Proceedings of the 8th International Conference on Music Perception & Cognition, Evanston, IL, 2004
ICMPC8
S.D. Lipscomb, R. Ashley, R.O. Gjerdingen, & P. Webster (Eds.)
Adelaide, Australia: Causal Productions

RETHINKING GESTALT INFLUENCES ON MELODIC EXPECTANCY

M. T. Pearce *G. A. Wiggins*

Centre for Computational Creativity,
City University, London, EC1V 0HB, UK
{m.t.pearce,geraint}@city.ac.uk

ABSTRACT

The implication-realisation model of melodic expectancy (Narmour, 1990) posits two independent cognitive systems: one which reflects the top-down influences of acquired stylistic knowledge; and another which consists of innate melodic schemata based on the Gestalt principles. While there exists empirical support for both aspects of the model, the results presented here question the need to invoke innate representations to account for observed patterns of melodic expectancy. Gestalt-like patterns of expectancy can be accounted for largely in terms of the induction of statistical regularities in existing repertoires of music.

1. INTRODUCTION

The generation of expectations is recognised as being an especially important factor in music cognition. From a music-analytic perspective, it has been argued that the generation and subsequent confirmation and violation of expectancies is critical to aesthetic experience and the expression of emotional meaning in music (Meyer, 1956; Narmour, 1990). From a psychological perspective, expectancy has been found to influence both the perception (Krumhansl, 1995b; Krumhansl et al., 1999, 2000; Schellenberg, 1996, 1997; Schellenberg et al., 2002) and production (Thompson et al., 1997) of music.

Narmour (1990) has proposed an influential theory of expectancy in melody which attempts to characterise the set of implied continuations to an incomplete melodic pattern. According to the theory, the expectancies of a listener are influenced by two independent cognitive systems: the first consists of a small number of Gestalt-like principles which are held to be innate and universal; the second consists of style-specific influences on expectancy that are acquired through extensive exposure to music in a given style. According to the theory developed here, observed patterns of melodic expectancy can be accounted for in terms of the induction of statistical regularities in existing repertoires of music. In order to examine this claim, previous data on melodic expectancy are reanalysed in terms of a computational model embodying the theory.

2. BACKGROUND

2.1. The Implication-Realisation Model

The *Implication-Realisation* (IR) model (Narmour, 1990) of melodic perception posits two independent perceptual systems – the *bottom-up* and *top-down systems* of melodic implication. While the principles of the former are held to be hard-wired, innate and universal, the principles of the latter are held to be learned and hence dependent on musical experience. The top-down influences on expectancy 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 aspects of a particular composition such as distinctive motivic and rhythmic patterns. 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.

In the bottom-up system, melodic pitch intervals vary in the degree of *closure* that they convey. 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 in terms of several principles of continuation which are influenced by the Gestalt principles of proximity, similarity, good continuation and symmetry. The principle of **registral direction** states that small intervals imply continuations in the same registral direction whereas large intervals imply a change in registral direction. 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. The principle of **intervallic difference** states that small intervals imply a subsequent interval that is similar in size, while large intervals imply a consequent interval that is smaller in size. 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 of the first tone of an implicative interval in cases where the realised interval reverses the registral direction of the implicative interval. This principle can be viewed as an application of the Gestalt principle of proximity. The principle of **proximity** describes a general implication for small intervals between any two tones. The implication is graded according to the size of the interval and can again be viewed as an application of the Gestalt principle of proximity. **Closure** is determined two conditions: first, a change in registral direction; and second, movement to a smaller sized interval. The first condition is related to the Gestalt principle of symmetry. Degrees of closure exist corresponding to the satisfaction of both, one or neither of the conditions.

Although the theory 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, 1995b; Schellenberg, 1996).

2.2. Empirical Studies of the IR Model

Empirical work to date has examined both the top-down and bottom-up systems of the IR model. Regarding the former, support has been found for the influence of top-down factors such as tonality and style-dependent statistical knowledge (Krumhansl, 1995b; Krumhansl et al., 1999, 2000; Schellenberg, 1996). Here, however, we focus on the influence of the bottom-up principles and, in particular, the claim that they reflect the action of innate predispositions.

Krumhansl (1995b) reports a series of three experiments examining expectancies generated in the context of melodic fragments taken from British folk songs, Webern's *Lieder* and Chinese folk songs respectively. 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. Analysis of the results yielded support for all of the bottom-up principles (with the exception of intervallic difference for the second experiment). Overall, the weakest contribution was made by intervallic difference and the strongest by proximity. Support was also found for two additional principles which assigned low ratings to realised intervals forming unisons and high ratings to those forming octaves.

