

Dissociating sensory and cognitive theories of harmony perception through computational modeling

Peter M. C. Harrison¹ and Marcus T. Pearce²

^{1,2}*Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, UK*

¹p.m.c.harrison@qmul.ac.uk, ²marcus.pearce@qmul.ac.uk

Abstract

Two approaches exist for explaining harmonic expectation. The *sensory* approach claims that harmonic expectation is a low-level process driven by sensory responses to acoustic properties of musical sounds. Conversely, the *cognitive* approach describes harmonic expectation as a high-level cognitive process driven by the recognition of syntactic structure learned through experience. Many previous studies have sought to distinguish these two hypotheses, largely yielding support for the cognitive hypothesis. However, subsequent re-analysis has shown that most of these results can parsimoniously be explained by a computational model from the sensory tradition, namely Leman's (2000) model of auditory short-term memory (Bigand, Delbé, Poulin-Charronnat, Leman, & Tillmann, 2014). In this research we re-examine the explanatory power of auditory short-term memory models, and compare them to a new model in the Information Dynamics Of Music (IDyOM) tradition, which simulates a cognitive theory of harmony perception based on statistical learning and probabilistic prediction. We test the ability of these models to predict the surprisingness of chords within chord sequences ($N = 300$), as reported by a sample group of university undergraduates ($N = 50$). In contrast to previous studies, which typically use artificial stimuli composed in a classical idiom, we use naturalistic chord sequences sampled from a large dataset of popular music. Our results show that the auditory short-term memory models have remarkably low explanatory power in this context. In contrast, the new statistical learning model predicts surprisingness ratings relatively effectively. We conclude that auditory short-term memory is insufficient to explain harmonic expectation, and that cognitive processes of statistical learning and probabilistic prediction provide a viable alternative.

Introduction

Western tonal harmony obeys a complex set of conventions concerning how notes combine to form chords, and how chords combine to form chord sequences. These conventions are termed *harmonic syntax*. Listeners are sensitive to aspects of harmonic syntax, and implicitly judge the syntactic implications of successive chords as they hear a piece of music. This online sensitivity to harmonic syntax is termed *harmonic expectation*.

Theories of the psychological origins of harmonic expectation can be organized along a *sensory-cognitive* spectrum (Collins, Tillmann, Barrett, Delbé, & Janata, 2014). *Sensory* theories hold that harmonic expectation is driven by low-level sensory responses to acoustic properties of musical sounds. Conversely, *cognitive* theories posit that harmonic expectation is driven by high-level cognitive processes similar to those involved in processing linguistic syntax.

Many empirical studies have accumulated in favor of cognitive theories of harmonic expectation. These studies

typically report that listeners display sensitivity to syntactic violations that could not be detected by sensory cues alone. However, much of this work was recently undermined by Bigand et al. (2014): the authors reanalyzed data from 18 tonal expectation studies, including 17 harmonic expectation studies, and found that most results could be explained with a sensory model of auditory short-term memory (Leman, 2000). The conflict between sensory and cognitive theories therefore remains largely unresolved.

The present research takes a new approach to the problem of disentangling sensory and cognitive theories of harmonic expectation. We avoid the problematic task of hand-constructing sequences where sensory cues contradict syntactic rules, and avoid the problematically small stimulus sets associated with this approach. Instead, we separate sensory and cognitive theories through computational modeling, giving us the flexibility to use a large dataset of naturalistic stimuli derived from real music.

Many sensory and cognitive theories of harmonic expectation exist in the literature, but here we restrict comparison to two types that have received particular empirical support in recent decades. *Auditory short-term memory* theories explain expectation through the retention and comparison of auditory images in short-term memory. These theories are typically sensory in nature. *Statistical learning* theories, meanwhile, claim that harmonic expectation corresponds to probabilistic predictions made by listeners who have internalized the statistical structure of musical styles through exposure. Statistical learning provides a cognitive alternative to auditory short-term memory theories.

