OPEN

Music Cognition as Mental Time Travel

Freya Bailes[1], Roger T. Dean[2] & Marcus T. Pearce[3]

[1]Department of Drama and Music, University of Hull, [2]MARCS Institute, University of Western Sydney, [3]Queen Mary, University of London.

As we experience a temporal flux of events our expectations of future events change. Such expectations seem to be central to our perception of affect in music, but we have little understanding of how expectations change as recent information is integrated. When music establishes a pitch centre (tonality), we rapidly learn to anticipate its continuation. What happens when anticipations are challenged by new events? Here we show that providing a melodic challenge to an established tonality leads to progressive changes in the impact of the features of the stimulus on listeners' expectations. The results demonstrate that retrospective analysis of recent events can establish new patterns of expectation that converge towards probabilistic interpretations of the temporal stream. These studies point to wider applications of understanding the impact of information flow on future prediction and its behavioural utility.

We choose to study music to understand the real-time modification of temporal prediction because of its basis in well-learned stylistic regularities and its dynamic unfolding[1–3]. Statistical analyses of musical repertoires reveal more or less probable transitions from note(s) to note(s) within a given style, and these probabilities have been used successfully in computational models to predict the melodic expectations of listeners e.g.[4], including the strong influence of pitch proximity e.g.[5,6]. Western music typically comprises both melody (the single voice tune) and harmony (the multiple voice progression of chords which accompany it). The perceptual processing of probable events in music is facilitated[7,8], so the formation of accurate expectations is behaviourally beneficial.

There has been little consideration of how expectations change through time. A given musical event necessarily serves a dual function as the fulfilment or violation of previous expectations, and towards establishment of future expectations[9]. Huron[2] developed the 'ITPRA' theory of expectation in relation to music. The 'Imagination' and 'Tension' responses occur prior to a musical event, while the 'Prediction', 'Reaction' and 'Appraisal' responses occur thereafter. Correct predictions are rewarded and incorrect ones penalised by the limbic system. Appraisal, which involves a cognitive assessment of the whole context, has been little investigated. Musical appraisal is an example of processes that are significant in the assessment of any event stream of biological relevance. We refer here to 'retrospection' to address a component of appraisal that, in the current experiment, deals with musical tonality (pitch-centeredness).

Models of tonality assume the importance of a recency effect in music listening: the most recent musical events in a sequence influence expectations more than prior events, which decay in sensory memory. Estimates of the rate of exponential sensory decay range from a one- to a four-second half-life[10], but this may also involve the influence of intervening events[11]. Thus expectations may derive from long-term knowledge acquired through prior exposure to the statistical properties of music within a given tradition, and also from short-term memory of hearing the recent distribution of notes[15]. In sum, expectations may be sensory or cognitive, stimulus-driven or a result of musical acculturation[7,10].

We assessed whether the influence of the passage of time upon expectations can be demonstrated empirically by defining conditions in which both expectation and retrospective appraisal contribute in musical processing. Listeners rated the goodness of fit of a probe tone (a single pitch) to a preceding melodic context, and by manipulating the time delay of the probe presentation we tested the robustness of expectations established by the context and the evidence for retrospective appraisal or sensory memory consolidation (such as decay or recoding[12]). We interposed silences, as often occur in music, between the melodic context and the probe tone[13]. We probed listeners' expectations at delays ranging from 0 to 19.2 seconds, to capture any shifts in expectedness resulting from reappraisal or memory consolidation. One previous musical study interposed silences of 7.07 s to 13.3 s between prime and probe, and found that even across such an extensive span, the proximity in pitch of the last note of the prime to the subsequent probe significantly facilitated its perception[14]. Other studies of musical expectation more typically separate context from probe tone by only ~ 1 s e.g.[15].

Figure 1 | The 'More Unstable' stimulus. Sample stimulus comprising 4-part chords in F# major (first dotted box) with an arpeggio continuation outlining a new tonal centre in G major (second dotted box). This is more unstable as the final arpeggio has a mean information content (based on IDyOM) of 5.64. A piano timbre was used for stimulus presentation. Note that both stimuli were heard starting in all twelve possible keys. In both figures, accidentals apply only to the note that immediately follows; in this figure advisory accidentals are shown in the second bar for G and D, confirming that they are not sharpened (unlike their corresponding notes in the opening key of F# major).