Schellenberg (1996) argues that the bottom-up system of the IR model is over-specified and contains redundancy due to collinearities between the bottom-up principles. As a result, the model may be expressed more simply and parsimoniously without loss of predictive power. Support was found for this argument in an independent analysis of the experimental data first reported in Krumhansl (1995b) using a model consisting of registral return as well as revised versions of registral direction and proximity. In a further analysis, Schellenberg (1997) applied principal components analysis to this revised model which resulted in the development of a two-factor model. The first factor is the bottom-up principle of proximity which predicts that listeners expect consecutive tones to be proximate in pitch; the second, *pitch reversal* predicts an expectation for a change of registral direction (or a unison). This model is considerably simpler than the revised model of Schellenberg (1996) and yet does not compromise the predictive power of that model in accounting for the data reported by Krumhansl (1995b).

Similar experiments with Finnish spiritual folk hymns (Krumhansl et al., 1999) and indigenous folk melodies (yoiks) of the Sami people of Scandinavia (Krumhansl et al., 2000) have, however, questioned the cross-cultural validity of such revised models. In both studies, it was found that the IR model (revised and extended by Krumhansl 1995a) provided a much better fit to the data than those of Krumhansl (1995b), Schellenberg (1996) and Schellenberg (1997). By contrast, Schellenberg et al. (2002) have found the opposite to be true in experiments with adults and infants in a task involving the rating of continuation tones following contexts taken from Acadian (French Canadian) folk songs. They suggest that the difference may be attributable to the fact that none of the musical contexts used the experiments of Krumhansl et al. (1999, 2000) ended in unambiguously large and implicative intervals. While Schellenberg et al. (2002) found strong support for proximity and limited support for pitch reversal, Krumhansl et al. (1999) found that the strongest influence came from the principle of proximity with some influence of registral return and intervallic difference. Krumhansl et al. (2000), on the other hand, found the strongest bottom-up influence came from the principle of intervallic difference with weak support for the principles of proximity and registral return. The consonance predictors of Krumhansl (1995a) made a strong contribution to both models though slightly more so in the case of the folk hymns.

The pattern emerging from this body of research is that the bottom-up IR principles constitute too simple and inflexible a model to account for the effects of differences in musical exposure, age, musical style and experimental task (Krumhansl et al., 2000; Schellenberg et al., 2002). While research has typically failed to find differences between bottom-up influences across levels of musical training (Krumhansl et al., 1999; Schellenberg, 1996) and relevant stylistic experience (Krumhansl et al., 2000; Schellenberg, 1996), it is important to note that ubiquitous patterns of behaviour do not necessarily imply innate cognitive mechanisms. The experimental evidence is broadly consistent with the argument that the bottom-up principles reflect innate predispositions but it is also consistent with an account that they reflect the influence of long-term informal exposure to simple and ubiquitous regularities of Western tonal-harmonic music (Thompson et al., 1997).

Recent research has given reason to favour the latter account. First, there is evidence that existing repertoires of music exhibit regularities that satisfy the bottom-up IR principles (Thompson & Stainton, 1996, 1998) and that these regularities may reflect universal physical constraints of performance such as the relative difficulty of singing large intervals accurately and the fact that large intervals will tend towards the limits of a singer's vocal range (Schellenberg, 1997; von Hippel & Huron, 2000). Second, Schellenberg et al. (2002) report developmental research demonstrating that expectancies are better explained by (revised) IR models with increasing age and musical exposure. While pitch proximity was a strong influence for listeners of all ages, the influence of more complex predictors such as pitch reversal (Schellenberg, 1997) and registral return (Narmour, 1990) only became apparent with the older listeners. These results suggest that mature patterns of expectancy in melody are acquired through exposure to music which results in the progressive induction of increasingly complex musical regularities.

3. AIMS

The first aim of this research was to test the hypothesis that a statistical learning model of melodic expectancy trained on a reasonable corpus of melodic music can account for experimentally observed patterns of expectancies as well as, or better, than the IR model. Since the model acquires all its knowledge of musical structure through exposure to melodic music, corroboration of the hypothesis would demonstrate that it is not necessary to posit innate and universal musical schemata to account for existing data on melodic expectancy. The alternative account, put forward here, is that melodic expectancy can be accounted for in terms of statistical induction of both intra- and extra-opus regularities in the music that listeners are exposed to. The methodological approach followed compares the pattern of melodic expectancies generated by a computational model to those of human subjects observed in previously reported experiments. The second aim was to examine which melodic features present in or simply derivable from the musical surface afford regularities that might be used to acquire the observed patterns of melodic expectancy.

