted for 64% of the variance in the ratings with all predictors making significant contributions.

Krumhansl (1995b) repeated the study of Cuddy & Lunny (1995) with sixteen musically trained subjects using a more complete set of two-tone contexts ranging from a descending major seventh to an ascending major seventh. The aim was to address a number of issues raised by the previous studies: the first concerns mutual incompatibility between the principles (e.g., for most small implicative intervals, a realised interval cannot simultaneously satisfy registral direction and closure); the second concerns the

[7] The key-finding algorithm developed by Krumhansl and Schmuckler (Krumhansl, 1990) was used to rate each of the patterns for tonal strength.

independence of the principles (e.g., in some contexts the principle of closure make the same predictions as the principle of intervallic difference); the third, concerns the possibility of deriving simpler models with fewer predictive variables; the fourth concerns the development of a more parsimonious model of the top-down effects of tonality on expectation. A series of multiple regression analyses supported the inclusion of modified versions of proximity, registral return and registral direction but not closure or intervallic difference. In particular, the results supported a modification of proximity such that it is linearly graded over the entire range of intervals used and a modification of registral return such that it varies as a linear function of the proximity of the third tone to the first. Finally, the principle of registral direction was supported by the results except for the data for the major seventh which carried strong implications for octave completion (see also Carlsen, 1981). Krumhansl (1995b) notes that the three remaining principles are both logically and statistically independent of one another. Krumhansl (1995b) also examined the effects of predictors for top-down effects of tonality and consonance on the data modelling. Regarding the former, the tonal region predictor developed by Cuddy & Lunny (1995) was extended by averaging the tonal hierarchy data (Krumhansl & Kessler, 1982) for all keys in which each interval was diatonic. The resulting variable correlated strongly with the data for all intervals bar the two tritones. No support was found for the tonal strength variable of Cuddy & Lunny (1995). Predictors were also added for (octave equivalent) consonance with the first and second tones of the context and these were found to correlate significantly with the data.

Schellenberg (1996) conducted a reanalysis of the experimental data first reported by Krumhansl (1995a) and concluded that the unique contributions of each predictor to the fit of the model revealed a high degree of overlap between the predictors. On the basis of this redundancy, Schellenberg derived a revised model containing only four predictors: the first was the revised version of registral direction of Cuddy & Lunny (1995); the second was registral return; the third was a revised version of proximity such that it varied indefinitely with the size of the interval; and the fourth was the tonality predictor of Krumhansl (1995a). This revised model significantly reduced the redundancy of the model while accounting for approximately 75% of the variance in the averaged responses and was consistent across training levels. On the basis of this experiment, Schellenberg (1996) that the model is over-specified and may be simplified without loss of predictive power. Schellenberg (1997) obtained further support for this conclusion in a reanalysis of the data reported by Cuddy & Lunny (1995). The revised IR model still exhibited significant correlation between the revised principles of registral direction and registral return reflecting the fact that for large implicative intervals, instances that satisfy the revised registral return principle involve a change of direction, thereby satisfying the revised registral direction principle as well, and suggesting that the model can be simplified further still (Schellenberg, 1996). Schellenberg (1997) applied principal components analysis to both the original IR model and the revised model. The latter experiment inspired the development of a model consisting of two factors: *pitch proximity* and *pitch reversal*. The former states that an implicative interval creates expectations whose strength decreases with the size of the realised interval. The latter is essentially a linear combination of the revised principles of registral direction and registral return and is roughly characterised as the implication for a change in register conditional on the size of the implicative interval. The model is considerably simpler than Schellenberg's revised model and yet does not compromise the predictive power of that model in accounting for the data obtained by Krumhansl (1995a) and Cuddy & Lunny (1995).

Krumhansl *et al.* (1999) have studied melodic expectations in the context of Finnish spiritual folk hymns. Twenty six subjects, of which 12 were experts (members of a Finnish religious prayer movement) and 14 were non-experts (Finnish musicology undergraduates), were asked to listen to eight melodic contexts transposed to C major or C minor and a variable continuation tone and to judge how well the continuation tone fitted with their expectations on a scale of 1 to 7. The results exhibited strong correlations both within and between the two groups. A multiple regression analysis of the data was conducted for each group using four models: the first includes all five principles of the IR model as coded by Krumhansl (1995a); the second, includes the five principles as coded by Krumhansl (1995b) with proximity and registral return coded as linearly decreasing functions of interval size and two consonance variables coding the consonance of the continuation tone with the last and penultimate tones of the context respectively; the third, was the revised model of Schellenberg (1996); and the fourth was the two variable model of Schellenberg (1997). The results of the multiple regression analysis indicated that the second model significantly outperformed the other models with each predictor, bar registral direction and closure, showing a significant correlation with the data of both expert and non-expert groups. Krumhansl *et al.* (1999) conclude that it may be premature to reduce or revise the bottom-up principles of implication until additional musical styles have been examined.

