Cognitive Music Modelling: an Information Dynamics Approach

Samer A. Abdallah, Henrik Ekeus, Peter Foster
Andrew Robertson and Mark D. Plumbley
Centre for Digital Music
Queen Mary University of London
Mile End Road, London E1 4NS
Email:

***Abstract*—People take in information when perceiving music. With it they continually build predictive models of what is going to happen. There is a relationship between information measures and how we perceive music. An information theoretic approach to music cognition is thus a fruitful avenue of research. In this paper, we review the theoretical foundations of information dynamics and discuss a few emerging areas of application.**

I. Introduction

A. Expectation and surprise in music

One of the effects of listening to music is to create expectations of what is to come next, which may be fulfilled immediately, after some delay, or not at all as the case may be. This is the thesis put forward by, amongst others, music theorists L. B. Meyer [1] and Narmour [2], but was recognised much earlier; for example, it was elegantly put by Hanslick [3] in the nineteenth century:

> 'The most important factor in the mental process which accompanies the act of listening to music, and which converts it to a source of pleasure, is ... the intellectual satisfaction which the listener derives from continually following and anticipating the composer's intentions—now, to see his expectations fulfilled, and now, to find himself agreeably mistaken.

An essential aspect of this is that music is experienced as a phenomenon that 'unfolds' in time, rather than being apprehended as a static object presented in its entirety. Meyer argued that musical experience depends on how we change and revise our conceptions *as events happen*, on how expectation and prediction interact with occurrence, and that, to a large degree, the way to understand the effect of music is to focus on this 'kinetics' of expectation and surprise.

Prediction and expectation are essentially probabilistic concepts and can be treated mathematically using probability theory. We suppose that when we listen to music, expectations are created on the basis of our familiarity with various styles of music and our ability to detect and learn statistical regularities in the music as they emerge, There is experimental evidence that human listeners are able to internalise statistical knowledge about musical structure, e.g. [4], [5], and also that statistical models can form an effective basis for computational analysis of music, e.g. [6]–[8].

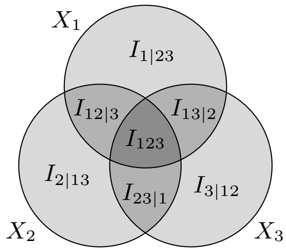
B. Music and information theory

With a probabilistic framework for music modelling and prediction in hand, we are in a position to apply Shannon's quantitative information theory [9]. The relationship between information theory and music and art in general has been the subject of some interest since the 1950s [1], [10]–[14]. The general thesis is that perceptible qualities and subjective states like uncertainty, surprise, complexity, tension, and interestingness are closely related to information-theoretic quantities like entropy, relative entropy, and mutual information. Berlyne [15] called such quantities 'collative variables', since they are to do with patterns of occurrence rather than medium-specific details, and developed the ideas of 'information aesthetics' in an experimental setting.

C. Information dynamic approach

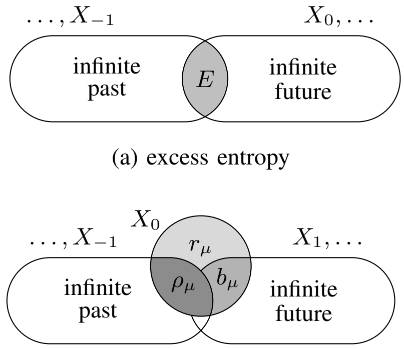
Bringing the various strands together, our working hypothesis is that as a listener (to which will refer as 'it') listens to a piece of music, it maintains a dynamically evolving probabilistic model that enables it to make predictions about how the piece will continue, relying on both its previous experience of music and the immediate context of the piece. As events unfold, it revises its probabilistic belief state, which includes predictive distributions over possible future events. These can be characterised in terms of a handful of information theoretic-measures such as entropy and relative entropy. By tracing the evolution of a these measures, we obtain a representation which captures much of the significant structure of the music.

One of the consequences of this approach is that regardless of the details of the sensory input or even which sensory modality is being processed, the resulting analysis is in terms of the same units: quantities of information (bits) and rates of information flow (bits per second). The probabilistic and information theoretic concepts in terms of which the analysis is framed are universal to all sorts of data. In addition, when adaptive probabilistic models are used, expectations are created mainly in response to to *patterns* of occurence, rather the details of which specific things occur. Together, these suggest that an information dynamic analysis captures a high level of *abstraction*, and could be used to make structural comparisons between different temporal media, such as music, film, animation, and dance.



$$
\begin{aligned}
I_{1|23} &= H(X_1|X_2,X_3) \\
I_{13|2} &= I(X_1;X_3|X_2) \\
I_{1|23}+I_{13|2} &= H(X_1|X_2) \\
I_{12|3}+I_{123} &= I(X_1;X_2)
\end{aligned}
$$

Fig. 1. I-diagram visualisation of entropies and mutual informations for three random variables X_1, X_2 and X_3. The areas of the three circles represent $H(X_1)$, $H(X_2)$ and $H(X_3)$ respectively. The total shaded area is the joint entropy $H(X_1, X_2, X_3)$. The central area I_{123} is the co-information [17]. Some other information measures are indicated in the legend.