Several auditory short-term memory models exist in the literature. We evaluate three here: Leman's (2000) periodicity-pitch model, Milne, Sethares, Laney & Sharp's (2011) spectral distance model, and Collins et al.'s (2014) tonal expectation model.

Fewer statistical learning models exist in the literature. The Information Dynamics Of Music (IDyOM) model (Pearce, 2005, 2018) is a prominent candidate, but so far it has largely been limited to the melodic domain. We therefore introduce an extension of the model to harmonic expectation and compare it with the auditory short-term memory models.

Our model evaluation involves an explicit behavioral measure of harmonic expectation where listeners rate the surprisingness of particular chords within chord sequences. We evaluate each model in terms of its ability to predict these surprisingness ratings: good models should deliver accurate predictions.

Models

Leman's (2000) Periodicity-Pitch Model

Leman's (2000) model is a sensory auditory short-term memory model. It was recently shown to explain a wide variety of experimental findings previously thought to support a cognitive account of harmony perception (Bigand et al., 2014).

The model takes an audio signal as input and simulates the acoustic filtering of the outer and middle ear, the resonance of the basilar membrane in the inner ear, and the conversion of the resulting signal into neural rate-code patterns. A periodicity analysis produces a series of *pitch images* representing the instantaneous pitch patterns perceived at every point in the audio signal. The model then simulates the persistence of these pitch images in auditory short-term memory to produce *echoic pitch images*. Echoic pitch images are created by *leaky integration*, where the echoic pitch image at each timestep is created by adding the current non-echoic pitch image to the echoic image from the previous time-step. The length of the 'echo' is determined by a time constant defining the echo's half-life.

Two echoic pitch images are created: a *local* image and a *global* image. The *local* pitch image summarizes the pitch content of the immediate temporal context (e.g. the last 0.5 seconds), whereas the *global* pitch image summarizes pitch over a longer temporal context (e.g. 5 seconds). The moment-to-moment similarity between these local and global pitch images is summarized in a *tonal contextuality* profile. Applied to chord sequences, points of high tonal contextuality correspond to chords that are tonally consistent with their recent musical context, whereas low tonal contextuality reflects low tonal consistency.

For this study we defined the model's estimate of a chord's surprisingness as the negative mean tonal contextuality during the time that chord was playing. We used the MATLAB model implementation as created by the original author and available in the IPEM toolbox.¹ Decay constants for the local and global pitch images were set to 0.1 s and 1.5 s, as optimized in Leman (2000). All other parameters were left at their default values.

Milne et al.'s (2011) Spectral Distance Model

This is a second sensory model that also embodies an auditory short-term memory theory of tonal perception. It avoids much of the complexity of Leman's (2000) model: it has no explicit modeling of the peripheral auditory system and does not model the time-course of echoic memory. Nonetheless, the model has demonstrated best-in-class results in modeling certain important results from the psychological literature (Milne & Holland, 2016; Milne, Laney, & Sharp, 2015).

Milne et al.'s (2011) model estimates the perceptual dissimilarity of pairs of pitch or pitch-class sets. It combines each harmonic series implied by every pitch(-class), smooths the resulting spectra to account for perceptual imprecision, and then computes the cosine distance between these spectra. This cosine distance has been shown to predict perceptual

judgements of triadic similarity rather effectively (Milne & Holland, 2016).

We applied this model to harmonic expectation by using it to model the perceptual dissimilarity of a chord and its context. The model's dissimilarity estimate was taken as an estimate of chord surprisingness. Context could be defined in several ways; here we defined the context as the immediately preceding chord, but future work should explore alternative definitions.