In a further analysis of the data, Krumhansl *et al.* (1999) sought to examine the influence of top-down factors on expectation and, in particular, to distinguish between schematic and veridical expectations using the following variables: the two-tone goodness-of-fit data obtained by Krumhansl (1995b), the major and minor key-profiles Krumhansl & Kessler (1982), and monogram, bigram and trigram distributions of tones in the entire corpus of 18 spiritual folk hymns and the correct continuation tone. The results demonstrated that the correlation with schematic variables such as the two-tone goodness-of-fit data and major and minor key profiles were stronger for the non-experts than for the experts. By contrast, the correlation with veridical variables such as the correct next continuation tone and trigram distributions was stronger for the experts than the non-experts. When these top-down variables were included in the multiple regression analysis the resulting correlation was significantly higher for the experts than the non-experts. The reason was that while both groups had significant contributions of both key profiles and monogram statistics only the expert group had significant contributions from the higher-order transitions and the correct continuation tone variable. Krumhansl *et al.* (1999) conclude from these results that whereas stylistic familiarity has no effect on bottom-up principles of melodic expectancy, it does influence top-down principles by reducing the effects of schematic knowledge of Western tonality and increasing the veridical effects of structural and statistical knowledge of the style.[8]

Eerola *et al.* (2002) criticise the goodness-of-fit perceptual paradigms discussed for using limited contexts, only probing expectations at one point in time and notes the possibility of distortion due to repetition of the context sequence. Such paradigms cannot be applied to the study of the manner in which expectations dynamically change while listening to a melodic composition. To address these limitations, Eerola *et al.* designed a methodology in which subjects were asked to listen to complete melodies and provide continuous judgements on a mouse-controlled slider of the perceived predictability of the melody. Twenty five music students took part in the experiment and the stimuli were composed of 30 tonal folk songs from the Essen Folk Song Collection (Schaffrath, 1992, 1994), 10 tonally ambivalent melodies from the music of Charles Ives and 27 artificially generated isochronous sequences. The listener ratings were sampled at 100ms intervals. The results exhibited strong intersubject correlations and were averaged over subjects. The listener's ratings were predicted with three models: first, Narmour's five principles of melodic implication supplemented by a predictor for tonal stability based on the data obtained by Krumhansl & Kessler (1982) and a predictor for melodic anchoring which increases the stability of a tone if it is anchored to adjacent, proximate and more stable tones; second, the per-event entropy estimated from a bigram model whose initial state was derived from the entire Essen Folk Song Collection; and finally, the per-event entropy of the distribution of pitch classes within a sliding window where the initial state of the sliding window model was derived from the key profiles of Krumhansl & Kessler (1982). Initial results suggested that the third model yielded the strongest correlations with the human data.

Another problem with the perceptual paradigms so far discussed is that all have required subjects to

[8] Krumhansl *et al.* (2000) have applied the same methodology to indigenous folk melodies (yoiks) of the Sami people of Scandinavia and arrived at largely the same conclusions with the exception of a shift in relative importance from proximity in the folk hymns to consonance in the yoiks.

provide explicit ratings of continuation tones as the dependent variable while their expectations need not necessarily be explicit (Meyer, 1956). Aarden (2002) has used a different methodology that uses *reaction time* to judgements of contour continuations as the dependent variable on the hypothesis that expectations help in planning responses to events such that accurate expectations facilitate faster responses. Aarden (2002) argues that reaction times are preferable to goodness-of-fit judgements for measuring expectancies because they are less dependent on conscious introspection. In Aarden's experiments, 27 music students were asked to listen to melodic phrases from the Essen Folk Song Collection (Schaffrath, 1992, 1994) of between eight and 20 notes and for each note indicate whether the interval formed with the previous note was rising, falling or a unison. A multiple linear regression analysis was conducted on the results with four predictors: pitch proximity and pitch proximity Schellenberg (1997), tonal hierarchy (Krumhansl & Kessler, 1982) and a dummy variable indicating whether the tone occurred on a pulse in the highest two levels of any bar-length metric hierarchy. Pitch proximity and tonal hierarchy were significant predictors while pitch proximity and metre approached significance.