(a) excess entropy

(b) predictive information rate b_μ

Fig. 2. I-diagrams for several information measures in stationary random processes. Each circle or oval represents a random variable or sequence of random variables relative to time $t = 0$. Overlapped areas correspond to various mutual information as in Fig. 1. In (c), the circle represents the 'present'. Its total area is $H(X_0) = H(1) = \rho_\mu + r_\mu + b_\mu$, where ρ_μ is the multi-information rate, r_μ is the residual entropy rate, and b_μ is the predictive information rate. The entropy rate is $h_\mu = r_\mu + b_\mu$.

Another consequence is that the information dynamic approach gives us a principled way to address the notion of *subjectivity*, since the analysis is dependent on the probability model the observer starts off with, which may depend on prior experience or other factors, and which may change over time. Thus, inter-subject variablity and variation in subjects' responses over time are fundamental to the theory.

II. Theoretical review

A. Entropy and information

Let X denote some variable whose value is initially unknown to our hypothetical observer. We will treat X mathematically as a random variable, with a value to be drawn from some set (or *alphabet*) $\mathcal{A}$ and a probability distribution representing the observer's beliefs about the true value of X. In this case, the observer's uncertainty about X can be quantified as the entropy of the random variable $H(X)$. For a discrete variable with probability mass function $p : \mathcal{A} \to [0, 1]$, this is

$$H(X) = \sum_{x \in \mathcal{A}} -p(x) \log p(x) = \mathrm{E} - \log p(X), \qquad (1)$$

where E is the expectation operator. The negative-log-probability $\ell(x) = -\log p(x)$ of a particular value x can usefully be thought of as the *surprisingness* of the value x should it be observed, and hence the entropy is the expected surprisingness.

Now suppose that the observer receives some new data $\mathcal{D}$ that causes a revision of its beliefs about X. The *information* in this new data *about* X can be quantified as the Kullback-Leibler (KL) divergence between the prior and posterior distributions $p(x)$ and $p(x|\mathcal{D})$ respectively:

$$\mathcal{I}_{\mathcal{D}\to X} = D(p_{X|\mathcal{D}} || p_X) = \sum_{x \in \mathcal{A}} p(x|\mathcal{D}) \log \frac{p(x|\mathcal{D})}{p(x)}. \qquad (2)$$

If there are multiple variables X_1, X_2 etc. which the observer believes to be dependent, then the observation of one may change its beliefs and hence yield information about the others. The relationships between the various joint entropies, conditional entropies, mutual informations and conditional mutual informations can be visualised in Venn diagram-like *information diagrams* or I-diagrams [16], for example, the three-variable I-diagram in fig. 1.

B. Entropy and information in sequences

Suppose that $(\ldots, X_{-1}, X_0, X_1, \ldots)$ is a stationary sequence of random variables, infinite in both directions, and that μ is the associated shift-invariant probability measure over all configurations of the sequence—in the following, μ will simply serve as a label for the process. We can indentify a number of information-theoretic measures meaningful in the context of a sequential observation of the sequence, during which, at any time t, there is 'present' X_t, a 'past' $\overleftarrow{X}_t \equiv (\ldots, X_{t-2}, X_{t-1})$, and a 'future' $\overrightarrow{X}_t \equiv (X_{t+1}, X_{t+2}, \ldots)$. Since the sequence is assumed stationary, we can without loss of generality, assume that $t = 0$ in the following definitions.

The *entropy rate* of the process is the entropy of the next variable X_t given all the previous ones.

$$h_\mu = H(X_0 | \overleftarrow{X}_0). \qquad (3)$$

The entropy rate gives a measure of the overall randomness or unpredictability of the process.

The *multi-information rate* ρ_μ (following Dubnov's [19] notation for what he called the 'information rate') is the mutual information between the 'past' and the 'present':

$$\rho_\mu(t) = I(\overleftarrow{X}_t; X_t) = H(X_0) - h_\mu. \qquad (4)$$

It is a measure of how much the context of an observation (that is, the observation of previous elements of the sequence) helps in predicting or reducing the suprisingness of the current observation.

The *excess entropy* [18] is the mutual information between the entire 'past' and the entire 'future':

$$E = I(\overleftarrow{X}_0; X_0, \overrightarrow{X}_0). \qquad (5)$$

The *predictive information rate* (or PIR) [21] is the average information in one observation about the infinite future given

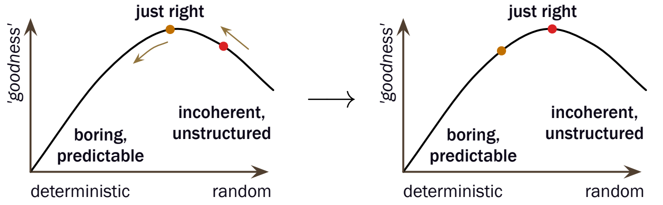


Fig. 3. The Wundt curve relating randomness/complexity with perceived value. Repeated exposure sometimes results in a move to the left along the curve [15].