The model has three important free parameters: the number of harmonics implied by each pitch(-class), the degree of amplitude roll-off as a function of harmonic number, and the degree of spectral smoothing. Additionally, there is freedom to choose either a pitch representation or a pitch-class representation. This study followed the model configuration as psychologically optimized in Milne and Holland (2016): 12 harmonics, roll-off parameter of 0.75, spectral smoothing parameter of 6.83, and a pitch-class representation. The model was given a new Common Lisp implementation, with a selection of results verified against the original author's MATLAB implementation.²

Collins et al.'s (2014) Tonal Expectation Model

While the two previous models (Leman, 2000; Milne et al., 2011) describe sensory accounts of tonal perception, this model describes an intermediate sensory-cognitive account of tonal perception. It is still an auditory short-term memory model, but some of its auditory representations involve cognitive abstractions.

The model centers on three representations of the musical input. The first is a *periodicity-pitch* representation corresponding to the pitch images of Leman's (2000) model. A *chroma-vector* representation is derived from the periodicity-pitch representation by collapsing pitches to pitch classes. Lastly, a *tonal space* representation is produced by projecting the periodicity-pitch representation onto a toroidal self-organizing map, after Janata et al. (2002). This last representation allows the model to learn a map of tonal space stored in long-term memory.

Incoming audio activates these representations in a cascade. These activations are blurred by echoic memory, analogous to Leman's (2000) model. For each of the three representations (periodicity pitch, chroma vector, tonal space), local and global images are created which evolve over the course of the stimulus. Local images are created by leaky integration with a short time constant (0.1 s); global images are produced by leaky integration with a longer time constant (4.0 s). These images summarize the recent activation of the respective representational space over the given time period.

Various features are then derived from these blurred activations. These features include both the correlation between local and global images (after Leman, 2000) and the peak activation within the global image. Extending Leman (2000), the model considers both absolute values of the correlations/peak activations and relative changes in these values. Different features are constructed that apply these computations to different time windows within the stimulus.

These features are then combined by linear regression, using coefficients optimized on a set of seven empirical

¹ <http://www.ipem.ugent.be/Toolbox>

² http://www.dynamictonality.com/probe_tone_files/

studies (Collins et al., 2014). The model's output corresponds to predicted deviations in reaction times in priming studies, with slower reaction times being thought to reflect lower expectedness. We therefore used these predicted reaction times as the model's estimate of chord surprisingness.

We used the original authors' implementation as available in the Janata lab music toolbox (JLMT).³ All parameters were left at their default values.

Information Dynamics Of Music

The Information Dynamics Of Music (IDyOM) model (Pearce, 2005, 2018) is a statistical learning model of musical expectation. It asserts that listeners acquire statistical knowledge of musical syntax through experience, and use this knowledge to generate probabilistic predictions as a piece unfolds.

The model takes a musical score as its input. It generates predictions in an incremental manner, predicting each event based on the preceding events. Its predictions are based on a variety of representations derived from the musical score, termed *viewpoints*. These viewpoints are intended to mirror the psychological representations of music available to listeners (e.g. Shepard, 1982). Predictions are generated by combining the output of a *long-term model*, which captures syntactic structure common to an entire musical style, and a *short-term model*, which captures statistical patterns heard so far in the current musical piece.

The original IDyOM model was restricted to melodies (Pearce, 2005). Here we extended the model to process chord sequences, substantially expanding the range of music to which the model can be applied.

The primary challenge in extending the IDyOM model to harmony is to develop new viewpoints for capturing harmonic structure. Our viewpoint set captures a variety of well-established psychological and music-theoretic principles of harmony, including inversion invariance, transposition invariance, chord roots, scale degrees, and intervallic relationships between successive chords. Chord roots are inferred using the General Chord Type algorithm of Cambouropoulos, Kaliakatsos-Papakostas, & Tsougras (2014), and scale degrees are inferred using the key-finding algorithm of Albrecht & Shanahan (2013). This explicit modeling of root-finding and key-finding improves cognitive realism compared to previous multiple-viewpoint models of harmony, which read key information and chord roots directly from the score (Hedges & Wiggins, 2016; Sears, Pearce, Caplin, & McAdams, 2018; Whorley & Conklin, 2016).