In a second linear regression analysis, pitch proximity and pitch reversal were used as predictors along with a third predictor called *step momentum* which states that small pitch intervals imply a continuation in the same registral direction. The addition of step momentum to the model had a significant impact on the predictive power of the model. Surprisingly, adding the tonal hierarchy predictor did not improve model performance. Aarden (2002) suggested that the key profiles derived by Krumhansl & Kessler (1982) might apply only to the distribution of pitches at the end of phrases. Accordingly they found that while a strong correlation existed between the frequency of tones ending phrases in the Essen Folk Song Collection and the Krumhansl-Kessler key profiles, a predictor based on a monogram analysis of the pitches in the entire Essen Folk Song Collection did prove to be a significant predictor. Aarden (2002) concludes that while the goodness-of-fit studies (e.g., Cuddy & Lunny, 1995) measure the expectancy for pitches implied by closure, the reaction-time method engages the listener in expectancies generated by implicative intervals which results in a flatter key profile.

3.3 Generative Paradigms

The first experimental studies of expectation in music were carried out by Carlsen (1981) who hypothesised that listener's expectancies would vary as a function of cultural milieu, amount of musical training and vocal register. In order to test these hypotheses, Carlsen designed an experiment in which subjects were asked to listen to sets of melodic beginnings (consisting of two tones) and sing the expected completion of the melody. The subjects were 91 music students from Germany, Hungary and the USA who were screened to ensure an ability to conceive a melody and sing it accurately. The subjects were divided into two groups reflecting levels of musical training and all four vocal registers were represented from each country and training level. The stimuli consisted of the 25 two-pitch intervals found within the equal tempered 12-tone octave (12 ascending, 12 descending and the unison) and 15 sets of these beginning were constructed such that no two melodic beginnings would be in the same contiguous relationship in any two sets. The 15 sets were recorded in each vocal register such that the second tone of the melodic beginning occurred in the mid range of the register.

For the purposes of analysis, Carlsen (1981) examined only the first pitch sung after the melodic beginnings. The results demonstrated that the different melodic beginnings generated different expectancies but also that they did so with varying strength and specificity (although no context produced unanimous results for a single continuation). For example, an ascending major second elicited a strong expectancy for another ascending major second (64% of cases) while a descending octave elicited 11 different expectancies with greater than chance frequency. In general, small response intervals (of three semitones or less) were more frequent than large intervals. Regarding the hypothesised determinants of expectancy, the results demonstrated a significant effect of cultural milieu, but not of musical training or vocal register. Carlsen (1981) provided explanations of many of the expectation results in terms of completion of schematic patterns such as major and minor triads, completion of the octave, return to the first pitch of the melodic beginning and a unison with the second. The failure to find an effect of training level and vocal register was attributed to the probable overlap in training and performance medium between the respective groups.

Unyk & Carlsen (1987) repeated the study of Carlsen (1981) with 27 American music students and obtained largely similar results. The data were used to generate six brief melodies that varied in their relationships to individual subject's expectations in terms of fulfillment of strong and weak expectancies, interval size violation of strong and weak expectancies and combined interval size and contour violation of strong and weak expectancies. These test melodies were presented aurally to the subjects for transcription. It was hypothesised that the transcription errors would be greater for melodies containing violated expectancies than for fulfilled expectancies and that this effect would be greater for strong expectancies than weak expectancies. It was also predicted that transcription errors would vary as a function of the degree of expectancy violation (i.e., no violation, interval size violation and interval size and contour violation). The results demonstrated significant effects of both expectancy fulfillment and expectancy strength. The effects of the degree of expectancy violation were only apparent in the case of strong expectancies. Unyk & Carlsen (1987) suggest that these results support the theory that listeners entertain expectancies about future events, that violations of these expectancies have a psychological effect in music perception and that this effect is dependent on the strength of the expectancies (Meyer, 1956).

Schellenberg (1996) has conducted a reanalysis the data of Carlsen (1981) and Unyk & Carlsen (1987) using multinomial log-linear analyses using the same predictor variables as for his own experimental data (see §3.2). There were strong correlations between the data from the three cultural groups in the study of Carlsen (1981) and the data obtained by Unyk & Carlsen (1987) with American subjects correlated strongly with all three set of data. The results of the multinomial log-linear analyses were largely comparable to those in Schellenberg's own perceptual experiments with the exception of the predictors for closure and tonality. While tones that did not create closure were sung more frequently in in the generative tasks, tones that created closure were judged as better continuations in the perceptual task. Schellenberg suggests that the brief contexts used by Carlsen (1981) and Unyk & Carlsen (1987) may have generated expectancies for extended continuations while the extended contexts of his own experiments might have created expectancies for more immediate closure.Tonality consistently explained more of the data in the generative task than in the perceptual task and Schellenberg (1996) suggests that this may be due to the use of chromatic continuation tones in the former and diatonic tones in the latter. Cuddy & Lunny (1995) also conducted a reanalysis of the data of Carlsen (1981) with the result that their own model provided an excellent fit to the data. In contrast to Schellenberg (1996), they found that their tonal region and tonal strength predictors had a greater influence in their own perceptual task than in the generative task. This discrepancy is likely to be due to the fact that the two reanalyses used different subsets of Carlsen's data.