Results

We examine harmonic expectations in the context of 4-part chord sequences with arpeggiated continuations (3 single notes) that outline a new tonal centre (or 'key': see Figures 1 & 2) challenging that of the chords. Note that, by design, we do not modulate with transitional material (e.g. a cadence) from one key to another; our purpose was to introduce levels of instability in the stimulus. Following presentation of the stimulus (chord and arpeggio portions), a probe tone is presented, and participants rate the goodness of fit of this probe to the stimulus. In a previous study[16], probe tones were presented after 500 ms, and their ratings suggest that listeners were influenced by the temporal course of a change (modulation) from one key to another: with no delay to probe presentation, expectations reflected those of the final key. Here we determine whether this pattern remains after intervening periods of silence, or whether a process of reappraisal occurs whereby expectations increasingly fit the stimulus opening or the stimulus as a whole as the probe presentation is delayed.

Evidence for the establishment of expectations formed by the most recent musical context (the arpeggio) was found with two kinds of stimulus. One was 'very unstable' harmonically (stimulus 1, Figure 1: all three notes of the concluding arpeggio are foreign to the starting key), the other 'less unstable' (stimulus 2, Figure 2: the first arpeggio note is intrinsic to the starting key, the second and third foreign; see Methods). For the very unstable stimulus, mean ratings of the probe tone were correlated with the pitch proximity (in semitones) from the last note of the preceding stimulus context (i.e., of the arpeggio). This correlation was significant regardless of the intervening time between the stimulus context and probe (immediate, $r = .67$, $p = .02$; 1.8 s, $r = .80$, $p = .002$; 6 s, $r = .68$, $p = .01$; 19.2 s, $r = .70$, $p = .01$). For the less unstable stimulus, mean probe tone ratings correlated with those documented for the tonality[17] of the key of the final arpeggio of the stimulus (not the opening tonality). Again, this correlation was significant at every time point (immediate, $r = .58$, $p = .05$; 1.8 s, $r = .66$, $p = .03$; 6 s, $r = .66$, $p = .03$; 19.2 s, $r = .64$, $p =$
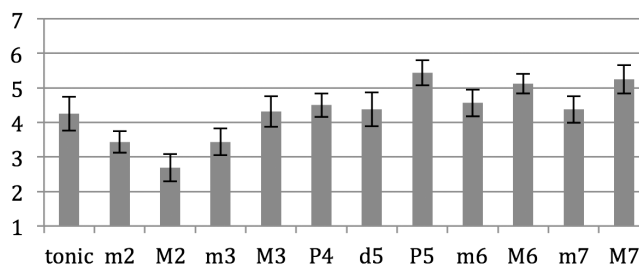

Figure 2 | The 'Less Unstable' stimulus. Sample stimulus comprising 4-part chords in G major with an arpeggio continuation outlining a new tonal centre in F# major. This is less unstable as the final arpeggio has a mean information content (based on IDyOM) of 4.29.
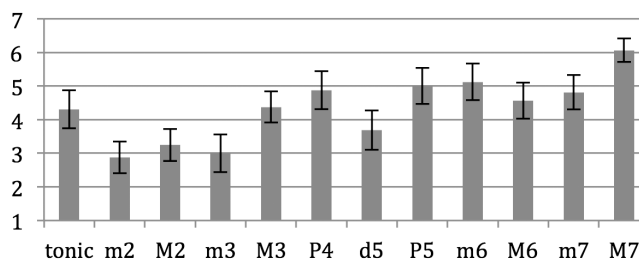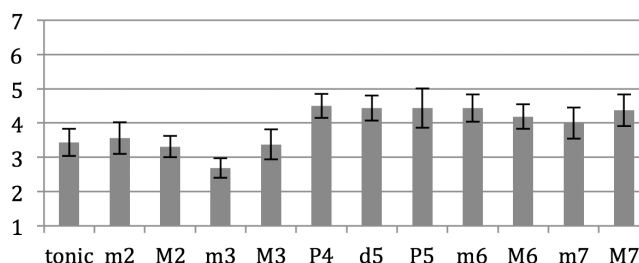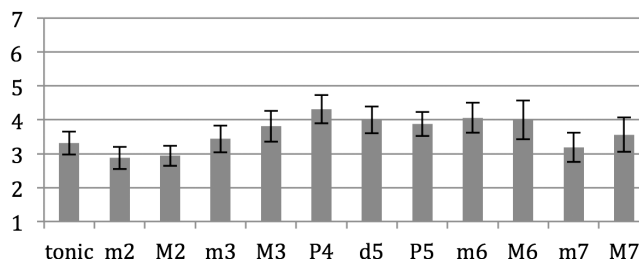

Figure 3 | Distribution of probe ratings following the More Unstable stimulus at each probe delay. Mean ratings of the goodness of fit of the probe tone to the unstable stimulus. 1 – the fit is very bad, 7 the fit is very good. Error bars represent the standard deviation. Labels along the x axis describe the tonal relationship of the probe tone to the initial (chordal) tonality (or in brackets the number of semitones separating them), as follows: tonic – root of the scale (0), m2 – minor second (1), M2 – major second (2), m3 – minor third (3), M3 – major third (4), P4 – perfect fourth (5), d5 – diminished fifth (6), P5 – perfect fifth (7), m6 – minor sixth (8), M6 – major sixth (9), m7 – minor seventh (10), M7 – major seventh (11).