4. METHOD

Schellenberg (1996, Experiment 1) reports an experiment in which subjects were asked to rate continuation tones following eight melodic fragments taken from British folk songs (Palmer, 1983; Sharp, 1920). The subjects were 20 members of the community of Cornell University in the USA of whom half had limited musical training and half had moderate musical training. The eight contexts were chosen such that they ended on an implicative interval. Half of the fragments are in a minor mode and half in a major mode. Four of the fragments end with one of two small intervals (2 or 3 semitones) in ascending and descending forms while four end with one of two large intervals (9 or 10 semitones) in ascending and descending forms. Continuation tones consisted of the 15 diatonic tones in a two octave range centred on the final tone of the melodic context. The subjects were asked to rate how well the continuation tone continued the melody on a scale from 1 (extremely bad continuation) to 7 (extremely good continuation). The experiment yielded 120 continuation tone ratings for each subject. Significant inter-subject correlation for all subjects warranted the averaging of the data across subjects and training levels. The mean continuation tone ratings are available in Schellenberg (1996, Appendix A).

Our computational model of expectancy is based on statistical n-gram techniques which induce transition frequencies through exposure to example sequences. The n-gram model used in this experiment was enhanced in several ways. First, it can combine models of several orders to generate expectancies for forthcoming musical events. Second, it can operate with a variable global order bound. Third, it is capable of combining information derived through exposure to a large body of example melodies with specific knowledge dynamically induced during exposure to a particular melody. Finally, it is capable of inducing and making use of statistical regularities in a number of melodic attributes both present in and derivable from the musical surface. In order to achieve this, the model operates in a *multiple-viewpoint* framework (Conklin & Witten, 1995) which is capable of representing basic melodic attributes such as chromatic pitch, derived attributes such as pitch interval, attributes which are defined non-sequentially and interactions between attributes. For a more detailed description of the model see Pearce & Wiggins (2004) and Pearce et al. (2004).

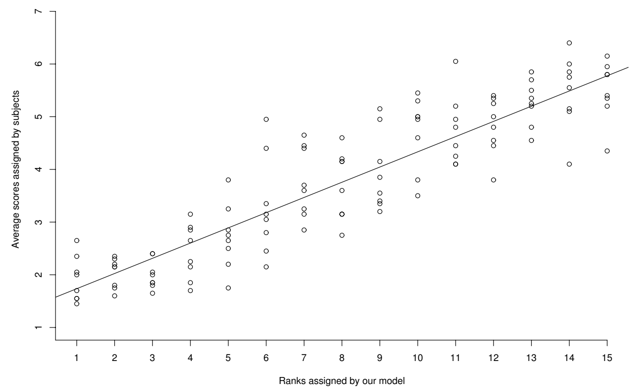

Figure 1: Correlation between the average scores assigned by subjects and the distribution ranks assigned by the model to continuation tones in the experiments of Schellenberg (1996).

Each melodic attribute of interest is modelled by a *viewpoint*. For our purposes, the use of a viewpoint corresponds to a hypothesis that the melodic dimension modelled by that viewpoint affords regularities which are capable of supporting the acquisition of observed patterns of melodic expectancy. A small set of viewpoints was selected to reflect a range of findings in the psychological literature. Viewpoints were included for simple attributes of pitch including absolute chromatic pitch (`cpitch`), pitch class (`cpitch-class`), pitch interval (`cpint`), interval class (`cpcint`) and contour (`contour`). Another set of viewpoints included in our experiments investigated the influence of *relative* pitch structure. Viewpoints were included which modelled pitch interval relative to the tonic of a melody (`cpintfref`), the first note in a melody (`cpintfip`) and the first note in the current bar (`cpintfib`). In order to examine the influence of rhythmic structure on expectancy, a handful of viewpoints were created to reflect the conjunction ($\otimes$) of several simple dimensions of pitch structure (`cpitch`, `cpint`, `contour` and `cpintfref`) and rhythmic structure represented by event durations (`dur`). In order to examine the influence of regularities in joint melodic and metrical structure on melodic expectancy, a viewpoint was developed which models pitch interval between events occurring on tactus beats (`thrtactus`). Finally, the influences of tonality were further investigated by a viewpoint `inscale`, which models whether or not a note is in the major or minor diatonic scale based on the key, and by conjoining `cpintfref` with `cpint` and `cpintfip` (see also Conklin, 1990; Conklin & Witten, 1995).

The data used to train the model were chosen to reflect a range of melodic styles and consisted of 152 Canadian ballads (Creighton,

Viewpoints	R
`cpitch`	0.847
`cpitch`, `cpintfib`	0.865
`cpitch`, `cpintfib`, `cpintfip`	0.897
`cpitch`, `cpintfib`, `cpintfip`, `cpintfref`⊗`cpint`	0.901

Table 1: The results of viewpoint selection.