Schmuckler (1989) points out that the two tone beginnings in the methodology used by Carlsen (1981) and Unyk & Carlsen (1987) might not be sufficient to generate strong expectations and that asking subjects to sing the continuations could lead to transcription errors. These issues were addressed in an experiment (Schmuckler, 1989, experiment 4) in which six pianists were provided with the scores of seven of the melodic contexts used in Experiment 1 (see §3.2) and asked to compose and/or improvise the expected melodic continuation of the extract. Subjects were given as much time as they wanted to work with the composition before performing the continuation on an electronic piano. The first note of the continuations thus obtained was compared to expectancy profiles obtained in the earlier experiment (Schmuckler, 1989, experiment 1). Of the seven extracts four of the performed continuations showed significant correlations with the expectancy profiles from experiment 1 with another achieving only marginal significance. Furthermore, there was a strong negative correlation between the length of the composed continuation and the expectation strength/specificity results obtained in Experiment 1 indicating that extracts giving rise to strong and specific expectations tended to induce shorter completions. Schmuckler (1990) reports further analyses of the continuations collected in this experiment. In particular, the continuations demonstrated strong influences of tonal hierarchies (Krumhansl & Kessler, 1982), metric hierarchies (Palmer & Krumhansl, 1990) and the metric content of the context but only moderate influence of the pitch content of the context.

Thompson *et al.* (1997) used a similar methodology to that of Carlsen (1981) and Unyk & Carlsen (1987) to test the bottom-up principles of the IR model. One hundred subjects, of whom 50 were musically trained (they had five or more years of formal music training and were currently playing a musical instrument) and 50 were untrained (they had less than five years of musical training and weren't currently playing a musical instrument), were asked to provide melodic completions of a two-tone context on a piano. Eight implicative contexts were used consisting of ascending and descending intervals of a major second, a minor third, a major sixth and a minor seventh for all of which the first tone was middle C. Thompson *et al.* (1997) hypothesised that the tones generated immediately after the context would more often realise than deny the five bottom up principles of melodic implication and other top-down principles related to tonal structure. They also hypothesised that if the principles reflect universal perceptual laws, evidence for their existence should vary as a function of musical experience. The results demonstrated that the continuations adhered to the tonal framework of C major and that with the exception of registral return for musically trained listeners, each implicative principle was fulfilled significantly more often than would be expected by chance. Furthermore, there was a strong correlation between the continuations produced by musically trained and untrained subjects.

Finally, Thompson *et al.* (1997) conducted a multinomial log linear analysis which revealed that the five implicative principles together (all had significant predictive power) accounted for 48% of the deviance in the combined results. Thompson *et al.* hypothesised that some of the remaining unexplained deviance could be accounted for by adding two principles based on implied tonal structure: *tonal hierarchy* and *diatonicism*. The former predicts that responses should reflect the tonal stability of notes in key of C major using the tonal hierarchy data of Krumhansl & Kessler (1982). The latter predicts that listeners should select melody notes that are diatonic in C major. These predictors both had significant predictive power and the improved model accounted for 79% of the variance in the combined results. Thompson *et al.* (1997) also conducted a multinomial log linear analysis of the corresponding data reported by Carlsen (1981) with the result that the improved model accounted for 63% of the deviance in responses with all predictors making a significant contribution to the model.

Manzara *et al.* (1992) have used a generative task to elicit the melodic expectations of listeners in a rather different way and for a rather different purpose. The goal of the research was to derive an estimate of the entropy of individual pieces within a style according to the predictive models used by human listeners. Such a measure could be used for several tasks including the comparative stylistic analyses of pieces within and between genres and the evaluation of computational models of prediction in monophonic music. The experiment followed a gambling methodology developed by Cover & King (1978) for estimating the entropy of printed English. Subjects interacted with a computer program that displayed the score of the piece with all notes placed on the centre of the score. Given an initial capital of $S_0 = 1.0$, the subjects were asked to move through the piece sequentially, selecting the pitch of each note and betting a proportion p of their capital on the selection repeatedly until the selected pitch was correct. At each stage n, the subjects capital was incremented by $20pS_{n-1}$ if the selection was correct and decremented by the proportion bet if it was incorrect.[9] This proportional betting scheme was designed to elicit intuitive probability estimates for the next symbol to be guessed and rewards not only the correct guess but also accurate estimates of the symbol's probability. Given this data, the cumulative average

[9] There were 20 chromatic pitches to choose from.

entropy of the style can be estimated as:

$$H = \lim_{n\to\infty} \left[\log_2 20 - \frac{1}{n} \log_2 S_n \right]$$

and the per-event entropy can be estimated as:

$$H = \log_2 20 - \log_2 S$$

where S is the capital won by the subjects on this event.

