

# Time-series analysis of Music: Perceptual and Information Dynamics

Marcus T. Pearce[1]  
Goldsmiths, University of London, UK

**ABSTRACT:** Dean and Bailes (2010) provide a tutorial on the use of time-series analysis in research on music perception and a study of the influence of acoustic factors on real-time perception of music. They illustrate their approach with a detailed case study of an electroacoustic composition by Trevor Wishart. In this commentary, I discuss four aspects of Dean and Bailes' presentation: first, the importance of focusing on dynamic changes in musical structure; second, the benefits of computer-generated music for research on music perception; third, the need for caution in averaging responses from multiple listeners; and finally, the role of time-series analysis in understanding computational information-dynamic models of music cognition.

Submitted 2011 August 5; accepted 2011 August 9.

**KEYWORDS:** *perception, emotion, structure, modelling, information dynamics*

## THE DYNAMICS OF MUSICAL EXPERIENCE

GIVEN that music is fundamentally temporal in nature, it is surprising how much research in music perception and cognition has focused on static analyses of the experience of musical works. On the cognitive side, for example, Lerdahl and Jackendoff (1983) focus on the final state of understanding of a piece of music by a listener, while research on musical emotions has tended to focus on the static expression of particular emotions by musical excerpts without studying dynamic changes in the ongoing emotional experience (see, e.g., Gabrielsson & Lindström, 2010, for a review).

Dean and Bailes (2010) focus explicitly on the dynamic real-time changes in the listening experience, continuing in the tradition of Krumhansl (1997), Schubert (2001) and others. Time-series analysis provides an elegant and powerful way of relating continuous dynamic changes in psychological experience (or at least the observable behavioural and physiological effects of these changes) to continuous dynamic changes in the structure of the music. This is especially important when considering the induction of emotion by unfamiliar, instrumental music where affective responses are most likely to be linked to dynamic changes in musical and acoustic structure reflecting low-level brainstem responses and high-level effects of expectation (Juslin & Västfjäll, 2008). There is a pressing need to go beyond static representations of musical attributes such as tempo and mode in understanding how musical passages lead to emotional states differing in valence and arousal as well as more aesthetic emotions such as awe and frisson (Huron, 2006)

## COMPUTER-GENERATED MUSIC

Dean and Bailes (2010) give two reasons for using electroacoustic music in their study: first, to demonstrate that their methods generalise beyond Western tonal music which is more often used in empirical work on music perception; and second, *Red Bird* provides an opportunity to test their methods on idiosyncratic temporally-localised timbral features in addition to the continuous features which generalise to other musical genres (see, e.g., Dean, Bailes & Schubert, 2011). Interestingly, their timbre feature of choice is *spectral flatness*, which they view as a more global indicator of timbre than *spectral centroid*, which is more commonly used in research on music perception (though this is not true of research on audio signal processing and music information retrieval where spectral flatness is one of the standard descriptors used in the MPEG 7 standard).

Dean and Bailes also argue that electroacoustic music can be algorithmically generated in such a way that the acoustic and algorithmic parameters of interest are systematically varied in creating stimuli for research on music perception. In other work, for example, Dean et al. (2011) extend their approach to the effects of intensity on arousal in two pieces written by Roger Dean, one of which is composed in the minimalist style. We might legitimately ask what advantage such algorithmically generated music has over the stimuli often constructed artificially to create experimental conditions in empirical research on music perception. The most obvious advantage is that the results gain in ecological validity from using stimuli created by composers, using stylistically legitimate methods, with an artistic purpose. These results should generalise to the experience of similar music outside the laboratory, while results obtained with artificially created or altered musical stimuli are not guaranteed to do so. The advantage of computer-generated music over other musical styles is that it can be produced so as to conserve experimental control. This is an argument we have made previously in the context of our research on information-dynamic processes in minimalist music (Abdallah & Plumbley, 2009; Potter, Wiggins, & Pearce, 2007; Wiggins, Pearce, & Müllensiefen, 2009) which has high ecological validity as a significant tradition in Western art music.

On a more general note, empirical work on music would benefit from greater interdisciplinary collaboration between music psychologists and musicians in the creation of stimuli and hypotheses to avoid falling foul of both musical and scientific naïveté.

### MULTIPLE INTERPRETATIONS

Dean and Bailes (2010) discuss their decision to average series of perceptual responses across participants and point to future work in which they analyze the participants' responses individually. It can be dangerous to average data from individual participants without first checking for inter-participant agreement. In the present case, imagine that two listeners have contrasting valence responses to a particular passage in *Red Bird*; one experiences high valence, the other low valence. There are very good reasons why this might happen if, for example, the listeners have different stylistic experiences and preferences. Averaging their responses produces a moderate valence score which incorrectly reflects the experience of both listeners.