1966), 185 chorale melodies (Riemenschneider, 1941) and 566 German folk songs (Schaffrath, 1995). The trained model was exposed to each of the eight contexts used by Schellenberg (1996). For each context, the model returns a probability distribution over the set of 15 diatonic tones from an octave below to an octave above the final tone of the context. Each of the pitches was assigned a rank according to its estimated probability in inverse order (such that high probability pitches were assigned high ranks). *Forward stepwise selection* was used to select viewpoints which maximise the regression coefficient of the mean ratings obtained by Schellenberg (1996) regressed on the distribution ranks of the model. Given an initially empty viewpoint system, the algorithm considers, at each step, all possible single viewpoint additions and selects the one which yields the most improvement in the regression coefficient.

5. RESULTS

In a multiple regression analysis of the same data, Schellenberg (1996) reports that the (revised) IR model accounts for approximately 75% of the variance in the mean continuation tone ratings ($R^2 = 0.759$). This model consists of bottom-up predictors for registral direction (revised), proximity (revised) and registral return (revised) as well as a top-down predictor for tonality. In accordance with our first experimental hypothesis, the final model developed in this experiment provides a better fit to the data accounting for approximately 81% of the variance in the mean continuation tone ratings ($R = 0.901,\ R^2 = 0.812,\ F(1, 118) = 508.6,\ p < 0.0001$). The relationship between the patterns of expectancy exhibited by the model and by the subjects in the experiments of Schellenberg (1996) is plotted with the fitted regression line in Figure 1.

In viewpoint selection, `cpitch` made a strong contribution to the model (see Table 1) suggesting that the contexts used in this experiment are capable of invoking expectancies based on commonly occurring pitch sequences. The viewpoints `cpintfib` and `cpintfip` also contributed to improving the fit of the model to the human data, suggesting that regularities defined by reference to salient events (the first in the piece and the first in the current bar) may have influenced melodic expectancies. Finally, a viewpoint modelling the influence of tonality and interval structure (`cpintfref`⊗`cpint`) further improved the fit. It is surprising that viewpoints modelling tonality were not found to exert a stronger influence. Given the short contexts, if subjects were not able to induce a tonality for the context, it is not unreasonable to suppose that expectancies might rely on regularities based on intervals from salient events such as the first note in the current bar or the first in the piece (Thompson et al., 1997). Finally, we hypothesised that rhythmic features might interact with regularities in pitch, pitch interval and contour to influence expectancies. The results provide no support for this hypothesis although it is possible that other rhythmic features (such as inter-onset interval or duration ratio) might yield improved fits to the data.

6. CONCLUSIONS

The IR model of melodic expectancy posits a top-down system influencing expectancy through the action of learned, style-dependent musical knowledge and an independent bottom-up system influencing expectancy through the action of innate Gestalt-like predispositions to respond differentially to simple patterns of melodic intervals. The results reported here question the need to make strong claims about innate representations to account for Gestalt-like patterns of expectancy in melody. In contrast, the observed patterns of expectancy may be accounted for more simply and parsimoniously in terms of the combined influence of innate general purpose learning mechanisms and the structure of the musical environment.

7. REFERENCES

Conklin, D. (1990). Prediction and entropy of music. Master's thesis, Department of Computer Science, University of Calgary. Available as Technical Report 1990–390–14.

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

Creighton, H. (1966). *Songs and Ballads from Nova Scotia*. New York: Dover.

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., Louhivand, 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., Toiviainen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivand, J. (2000). Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition*, *76*, 13–58.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. 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.

Palmer, R. (Ed.). (1983). *Folk songs collected by Ralph Vaughan Williams*. London: Dent.

Pearce, M. T., Conklin, D., & Wiggins, G. A. (2004). Methods for combining statistical models of music. To appear in *Proceedings of the Second International Symposium on Computer Music Modelling and Retrieval*. Esbjerg, Denmark.

Pearce, M. T. & Wiggins, G. A. (2004). Improved methods for statistical modelling of monophonic music. To appear in *Journal of New Music Research*.

Riemenschneider, A. (1941). *371 harmonised chorales and 69 chorale melodies with figured bass*. New York: G. Schirmer, Inc.

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*. Menlo Park, CA: CCARH.

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

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

Schellenberg, E. G., Adachi, M., Purdy, K. T., & McKinnon, M. C. (2002). Expectancy in melody: tests of children and adults. *Journal of Experimental Psychology: General*, *131*(4), 511–537.

Sharp, C. J. (Ed.). (1920). *English folk songs*, volume 1-2, selected ed. London: Novello.

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. (1995/1996). 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.

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.