A further change introduced in this research was to replace the IDyOM model's heuristic entropy-weighting scheme with a supervised weighting scheme, where the contribution of each viewpoint is determined by a weight vector derived using maximum-likelihood estimation. The cognitive implications of this change are outside the scope of the present paper, but will be explored in future work.

Consistent with previous IDyOM research, we computed an estimate of surprisingness as the negative logarithm of the chord's conditional probability according to the statistical model. This quantity is termed *information content*.

The resulting model was implemented in the programming languages Common Lisp and R. The implementation extends the publicly available IDyOM codebase.⁴

Methods

Participants

Fifty psychology undergraduates (44 female, six male) participated in exchange for course credit or small financial reward. The mean age was 18.7 years ($SD = 1.7$). Most self-reported as frequent listeners to popular music. The mean musical training score as assessed by the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen, Gingras, Musil, & Stewart, 2014) was 15.1 ($SD = 8.0$), corresponding to the 22nd percentile of the original Gold-MSI sample.

Stimuli

Chord sequences were sourced from the Billboard corpus (Burgoyne, 2012). This dataset comprises a set of transcriptions of popular songs sampled from the Billboard magazine's United States "Hot 100" chart between 1958 and 1991. Three hundred eight-chord sequences were randomly sampled from this dataset, with repeated chords removed, under the constraint that no song appeared twice. Sequences were played with a piano timbre at a tempo of 60 beats per minute without metrical cues. Bass notes were played in the octave below middle C, non-bass notes in the octave above.

Procedure

Participants took the experiment individually in a quiet room at a desktop computer, navigating the experiment using keyboard and mouse. Audio was played over headphones.

The main part of the experiment comprised 150 trials for each participant. In each trial, the participant was played a sequence of eight chords and instructed to rate the sixth chord in this sequence, termed the *target*, for surprisingness. This chord was visually cued by a continuous clock-like animation, and surprisingness ratings were given on a scale from one to nine using the computer keyboard. Participants were given 10-second breaks every 25 trials. Each participant's 150 chord sequences were randomly chosen under the constraints that no participant heard the same sequence twice and that each of the 300 possible chord sequences was presented equally often to all participants over the course of the study.

The main part of the experiment was preceded by a training routine which included three practice trials. After the main part of the experiment, the participant completed a short questionnaire concerning basic demographic details and familiarity with popular music, and then completed the musical training component of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014). On average the procedure lasted approximately 40 minutes.