However, low inter-participant agreement need not necessarily signal the end of the analysis. A complementary approach to aggregating the dependent variable over participants is to explicitly examine possible causes of the lack of agreement. We have recently followed this approach in a study of melodic segmentation which suggested that any given melody can yield several different grouping structures corresponding to different perceptual strategies employed by different groups of participants (Pearce, Müllensiefen, & Wiggins, 2010). It remains to be seen whether the dynamics of emotion and change perception studied by Dean and Bailes (2010) exhibit multiple interpretations in similar way.

### COGNITIVE MODELS AND INFORMATION DYNAMICS

Dean and Bailes (2010) use the term *model* to refer to a time-series analysis of the relationships between particular series. Here a computer program is used as an analytical tool for testing a hypothesis: i.e., that changes in loudness generate changes in arousal in the listener.

There is a very different use of the word *model* in cognitive science to describe a computer program which is intended to reflect, and therefore explain, some aspect of cognitive processing. In this case, the program itself embodies hypotheses not about relationships between stimulus structure and subjective experience (or its observable concomitants) but about the cognitive processes involved in generating the subjective experience from the perceptual input. In this case, for example, we might write a program which embodies hypotheses about the cognitive processes responsible for increasing arousal when presented with auditory stimuli that increase in loudness.

In their discussion, Dean and Bailes (2010) review an ongoing collaboration with me and my colleagues in London, involving time-series analysis and information-dynamic models of music perception. Here the time-series models are analytical models while the information-dynamic models are cognitive models. In the information-dynamic approach, perception is viewed as a process of ongoing prediction, in which surrounding context informs our expectations and the actual outcomes (predicted or not) stimulate learning which informs future predictions (e.g. Bubic, von Cramon, & Schubotz, 2010; Friston, 2005; Levy, 2008; Pearce & Wiggins, 2011; Reynolds, Zacks, & Braver, 2007; Summerfield & Egner, 2009). Music is an

example of a perceptual domain in which these domain-general processes operate; the vast amount of experience that most individuals have with music means that we have highly developed, though implicit, predictive models which makes music an ideal domain for studying these cognitive processes. Expectations are modelled as conditional probability distributions over some aspect of forthcoming musical structure, such as the pitch, timing or loudness of the next note given both the previous musical context and a lifetime's listening to music. Such distributions yield formal information-theoretic quantities such as *entropy* and *information content* which can be linked directly to subjective experiences such as *uncertainty* and *unexpectedness*.<sup>2</sup> It has been demonstrated empirically that information-dynamic models predict listener's pitch expectations in a range of melodic contexts and styles better than existing models based on static rules (Pearce & Wiggins, 2006; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

Information-theoretic quantities such as entropy and information content characterise the uncertainty and unexpectedness in actual sequences of events (which might consist, for example, of a melody, which is a sequence of pitched sounding notes, or at a lower-level, a sequence of acoustic features computed over time-frames of an audio signal). It is also interesting to consider complex information-dynamic measures which characterise properties of the sequential random *process* that generates sequences of events (e.g., we might consider a musical style as a process that generates melodies in that style). One such information measure is the *information rate* (Dubnov, 2006) which reflects the average amount of information to be gained from the past about the present in a sequential random process. As noted by Dean & Bailes, (2010), the spectral flatness measure (SFM) they use is related to the information rate assuming that the frequency bands of the log-magnitude spectrogram of the audio signal are linear combinations of independent Gaussian processes (Abdallah & Plumbley, 2009). In fact, Dubnov computes the information rate over three second windows as the sum of the SFM applied to the (decorrelated) cepstral coefficients of the audio signal computed over 200ms time windows. Dubnov, McAdams, and Reynolds (2006) report significant correlations between the information rate and listeners' ratings of emotional force in *The Angel of Death* by Roger Reynolds.

Abdallah and Plumbley (2009) examine the application of a number of information measures (including the information rate, which they refer to as *redundancy*) to probabilistic Markov chains (digram models specifically) of symbolically represented melody. They introduce the *instantaneous predictive information* (IPI) which characterises the new information that an event carries about future events, given the previous events and the *predictive information rate* (PIR) which reflects the average of this quantity for a sequential random process. Abdallah and Plumbley (2009) investigate the extent to which these measures reflect the human experience of 'interestingness'. Processes with minimum entropy rates would generate completely deterministic sequences while those with maximum entropy rates would generate completely unpredictable sequences, both of which would be considered 'uninteresting' by a human listener. This implies some kind of inverted-U shaped function linking entropy rate (as a measure of uncertainty or complexity) and interestingness (or pleasantness) following theoretical considerations in empirical aesthetics (Berlyne, 1974). Abdallah and Plumbley (2009) show that for first-order Markov chains, PIR shows such an inverted-U shaped relationship with entropy rate while the information rate of Dubnov (2006) does not. Much remains to be done both in terms of our theoretical understanding of these information-dynamic measures and their relationship to cognitive and neural processes in human perception of music (see, e.g., Pearce et al., 2010). The time-series methods used by Dean and Bailes (2010) are perfectly suited to an analysis of how dynamic changes in cognitive, perceptual and emotional processing are affected by dynamic changes in information-dynamic models, applied both to the acoustic level as in the target article, but also at the symbolic level.