The data were two chorale soprano melodies harmonised by J.S. Bach (no. 61 and no. 151 Breitkopf edition) and the experiment was organised as a competition. Ten subjects took part in the first round with one chorale melody and the best four were selected for the second round with the second chorale melody. The overall winner of the competition won a prize. The capital data were combined in three ways:

- *average capital*: the average capital that accrues to each subjects for each event;
- *best subject*: take the capital won by the best performing subject on each event;
- *committee gambling*: compute a weighted average of accrued capitals where the weights are proportional to the subject's performance on that event.

Although these yield similar results, Manzara *et al.* (1992) suggest that the third method is the most reliable.

Although it was not possible to generalise the entropy estimates to the chorale genre in general due to the small scale of the study, it was possible to make some interesting observations about the entropy profiles derived. In particular, Manzara *et al.* (1992) found that the ultimate notes in cadences tended to be associated with lower entropy than the middle and beginning of phrases. High entropy was also associated with stylistically unusual cadential forms and intervals. The cumulative entropy profiles for both pieces showed rising entropy at the beginning of the piece due to lack of context, followed by decreasing entropy as the growing context supports more confident predictions which is in turn followed by a rise in entropy near the end of the piece before a steep decline to the final cadence. Finally, Witten *et al.* (1994) have compared the human data to that generated by a multiple viewpoint Markov model derived from 95 chorale melodies (Conklin & Witten, 1995). Although the human subjects slightly outperformed the model, the overall form of the entropy profiles were remarkably similar. The relative degrees of uncertainty elicited by events throughout the pieces was strikingly similar for both the subjects and the model.

In addition to empirical research on the melodic continuations supplied by listeners to a melodic context, there exists research which has examined theoretical models of expectation in analyses of existing repertoires of melodic music. Thompson & Stainton (1996), for example, identified implicative intervals in 50 Bach Chorales (soprano and bass voices) and 16 Schubert melodies as those in which the second tone of the interval satisfied all of the following three conditions:

1. it is equally or less tonally stable than the first tone (using the data of Krumhansl & Kessler 1982);
2. it occurs at a weaker metrical position than the first tone;
3. it has a duration equal to or less than that of the first tone.

The proportions of implicative intervals satisfying each IR principle were relatively consistent across the three data sets with large proportions of intervals satisfying intervallic difference and closure while smaller proportions satisfied the other three principles. These results are hard to interpret for two reasons: first, the principles differ in terms of the size of the sets of continuations they imply meaning that different proportions are expected by chance; and second, the principles make overlapping predictions (e.g., intervallic motion and proximity) so it is impossible to determine from this analysis which principles are responsible for the findings. To address these issues, Thompson & Stainton (1996) conducted a multinomial log-linear analysis with the five IR principles as predictor variables. The resulting model accounted for 60%, 53% and 44% of the variance in pitch of the tone following implicative intervals for the Bach chorale soprano melodies, Bach chorale bass melodies and Schubert melodies respectively. With the exception of intervallic difference for the Schubert dataset, all predictors add significant predictive power of the model.

Thompson & Stainton (1998) replicated this study with a collection of 513 Bohemian folk melodies from the Essen Folk Song Collection (Schaffrath, 1994, 1995). The goals were to assess the following hypotheses:

- the predictive power of the principles should be larger for strongly implicative intervals than for strongly closural intervals;
- the predictive power of the principles should be larger for large intervals than small intervals;
- the data can be modelled least redundantly by removing overlapping predictions between the principles;

An analysis of the frequencies with which the principles were satisfied for implicative and closural intervals demonstrated that intervallic difference, registral return and proximity were satisfied significantly more often than would be expected by chance and those satisfying registral direction, intervallic difference and proximity were lower for closural than for intervallic intervals. A multinomial log-linear analysis revealed that the relative predictive power of the principles differs for implicative and closural intervals although not always in the predicted direction (e.g., closural intervals were more likely than implicative intervals to be followed by a change in registral direction). A model consisting of the five principles as predictor variables accounted for 65% and 64% of the variance for implicative and closural intervals respectively and all predictors had significant predictive power. A comparison of results for small and large intervals indicated that the predictive power of the principles was slightly greater for large than small intervals. While the five-principle model contained a high degree of redundancy, Thompson & Stainton (1998) were able to reduce the redundancy while retaining predictive power by removing the principles of closure and intervallic difference and revising registral direction such that it applies only to large intervals. Finally, predictors for consonance and unison with the second tone of the implicative interval were also found to have significant predictive power in the model.