³ <http://atonal.ucdavis.edu/resources/software/jlmt/>

⁴ <https://code.soundsoftware.ac.uk/projects/idyom-project>

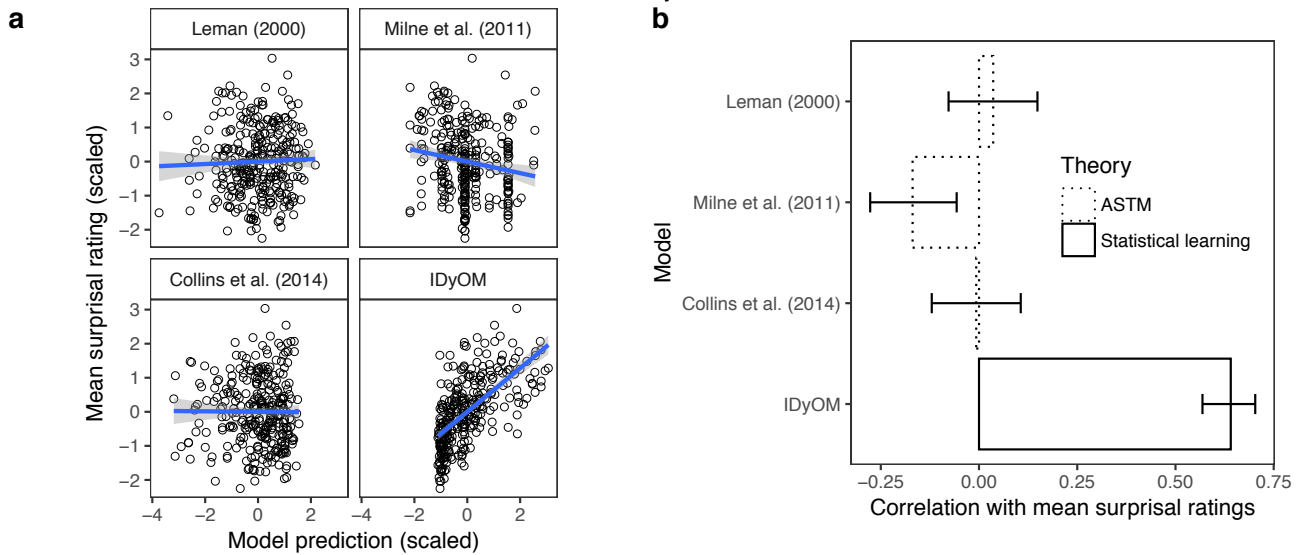


Figure 1. a) Scatterplots of mean surprisal rating (z-score) against model prediction (z-score) for the four computational models. Linear regression lines are plotted for each model, with the standard errors of these lines shaded in grey. b) Bar chart displaying Pearson correlation coefficients for the relationship between each model's predictions and mean surprisal ratings. The error bars denote 95% confidence intervals. 'ASTM' stands for auditory short-term memory.

Results

Data Preprocessing

One participant was found to give the same response for all 150 stimuli, and so their data were removed from further analyses. The remaining 49 participants had their ratings standardized to z-scores to normalize across individual differences in scale usage, and then these z-scores were averaged across participants to produce a mean surprisal rating for each chord sequence. These surprisal ratings were then z-transformed again across all stimuli, so that the mean surprisal rating would be zero and the standard deviation one.

Each of the four computational models was applied to the target chord in each chord sequence. Model outputs were converted to z-scores to facilitate comparison across models.

Model Predictive Performances

The predictive performance of each model was assessed in terms of its Pearson correlation with mean surprisal ratings: a high correlation means that the model predicted perceived surprisal well. These results are summarized in Figure 1 and the last column of Table 1. The three auditory short-term memory models did not display significant positive correlations with mean surprisal ratings. Surprisingly, Milne et al.'s model displayed a significant correlation in the opposite direction to that predicted by theory: greater spectral distance was significantly associated with lower surprisal ratings ($r(298) = -.169, p = .003, 95\% \text{ CI} = [-.277, -.057]$). In contrast, IDyOM model outputs exhibited a moderately large positive correlation with surprisal ratings ($r(298) = .641, p < .001, 95\% \text{ CI} = [.569, .703]$). A linear regression model predicting surprisal ratings from the four computational models found a significant coefficient for the IDyOM model ($p < .001$) but non-significant coefficients for the remaining models (all p -values $> .3$).

Model Correlations

Table 1 displays pairwise correlations within the set of computational models. Lemman's (2000) model outputs and Milne et al.'s (2011) model outputs were fairly well correlated ($r(298) = .591, p < .001, 95\% \text{ CI} = [.512, .660]$). Surprisingly, Collins et al.'s (2014) model outputs were not significantly correlated with any of the other model outputs (all p -values $> .3$). IDyOM model outputs were significantly negatively correlated with Milne et al.'s (2011) model outputs ($r(298) = -.203, p < .001, 95\% \text{ CI} = [-.309, -.092]$) but not significantly correlated with outputs of the other two models (both p -values $> .2$).

Table 1. Pairwise Pearson correlation coefficients for surprisal ratings and the four computational models.