The decision to focus on isochronous melody in some (but not all, see Pearce & Wiggins, 2006) of this research is based on experimental concerns (to ensure that the expectations we elicit from listeners and the model are based on pitch alone) rather than inherent limitations of the model. In the spirit of scientific reductionism, we decided to fully understand and test the model in one domain (pitch) before examining its behaviour with respect to other musical features. We have, however, investigated the effects of temporal structure on pitch expectations (Pearce & Wiggins, 2006) and we are currently exploring the extent to which information-dynamic models can account for listeners' temporal (Pearce, Müllensiefen, et al., 2010) and harmonic (Whorley, Wiggins, Rhodes, & Pearce, 2010) expectations in music. As noted by Dean and Bailes (2010), there are subtle issues involved in defining the alphabets over which the models compute their probability distributions; in particular, it is important that these reflect realistic perceptual and cognitive constraints such as those outlined in the target article. When looking at polyphonic music, it is

also necessary to address the question of when and how listeners segregate the auditory input into multiple parallel streams (Bregman, 1990) when constructing appropriate contexts for the models.

As the name implies, dynamic changes in information-theoretic quantities (and the experiences they model) as a result of changes in the structure of the auditory input are central to the information-dynamic approach (Pearce & Wiggins, 2006; Pearce, Müllensiefen, et al., 2010; Potter et al., 2007). In this context, time-series analysis as presented by Dean and Bailes (2010) will prove invaluable in investigating the relationships between series reflecting listeners' experience (e.g., behavioural ratings, reaction times or physiological measures), series reflecting aspects of performance (e.g., expressive timing and dynamics) and series reflecting the model's experience (e.g., the degree of uncertainty or unexpectedness), each recorded at each time point in the same compositions. Time-series analysis has several advantages over other methods in being able to rigorously identify (Granger) causal relationships between series and the magnitude of the lead or lag in these relationships.

## CONCLUSIONS

Dean and Bailes (2010) present both a tutorial and a genuine research contribution on the use of time-series analysis for examining real-time dynamic relationships between the changing structure of the auditory input and the ongoing musical experience. This is especially important for longer excerpts of unfamiliar, instrumental music where static emotional labels are unable to fully capture the changing nature both of the music and of the musical experience. In studying such dynamic responses, computer-generated music can be useful for striking a balance between experimental control and ecological validity, especially when the music is created with an aesthetic goal in mind (rather than a purely experimental one). One potential disadvantage of computer-generated music could be its relative inaccessibility to some individuals, which raises the question of individual differences in the experience of tension in Dean and Bailes' study, which remains an important topic for future research. Finally, it is necessary to differentiate analytic models such as the time-series models used by Dean and Bailes and cognitive models, which attempt computer implementation of a cognitive process to understand how it operates. I have introduced one particular set of cognitive models based on statistical learning and information dynamics, which are related to the spectral flatness measure used by Dean and Bailes. Future research will put time-series analysis to good use in further examining the relationships between the behaviour of these models and human responses in music perception and performance.

## NOTES

[1] Address for correspondence: Department of Computing, Goldsmiths, University of London, SE14 6NW, UK, <http://www.doc.gold.ac.uk/~mas01mtp>

[2] The foundations of modern information theory (e.g., MacKay, 2003) were developed by Claude Shannon in his seminal mathematical theory of communication (Shannon, 1948). Particularly relevant here, is the portion of Shannon's theory capturing discrete noiseless systems and their representation as stochastic Markov sources, the use of  $n$ -grams to estimate the statistical structure of the source and the development of entropy as a quantitative measure of the predictability of the source.

An  $n$ -gram model (of order  $n-1$ ) computes the conditional probability of an element  $e_i$  at index  $i \in \{n, \dots, j\}$  in a sequence  $e_1^j$  of length  $j$ , over an alphabet,  $\mathcal{E}$ , of elements from which the sequence is composed, given the preceding  $n-1$  elements,  $e_{i-n}^{i-1}$ :

$$p(e_i | e_{i-n}^{i-1}) = \frac{\text{count}(e_{i-n}^i)}{\text{count}(e_{i-n}^{i-1})} \quad (1)$$

where  $e_m^n$  is the contiguous subsequence (or substring) of sequence  $e$  between elements  $m$  and  $n$  inclusive,  $e_m$  is the element at index  $m$  of the sequence  $e$  and  $count(x)$  is the number of times that  $x$  appears in some corpus of sequences used to train the model. The  $n$ -gram model can be used to compute a conditional probability distribution over the alphabet  $\mathcal{E}$  governing the identity of the next element in the sequence.