These studies suggest that similar patterns of expectancy emerge both from perceptual and generative tasks but also that they tend to be followed in melodic composition. This raises the question of whether the bottom-up principles of the IR model might be acquired through everyday experience of music rather than being dependent on universal, low-level perceptual constraints. von Hippel & Huron (2000) have suggested that the explanations of *post-skip reversals* given by Meyer (1973) and Narmour (1990) in terms of expectancies are incorrect. Instead, they argue that the common occurrence of these structures in music is due to the effects of *regression towards the mean* caused by tessitura. Three experiments with a variety of melodic datasets supported this conclusion. In the first, it was found that evidence for post-skip reversals is limited to those skips (intervals of three semitones or more) which cross or move away from the median pitch of a given dataset. When skips approach the median pitch or land on it, there is no difference in the proportions of continuations and reversals of registral direction. In the second experiment, a series of multiple regression analyses demonstrated that a predictor for the relative extremity of the first pitch of an interval was a useful predictor of the following interval. Furthermore, when tessitura constraints were controlled the analyses failed to find any consistent function relating successive intervals in general and those predicting reversals following skips and continuation of motion following small intervals. In the final experiment, bigram models, which are unable to represent a rule for post-skip reversal, were constructed for each composition and used to generate a *Markov twin*. Analyses of the original compositions and their Markov twins demonstrated no evidence of a greater proportion of post-skip reversals in the originals nor of more specific hypothesised structures such as contrary step (a gap followed by a one or two semitone interval in the opposite direction), gap-fill Meyer (1973) or registral return Narmour (1990). These results held when a gap was defined to be an interval greater than two semitones or greater than six semitones.

In a perceptual task, von Hippel (2002) has found evidence that post-skip reversal and step momentum (the expectation that a melody will maintain its direction after a small interval) are stronger in trained than untrained subjects and that the listener's expectations only approximate musical structure. These results are taken to suggest that such principles of expectation are simple learned heuristics based on exposure to structural regularities in music.

4 Discussion

One of the major results of research on melodic expectancy has been general support for the basic tenets of the IR model: first, the bottom-up principles are generally capable of accounting for much of the data on melodic expectancies; second, top-down factors such as tonality and consonance improve the models fit to the data; and third, the patterns of expectancy do not appear to depend on amount of formal training.

All five of the bottom-up principles of the IR model have received some degree of empirical support. However, the empirical research has also suggested two kinds of revision to the model: first, that some of the principles may be modified or dropped to reduce mutual inconsistencies and correlations between them; second, additional principles may improve the match between the predictions of the model and the empirical data. Regarding the latter point there is evidence that other influences on expectancy include tonality (Cuddy & Lunny, 1995; Krumhansl, 1995a,b; Schellenberg, 1996; Thompson *et al.*, 1997), pitch height (Cuddy & Lunny, 1995), consonance (Krumhansl, 1995a,b) and unisons (Krumhansl, 1995a,b). Various tonality predictors, in particular, appear to have a strong top-down influence on expectancy. Furthermore, familiarity with the style appears to increase the relative importance of veridical influences on expectancy and lessen the impact of schematic factors and vice versa (Krumhansl *et al.*, 1999, 2000). Regarding the second point, several analyses have suggested the revision of the principle of registral direction to apply to large intervals only (Cuddy & Lunny, 1995; Schellenberg, 1996). Furthermore, the evidence seems to support the modification of proximity and registral return such that they apply in a continuous rather than discrete fashion (Krumhansl, 1995b; Krumhansl *et al.*, 1999, 2000). Such revisions in addition to dropping principles such as closure for which the empirical support is weak not only result in a better fit to the data but also reduce the redundancy of models (Krumhansl, 1995b; Schellenberg, 1996, 1997). However, cross-cultural studies indicate that such simplifications should be assessed with caution until further research on a range of different musical styles is carried out (Krumhansl *et al.*, 1999, 2000).

While the results of experimental studies of the IR principles have yielded results that are broadly consistent across musical and cultural backgrounds (although see von Hippel 2002 for contrary results), the results of correlational studies do not rule out the possibility that other models might account for the data just as well. Such models might include a tendency to expect small intervals or a tendency to expect notes near the middle of the tessitura (Thompson, 1996). It has been shown that the principles tend to be satisfied in a range of styles of monophonic music (Thompson & Stainton, 1996, 1998) and von Hippel & Huron (2000) have provided evidence that, in the case of post-skip reversal, this is due to the constraints of tessitura rather than an attempt to satisfy listener's expectations. This suggests an alternative account in which the principles of the IR model exist as regularities in music and are acquired as expectancies through the listener's exposure to music. Under this account, the failure to find effects of musical training in most studies would be due to casual (as opposed to formal) exposure to relatively simple melodic music. Further evidence for this view is provided by cross-cultural studies which have found that different variants and combinations of the principles best account for the data (Krumhansl, 1995a; Krumhansl *et al.*, 1999, 2000). As noted by (Cuddy & Lunny, 1995) we shall require further cross-cultural as well as developmental research to provide further support for or against either hypothesis (Cuddy & Lunny, 1995).