	Milne et al. (2011)	Collins et al. (2014)	IDyOM	Surprisal ratings
Leman (2000)	.591	-.004	.071	.036
Milne et al. (2011)		.053	-.203	-.169
Collins et al. (2014)			.04	-.007
IDyOM				.641

Musical Examples

Figure 2 displays four specific chord sequences from the dataset; corresponding model outputs are displayed in Table 2. These examples were selected as follows: a) the most and b) the least surprising stimuli according to Lemman's (2000) model; c) the most and d) the least surprising stimuli according to the IDyOM model.

Sequence a) comprises solely bare fifths. The progression from ‘EB’ to ‘GD’ on the sixth chord is considered somewhat unsurprising by listeners ($z = -0.101$): in this context, ‘EB’ implies E minor, and ‘GD’ implies G major, so the passage is simply a version of the common progression i-III with the third of each chord removed. However, this passage is considered particularly unexpected by Lemán et al.’s (2000) model and Milne et al.’s (2011) model. This is presumably because the missing thirds are the two common tones between these chords; without them, the chords are rather acoustically dissimilar.

Sequence b) alternates between two inversions of the same major triad. The target chord is very spectrally similar with its previous context, yielding low surprisal ratings from Lemán et al.’s (2000) model and Milne et al.’s (2011) model. Surprisingly, Collins et al.’s (2014) gives a relatively high surprisal rating for this chord. As expected, the IDyOM model finds the chord relatively predictable, in large part because the same transition occurs several times in the stimulus.

Sequence c) begins with conventional progressions along the circle of fifths. However, the target chord is very tonally distant from its context: it corresponds to a semitone displacement of the previous chord with a tonic pedal in the bass. Correspondingly, most of the models predict high surprisal. The exception again is Collins et al.’s (2014), which predicts only moderate surprisal.

Sequence d) was considered very unsurprising by the participants and by the IDyOM model, but very surprising by the auditory short-term memory models. It corresponds to a major-mode IV-V progression, which is very common in Western popular music. The IDyOM model therefore finds the progression very predictable, because similar progressions occur many times in the corpus and even at the start of the same chord sequence. However, the two chords are not particularly similar spectrally speaking, and so they are considered surprising by the auditory short-term memory models.

Table 2. Surprisal ratings as predicted by the models and as reported by the participants. All scores are z-scores where higher scores correspond to higher surprisal.

	Lemán (2000)	Milne et al. (2011)	Collins et al. (2014)	IDyOM	Participants
a)	2.17	2.51	0.49	0.06	-0.10
b)	-3.74	-2.16	0.35	-0.21	-1.50
c)	1.00	0.56	0.02	3.07	1.28
d)	1.35	1.54	0.79	-1.11	-1.56

Discussion

We tested two competing explanations of harmonic expectation: an auditory short-term memory explanation and a statistical learning explanation. According to the former, harmonic expectation is a low-level process driven by the accumulation of auditory images in short-term memory. According to the latter, harmonic expectation reflects

Figure 2. Four example stimuli. Target chords are indicated by asterisks. Sequence a) is the most surprising stimulus in the corpus, according to Lemán’s (2000) model. Sequence b) is the least surprising according to Lemán’s (2000) model. Sequence c) is the most surprising according to the IDyOM model. Sequence d) is the least surprising according to the IDyOM model. Precise model outputs and participant ratings are given in Table 2.

probabilistic predictions of listeners which derive from internalized statistical knowledge about musical styles.

The results were unambiguous. None of the auditory short-term memory models produced statistically significant correlations with surprisal ratings in the direction predicted by theory. In contrast, the statistical learning model predicted surprisal ratings moderately well. The results therefore strongly corroborate the statistical learning account over the auditory short-term memory account.

It is plausible that the statistical learning model might outperform the auditory short-term memory models, but highly surprising that the latter models should not outperform chance. Pre-existing literature gives the impression that harmonic syntax is ineluctably correlated with spectral similarity, with the result that auditory short-term memory models can explain the results of most existing harmonic expectation studies (Bigand et al., 2014). However, we found that these models had no explanatory power for our dataset. Moreover, their predictions did not correlate positively with the predictions of the statistical learning model, suggesting that the correlation between spectral similarity and harmonic syntax was minimal at best.