Given an  $n$ -gram model, the *information content*,  $h(e_i | e_{i-n}^{i-1})$ , of a given element appearing at index  $i$  of a sequence  $e$  is defined as:

$$h(e_i | e_{i-n}^{i-1}) = \log_2 \frac{1}{p(e_i | e_{i-n}^{i-1})} \quad (2)$$

and can be interpreted as the degree of unexpectedness (or surprisal) associated with an event appearing in a particular context. Highly improbable elements are not predicted to occur and, therefore, provide information about the environment when they do, whilst highly probable outcomes are uninformative as they were predicted to occur anyway.

The *entropy* (or average information content) of the predictive context itself is given by:

$$H(e_{i-n}^{i-1}) = \sum_{e \in \mathcal{E}} p(e_i | e_{i-n}^{i-1}) h(e_i | e_{i-n}^{i-1}) \quad (3)$$

and represents the contextual uncertainty about the identity of the element at index  $i$  in the sequence. For example, entropy is maximal for a flat conditional distribution in which each element is equiprobable given the context.

## REFERENCES

- Abdallah, S. A., & Plumbley, M. D. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, Vol. 21, No.2, pp. 89-117.
- Berlyne, D. E. (Ed.). (1974). *Studies in the New Experimental Aesthetics: Steps Towards an Objective Psychology of Aesthetic Appreciation*. Washington: Hemisphere Publishing Co.
- Bregman, A. S. (1990). *Auditory Scene Analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bubic, A., von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4, 25. doi:10.3389/fnhum.2010.00025
- Dean, R. T., & Bailes, F. (2010). Time series analysis as a method to examine acoustical influences on real-time perception of music. *Empirical Musicology Review*, Vol. 5, No. 4, pp. 152–175.
- Dean, R. T., Bailes, F., & Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation. *PLoS one*, Vol. 6, No. 4, e18591.
- Dubnov, S., McAdams, S., & Reynolds, R. (2006). Structural and Affective Aspects of Music from Statistical Audio Signal Analysis. *Journal of the American Society for Information Science and Technology*, Vol. 57, No. 11, pp. 1526 - 1536.
- Dubnov, S. (2006). Spectral Anticipations. *Computer Music Journal*, Vol. 30, No. 2, pp. 63–83.
- Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, Vol. 360, No. 1456, pp. 815-836.
- Gabrielsson, A. & Lindström, E. (2010). *The role of structure in the musical expression of emotions*. In P.N. Juslin & J.A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 367-400). Oxford: Oxford University Press.

- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, Vol. 31, No. 5, pp. 559-575.
- Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology*, Vol 51, No. 4, pp. 336-53.
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, Volume 16, No. 3, pp. 1126-1177.
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.
- Pearce, M. T., & Wiggins, G. A. (2006). Expectation in Melody: The Influence of Context and Learning. *Music Perception*, Vol. 23, No. 5, pp. 377-405.
- Pearce, M. T., & Wiggins, G. A. (2011). Auditory expectation: The information-dynamics of music perception and cognition. *TopiCS in Cognitive Science*. In press.
- Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception*, Vol. 39, No. 10, pp. 1365-1391.
- Pearce, M. T., Ruiz, M. H., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage*, Vol. 50, No. 1, pp. 302-313.
- Potter, K., Wiggins, G. A., & Pearce, M. T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Scientiae*, Vol. 11, No. 2, pp. 295-322.
- Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, Vol. 31, No. 4, pp. 613-643.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379-423 and 623-656.
- Schubert, E. (2001). Continuous measurement of self-report emotional responses to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and Emotion* (pp. 393-414). Oxford: Oxford University Press.
- Smith, E. C., & Lewicki, M. S. (2006). Efficient Auditory Coding. *Nature*, Vol. 439, pp. 978-982.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, Vol. 13, No. 9, pp. 403-9. doi:10.1016/j.tics.2009.06.003
- Whorley, R., Wiggins, G. A., Rhodes, C. S., & Pearce, M. T. (2010). Development of Techniques for the Computational Modelling of Harmony. In V. D (Ed.), *Proceedings of the International Conference on Computational Creativity*. Lisbon, Portugal.
- Wiggins, G. A., Pearce, M. T., & Müllensiefen, D. (2009). Computational Modelling of Music Cognition and Musical Creativity. In R. Dean (Ed.), *The Oxford Handbook of Computer Music* (pp. 383-420). Oxford, UK: Oxford University Press.