.03), showing that the terminating short-term context induced a relatively unambiguous tonal centre in the listeners.

The most interesting results demonstrate that probe tone ratings changed with probe delay. A significant interaction between probe delay and probe pitch was found in a repeated measures ANOVA of mean probe ratings, across both more and less unstable stimuli: $F(27.2, 408.3) = 1.75$, $p = .012$ (with Huynh-Feldt correction). With the more unstable stimulus, this retrospective change led to a correlation between the distribution of probe ratings (see Figures 3 and 4) and the pitch probabilities estimated by the statistical model (IDyoM, see[4]) on the basis not only of the final arpeggio but also of the initial tonality of the stimulus context (established by the opening chords). This occurred when the probe presentation was delayed by 6 s ($r = .61$, $p = .04$) or 19.2 s ($r = .65$, $p = .03$), but not with delays of only 0 or 1.8 s. Therefore, after a very unstable stimulus, listeners' judgements become increasingly consistent with the overall probabilistic structure of the music with increasing time. No correlation between probability structures involving the initial tonality and probe ratings was found for the less unstable stimuli, though the ratings changed with time.

Figure 3 (0 s) shows the highest fit (rating 6) perceived in the immediate responses was the major seventh (unlikely were there only the initial tonality without the melodic challenge). By 1.8 s the highest fit, with a lower rating is at the perfect 5th; and by the later times the ratings distribution is flattened, and no values exceed 5. Further analysis was necessary to determine the nature and significance of these changes. If changes in probe rating with delay merely represented sensory decay, their distribution would collapse toward the scale mid-point of '4' (i.e., where the fit of the probe is "neutral"). Response variability did decrease as the delay to the probe presentation increased, as evidenced by an effect of probe delay in a one-way repeated measures ANOVA of the standard deviation of probe tone ratings: more unstable stimulus, $F(3, 45) = 4.31$, $p = .009$; less unstable stimulus, $F(3, 45) = 4.70$, $p = .006$. However, Cramer tests for the equality of distributions showed that the distribution of probe ratings was always significantly different from this scale mid-point (immediate, Cramer Statistic = 2.34, $p = .005$; 1.8 s, Statistic = 1.47, $p = .003$; 6 s, Statistic = 0.97, $p = .002$; 19.2 s, Statistic = 2.03, $p = .005$). Thus for both stimuli, the process of change is better explained as an impact of consolidation than of decay.
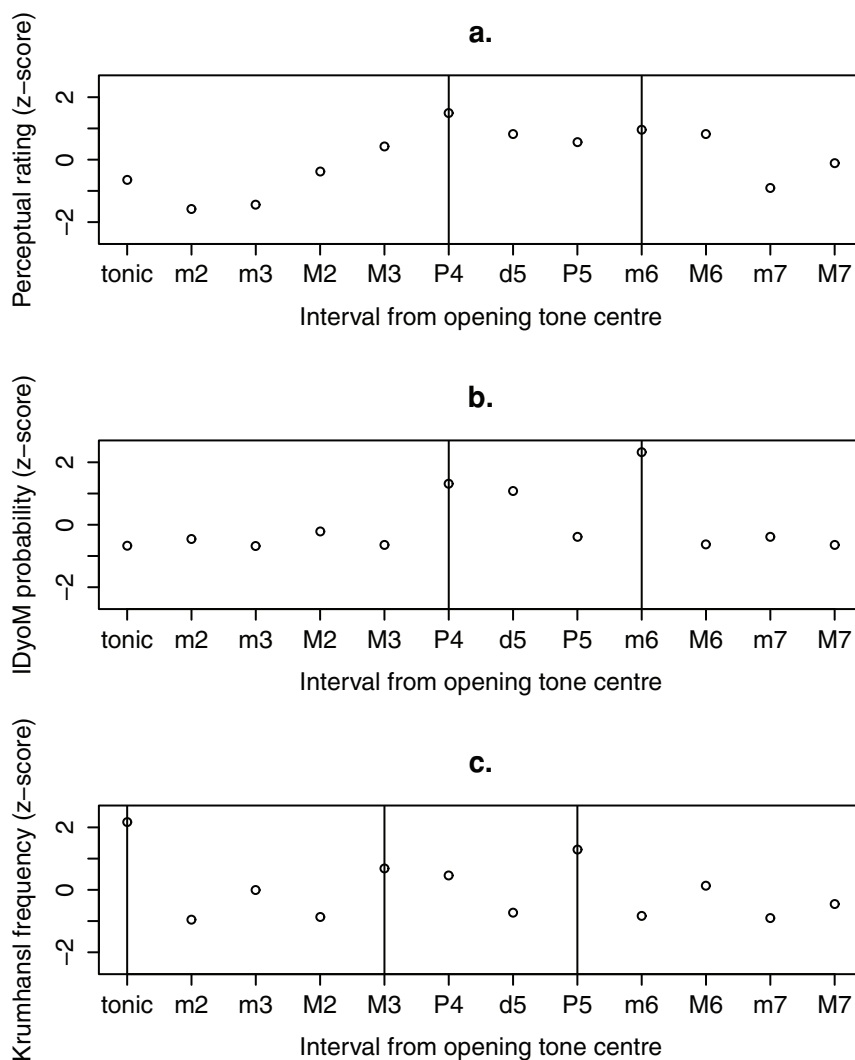