The methodologies used in research into melodic expectancies differ along three main dimensions:

1. *task*: perceptual or generative;
2. *contexts*: ranging from a single melodic interval to several measures of existing music;
3. *parameters*: ranging from melodic intervals to entire compositions;

We shall discuss each of these in turn.

Schmuckler (1989) argues that generative paradigms suffer from two major problems: first, the elicitation of completions requires some level of conscious attention to to melodic expectations while such expectations need not necessarily be explicit (Meyer, 1956); and second, the methodology limits the choice of subjects to those who are capable of generating continuations in some recordable fashion. Nonetheless, perceptual and generative paradigms generally indicate that similar processes of expectancy are operative in both perceptual and generative tasks (Schellenberg, 1996; Schmuckler, 1989; Thompson *et al.*, 1997) although production tasks appear to engender a weaker sense of closure (Schellenberg, 1996; Thompson *et al.*, 1997). However, a further problem with the generative paradigm is that analysis is limited to the continuations that were actually generated by the subjects which may be subject to constraints other than expectation alone Cuddy & Lunny (1995). Manzara *et al.* (1992) partially addressed this problem by requiring subjects to repeatedly produce continuations until they produced the correct continuation.

Schmuckler (1989) also suggests that single interval contexts might not be sufficient to generate strong expectations. This does in fact appear to be the case: those studies involving longer contexts (Krumhansl, 1995a; Schellenberg, 1996) have generally found stronger support for the IR principles than comparable studies using single interval contexts (Carlsen, 1981; Cuddy & Lunny, 1995; Krumhansl, 1995b; Thompson *et al.*, 1997). In addition, studies using longer contexts drawn from existing bodies of music have been able to assess the cross-cultural support for the IR principles in a more satisfactory manner. In particular, the research of Krumhansl *et al.* (1999, 2000) had demonstrated that although there is evidence for a core set of implicative principles influencing expectancies, they appear to differ in relative importance across styles and cultures. A problem with all these studies is that they study expectancies at just one moment in a melodic passage while it is likely that expectancies change dynamically during audition of a melody. In the perceptual studies, for example, the melody pauses while the listener responds and this pause may induce a greater sense of closure leading to results which reflect expectancies only at points of melodic closure (Aarden, 2002). The experimental design of Eerola *et al.* (2002) addresses these concerns to some extent but with a consequent loss of control of the precise musical parameters generating expectancies. The methodology adopted by Manzara *et al.* (1992) addresses these concerns and allows for the study of individual parameters but loses much of the immediacy of the perceptual paradigm.

Finally, the parameters studied have generally concerned pitch expectancies which suggests the extension of the research described above to other musical parameters, in particular onsets and duration. Many other issues also remain to be addressed. First, more detailed studies of the kinds of perceptual hierarchies formed by listeners using real musical examples and including analysis of segmentation strategies. Second, a more detailed delineation of top-down and bottom-up principles and their interaction. Third, the extension of the experimental research to more styles of Western and non-Western music. Fourth, further analysis of how expectancies develop as a piece of music unfolds. Fifth, a detailed study of the manner in which composers deny implications for aesthetic effect.

References

Aarden, B. (2002). Expectancy vs. retrospective perception: Reconsidering the effects of schema and continuation judgements on measures of melodic expectancy. In C. Stevens, D. Burnham, E. Schubert, & J. Renwick (Eds.), *Proceedings of the Seventh International Conference on Music Perception and Cognition* (pp. 469–472). Adelaide, Australia: Causal Productions.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, *5*(1), 1–30.

Boltz, M. G. & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology*, *18*(4), 389–431.

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

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73.

Cover, T. M. & King, R. C. (1978). A convergent gambling estimate of the entropy of English. *IEEE Transactions on Information Theory*, *24*(4), 413–421.

Cross, I. (1995). Review of *The analysis and cognition of melodic complexity: The implication-realization model*, Narmour (1992). *Music Perception*, *12*(4), 486–509.

Cuddy, L. L. & Lunny, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgements of continuity. *Perception and Psychophysics*, *57*(4), 451–462.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, *28*(5), 381–389.

Deutsch, D. & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review*, *88*(6), 503–522.