We suggest several possible reasons for this discrepancy. First, while previous studies typically used stimuli in the style

of Western classical music, this study used stimuli from Western popular music, where spectral distance seems to be a worse predictor of harmonic style (Harrison & Pearce, 2018). Second, many previous studies tested perception of final chords in sequences, whereas we tested the perception of non-final chords. Stylistic expectations will differ between these contexts. Third, much of the harmonic expectation literature relies on harmonic priming paradigms with the listener's task being to detect tuning or timbre deviants. Both of these tasks involve detecting deviations in acoustic spectra, and it is clear that this might be facilitated by maximizing spectral similarity between successive chords. Reliance on this paradigm might therefore overemphasize the role played by spectral similarity (and correspondingly auditory short-term memory) in harmonic expectation. Fourth, we have yet to exhaust the potential of each computational model. Leman's (2000) model might perform better with different time constants (e.g. Bigand et al., 2014). Milne et al.'s (2011) model might be improved by incorporating inharmonic partials, different degrees of spectral smoothing, or a continuously decaying echoic memory. The predictors used in Collins et al.'s (2014) model might still have useful explanatory power, even if the regression model doesn't generalize well. Likewise, many computational aspects of the harmonic IDyOM model remain to be psychologically optimized.

We intend to explore these computational models further in ongoing research. However, as the results stand, it seems that auditory short-term memory is insufficient to explain harmonic expectation. We have shown that one viable alternative is statistical learning. However, several other alternatives exist at the sensory end of the sensory-cognitive spectrum. In particular, earlier literature has emphasized the importance of roughness and voice-leading distance in harmonic expectation, both of which correspond to relatively low-level psychological processes (Bigand, Parncutt, & Lerdahl, 1996). Perhaps it is time to re-examine these alternative sensory models.

Acknowledgements. The authors would like to thank Emmanouil Benetos and Matthew Purver for useful feedback and advice regarding this project. PH is supported by a doctoral studentship from the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

References

- Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception*, *31*(1), 59–67.
- Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8*. <https://doi.org/10.3389/fnsys.2014.00094>
- Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*(1), 124–141.
- Burgoyne, J. A. (2012). *Stochastic Processes & Database-Driven Musicology*. (Unpublished doctoral dissertation). McGill University, Montréal, Québec, Canada.
- Cambouropoulos, E., Kaliakatsos-Papakostas, M., & Tsougras, C. (2014). An idiom-independent representation of chords for computational music analysis and generation. In *Proceedings of the joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC)*. Athens, Greece.
- Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, *121*(1), 33–65.
- Harrison, P. M. C., & Pearce, M. T. (2018). An energy-based generative sequence model for testing sensory theories of Western harmony. In *Proceedings of the 19th International Society for Music Information Retrieval Conference*. Paris, France.
- Hedges, T., & Wiggins, G. A. (2016). The prediction of merged attributes with multiple viewpoint systems. *Journal of New Music Research*, *45*(4), 314–332.
- Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, *298*(5601), 2167–2170.
- Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*(4), 481–509.
- Milne, A. J., & Holland, S. (2016). Empirically testing Tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, *10*(1), 59–85.
- Milne, A. J., Laney, R., & Sharp, D. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception*, *32*(4), 364–393.
- Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, *5*(1), 1–20.
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), 1–23.
- Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. (Unpublished doctoral dissertation). City University, London.
- Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1–18.
- Sears, D. R. W., Pearce, M. T., Caplin, W. E., & McAdams, S. (2018). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, *47*(1), 29–52.
- Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of Music* (pp. 343–390). New York, NY: Academic Press.
- Whorley, R. P., & Conklin, D. (2016). Music generation from statistical models of harmony. *Journal of New Music Research*, *45*(2), 160–183.