Figure 4 | **Comparison of Perceptual Ratings (4a), IDyoM probabilities (4b), and Krumhansl-Kessler Frequencies (4c) for the More Unstable Stimulus at 19.2 s delay.** All values were z-standardised (that is, expressed in terms of the standard deviation of the value set, becoming both positive and negative in relation to the mean) so that they can be viewed on the same scale. The upper panel (a) shows the perceptual ratings; the middle panel (b) the IDyoM probabilities; and the bottom panel (c) the Krumhansl-Kessler pitch frequencies for tonal music in a major key (as here). The vertical lines indicate the major peaks in the ratings: convergent for the Perceptual and IDyoM values (panels a, b: 2 peak values), and divergent in the third panel (c: Krumhansl-Kessler frequencies: 3 peak values). Labels along the x-axis describe the tonal relationship/semitone distance of the pitch to the opening tonal centre (as for Fig. 3).

The additional further analysis illustrated in Figure 4 uses z-standardised values to compare on the same scale the somewhat flatter perceptual values with the Krumhansl-Kessler frequencies and with the IDyoM probabilities. Figure 4 (top two panels) also shows that the two most preferred pitches after 19.2 seconds of appraisal of the more unstable stimulus were 5 and 8 semitones from the opening tone centre, in agreement with the probabilistic model. In contrast, the Krumhansl-Kessler[17] pitch frequencies shown in the third panel are quite different, emphasising intervals of 0, 4 and 7 semitones. This perceptual resolution and preference is similar to that of an optimal Bayesian decision maker performing a task based on a fixed amount of perceptual evidence, as developed for example in powerful models of continuous speech recognition such as Shortlist B and Merge[18]. Discussion of the presently more limited study of such Bayesian models in the context of musical pitch sequence predictions is given by Temperley[19].

Discussion

We argue that the observed change in judgement with time reflects reappraisal of the prior stimulus taken as a whole, which could also be construed as memory consolidation, extracting generalities from the information. The process could not be that of 'recoding' enforced by further incoming related musical events[12], because the delays were occupied by silence. We note that from a musicological perspective, the more unstable stimulus (Figure 1) has a codified form, where the melodic triad is a semitone above the opening tonic (home key): this is one kind of 'Neapolitan' harmony. This might make one think that it is familiar and hence stable. However, codification is not the same as either familiarity or 'stability', nor statistical probability; we identify stability as the latter and determine it quantitatively.

We conclude in accord with Huron's ITPRA (and with an optimal Bayesian process) that retrospective appraisal influences perceptions of musical fit (and this is not solely forgetting[11]). Stimulus properties and the stability of their mental representation interact with the shifting range of potential cognitive operations. Our results suggest that very unstable stimuli are likely to be progressively and then integrally reappraised, more stable stimuli less so. This result has potentially broad implications for behavioural therapy and learning processes in that unfamiliar stimuli (which are hence information rich) may be used to influence the statistical interpretation of familiar stimuli with which they can be seen to relate, and in turn the unfamiliar stimuli can be integrated within the resultant learned statistical field.