Eerola, T., Toiviainen, P., & Krumhansl, C. L. (2002). Real-time prediction of melodies: Continuous predictability judgements and dynamic models. In C. Stevens, D. Burnham, E. Schubert, & J. Renwick (Eds.), *Proceedings of the Seventh International Conference on Music Perception and Cognition* (pp. 473–476). Adelaide, Australia: Causal Productions.

Jones, M. R. (1981). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology*, *1*(2), 34–51.

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

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception and Psychophysics*, *41*(6), 621–634.

Jones, M. R. (1990). Learning and the development of expectancies: An interactionist approach. *Psychomusicology*, *9*(2), 193–228.

Jones, M. R. & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*(3), 459–491.

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

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. Oxford: Oxford University Press.

Krumhansl, C. L. (1995a). Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft*, *3*(2), 211–250.

Krumhansl, C. L. (1995b). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, *17*, 53–90.

Krumhansl, C. L. (1997). Effects of perceptual organisation and musical form on melodic expectancies. In M. Leman (Ed.), *Music, Gestalt and Computing: Studies in Cognitive Systematic Musicology* (pp. 294–319). Berlin: Springer.

Krumhansl, C. L. & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368.

Krumhansl, C. L., Louhivuori, J., Toiviainen, P., Järvinen, T., & Eerola, T. (1999). Melodic expectation in Finnish spiritual hymns: Convergence of statistical, behavioural and computational approaches. *Music Perception*, *17*(2), 151–195.

Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 579–594.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*(1), 13–58.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the entropy of music: An experiment with Bach chorale melodies. *Leonardo*, *2*(1), 81–88.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Meyer, L. B. (1957). Meaning in music and information theory. *Journal of Aesthetics and Art Criticism*, *15*(4), 412–424.

Meyer, L. B. (1967). *Music, the Arts and Ideas: Patterns and Predictions in Twentieth-century Music*. Chicago: University of Chicago Press.

Meyer, L. B. (1973). *Explaining Music: Essays and Explorations*. Chicago: University of Chicago Press.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-realisation Model*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-realisation Model*. Chicago: University of Chicago Press.

Palmer, C. & Krumhansl, C. L. (1990). Mental representations for musical metre. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728–741.

Rosner, B. S. & Meyer, L. B. (1982). Melodic processes and the perception of music. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 317–341). New York: Academic Press.

Rosner, B. S. & Meyer, L. B. (1986). The perceptual roles of melodic process, contour and form. *Music Perception*, *4*(1), 1–40.

Schaffrath, H. (1992). The ESAC databases and MAPPET software. *Computing in Musicology*, *8*, 66.

Schaffrath, H. (1994). The ESAC electronic songbooks. *Computing in Musicology*, *9*, 78.

Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide [computer database]*. Menlo Park, CA: CCARH.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realisation model. *Cognition*, *58*(1), 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, *14*(3), 295–318.

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

Schmuckler, M. A. (1990). The performance of global expectations. *Psychomusicology*, *9*(2), 122–147.

Schmuckler, M. A. (1997). Expectancy effects in memory for melodies. *Canadian Journal of Experimental Psychology*, *51*(4), 292–305.

Simon, H. A. & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgement* (pp. 219–250). New York: Wiley.

Thompson, W. F. (1996). Eugene Narmour: *The analysis and cognition of basic musical structures* (1990) and *The analysis and cognition of melodic complexity* (1992): A review and empirical assessment. *Journal of the American Musicological Society*, *49*(1), 127–145.

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception and Psychophysics*, *59*(7), 1069–1076.

Thompson, W. F. & Stainton, M. (1996). Using *Humdrum* to analyse melodic structure: An assessment of Narmour's implication-realisation model. *Computing in Musicology*, *12*, 24–33.

Thompson, W. F. & Stainton, M. (1998). Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Perception*, *15*(3), 231–252.

Unyk, A. M. & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology*, *7*(1), 3–23.

von Hippel, P. T. (2000). Questioning a melodic archetype: Do listeners use gap-fill to classify melodies? *Music Perception*, *18*(2), 139–153.

von Hippel, P. T. (2002). Melodic-expectation rules as learned heuristics. In C. Stevens, D. Burnham, E. Schubert, & J. Renwick (Eds.), *Proceedings of the Seventh International Conference on Music Perception and Cognition* (pp. 315–317). Adelaide, Australia: Causal Productions.

von Hippel, P. T. & Huron, D. (2000). Why do skips precede reversals? The effects of tessitura on melodic structure. *Music Perception*, *18*(1), 59–85.

Witten, I. H., Manzara, L. C., & Conklin, D. (1994). Comparing human and computational models of music prediction. *Computer Music Journal*, *18*(1), 70–80.