Methods

Four male and 12 female non-musicians studying undergraduate psychology (mean age 20.5 years, range 17–37) participated for course credit. Each stimulus consisted of a sequence of five chords in one tonality followed by three monophonic notes forming an arpeggio in another key a semitone away (inter-onset interval 600 ms throughout). The 'less unstable' stimulus commenced in G major and ended with an F# major triad (though no modulation was cadentially confirmed), while the corresponding 'more unstable' stimulus moved from F# major to G major. Importantly, each stimulus was presented in all twelve possible keys. The information content (IC)[4] of the arpeggio of the melodic triad in the context of the opening key, confirmed the much greater unexpectedness of the arpeggio in the more unstable stimulus (more unstable: Mean IC = 5.64; less unstable: Mean IC = 4.29). Following the stimulus, a probe tone (taken from each of the 12 pitches of the Western chromatic scale[17]) was presented immediately, 1.8 s, 6 s, or 19.2 s later by *Max/MSP* using a piano timbre. Participants listening over closed headphones rapidly rated the goodness of fit of the probe to the preceding context, using a 7-point scale ranging from '1 – the fit is very bad' to '7 – the fit is very good'. The pitch relationship of the probe to the starting tonal centre was manipulated. 96 trials were presented (2 stimuli × 4 probe delays × 12 probe pitches), with a mask of 16 randomly pitched 'Shepard tones' (125 ms tone durations) separating each trial. To avoid serial order effects, each trial was presented in a different starting key from the preceding and succeeding trials, selected at random.

The experiments were approved by the human ethics committee of the University of Western Sydney, and informed consent was obtained from all subjects.

1. Vuust, P. & Kringelbach, M. L. The pleasure of making sense of music. *Interdisciplinary Science Reviews* **35**, 166–182 (2010).
2. Huron, D. *Sweet anticipation: Music and the psychology of expectation.* (MIT Press, 2006).
3. Meyer, L. *Emotion and Meaning in Music.* (The University of Chicago Press, 1956).
4. Pearce, M. T. & Wiggins, G. A. Expectation in melody: The influence of context and learning. *Music Percept.* **23**, 377–405 (2006).
5. Ockelford, A. Implication and expectation in music: a zygonic model. *Psychol. Music* **34**, 81–142 (2006).
6. Narmour, E. *The analysis and cognition of basic melodic structures: The implication-realization model.* (The University of Chicago Press, 1990).
7. Marmel, F., Tillmann, B. & Delbé, C. Priming in melody perception: Tracking down the strength of cognitive expectations. *J. Exp. Psychol.: Hum. Percept. Perform.* **36**, 1016–1028 (2010).
8. Justus, T. C. & Bharucha, J. J. Modularity in musical processing: The automaticity of harmonic priming. *J. Exp. Psychol.: Hum. Percept. Perform.* **27**, 1000–1011 (2001).
9. Margulis, E. H. Surprise and listening ahead: analytic engagements with musical tendencies. *Society for Music Theory* **29**, 197–217 (2007).
10. Parncutt, R. The tonic as triad: key profiles as pitch salience profiles of tonic triads. *Music Percept.* **28**, 333–365 (2011).
11. Lewandowsky, S., Duncan, M. & Brown, G. D. A. Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review* **11**, 771–790 (2004).
12. Dowling, W. J., Tillman, B. & Ayers, D. F. Memory and the experience of hearing music. *Music Percept.* **19**, 249–276 (2001).
13. Margulis, E. H. Silences in music are musical not silent: An exploratory study of context effects on the experience of musical pauses. *Music Percept.* **24**, 485–506 (2007).
14. Bailes, F. & Delbé, C. Long-term melodic expectation: The unexpected observation of distant priming effects. *Musicae Scientiae* **8**, 315–336 (2009).
15. Curtis, M. E. & Bharucha, J. J. Memory and musical expectation for tones in cultural context. *Music Percept.* **26**, 365–375 (2009).
16. Cuddy, L. L. & Thompson, W. F. Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Psychological Research* **54**, 51–59 (1992).
17. Krumhansl, C. L. & Kessler, E. J. Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys. *Psychol. Rev.* **89**, 334–365 (1982).
18. Norris, D. & McQueen, J. M. Shortlist B: A Bayesian model of continuous speech recognition. *Psychol. Rev.* **115**, 357–395 (2008).
19. Temperley, D. *Music and Probability.* (MIT Press, 2007).

Author contributions

F.B. and R.T.D. conceived the experiments, which were run by F.B. M.T.P. undertook information theoretic analysis of the stimuli, F.B. and R.T.D. all other analyses, and all authors contributed to writing the manuscript.

Additional information