the infinite past, and is defined as a conditional mutual information:

$$b_\mu = I(X_0; \overrightarrow{X}_0 | \overleftarrow{X}_0) = H(\overrightarrow{X}_0 | \overleftarrow{X}_0) - H(\overrightarrow{X}_0 | X_0, \overleftarrow{X}_0). \quad (6)$$

Equation (6) can be read as the average reduction in uncertainty about the future on learning X_t, given the past. Due to the symmetry of the mutual information, it can also be written as

$$I(X_t; \overrightarrow{X}_t | \overleftarrow{X}_t) = H(X_t | \overleftarrow{X}_t) - H(X_t | \overrightarrow{X}_t, \overleftarrow{X}_t). \quad (7)$$

Now, in the shift-invariant case, $H(X_t|\overleftarrow{X}_t)$ is the familiar entropy rate h_μ, but $H(X_t|\overrightarrow{X}_t, \overleftarrow{X}_t)$, the conditional entropy of one variable given *all* the others in the sequence, future as well as past, is what we called the *residual entropy rate* r_μ in [22], but was previously identified by Verdú and Weissman [23] as the *erasure entropy rate*. Thus, the PIR is the difference between the entropy rate and the erasure entropy rate: $b_\mu = h_\mu - r_\mu$. These relationships are illustrated in Fig. 2, along with several of the information measures we have discussed so far.

C. Other sequential information measures

James et al [24] study the predictive information rate and also examine some related measures. In particular they identify the σ_μ, the difference between the multi-information rate and the excess entropy, as an interesting quantity that measures the predictive benefit of model-building (that is, maintaining an internal state summarising past observations in order to make better predictions). They also identify $w_\mu = \rho_\mu + b_\mu$, which they call the *local exogenous information*.

D. First order Markov chains

These are the simplest non-trivial models to which information dynamics methods can be applied. In [21] we, showed that the predictive information rate can be expressed simply in terms of the entropy rate of the Markov chain. If we let a denote the transition matrix of the Markov chain, and h_a it's entropy rate, then its predictive information rate b_a is

$$b_a = h_{a^2} - h_a, \quad (8)$$

where $a^2 = aa$, the transition matrix squared, is the transition matrix of the 'skip one' Markov chain obtained by leaving out every other observation.

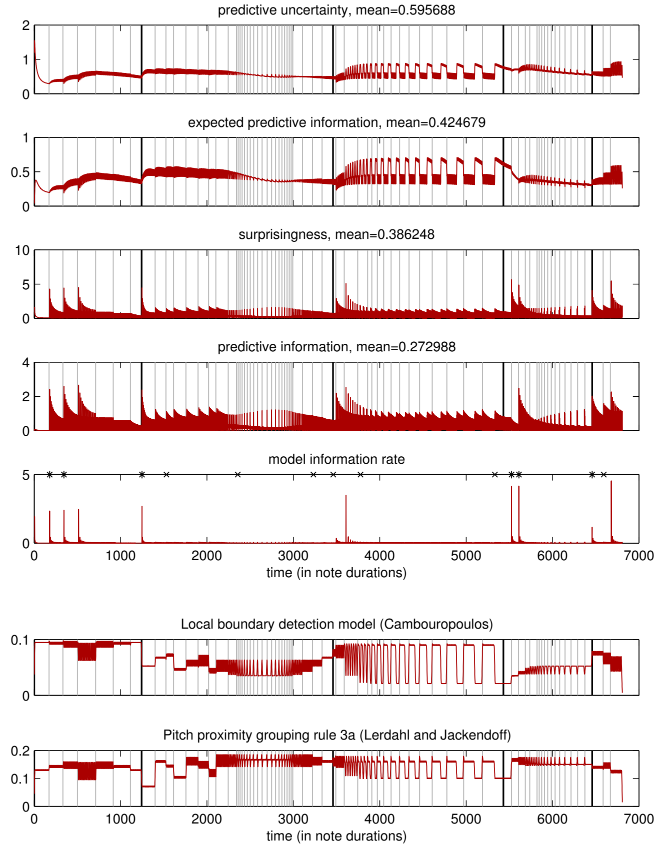


Fig. 4. Analysis of *Two Pages*. The thick vertical lines are the part boundaries as indicated in the score by the composer. The thin grey lines indicate changes in the melodic 'figures' of which the piece is constructed. In the 'model information rate' panel, the black asterisks mark the six most surprising moments selected by Keith Potter. The bottom panel shows a rule-based boundary strength analysis computed using Cambouropoulos' LBDM. All information measures are in nats and time is in notes.

E. Higher order Markov chains

Second and higher order Markov chains can be treated in a similar way by transforming to a first order representation of the high order Markov chain. If we are dealing with an Nth order model, this is done forming a new alphabet of possible observations consisting of all possible N-tuples of symbols from the base alphabet. An observation in this new model represents a block of N observations from the base model. The next observation represents the block of N obtained by shift the previous block along by one step. The new Markov of chain is parameterised by a sparse $K^N \times K^N$ transition matrix $\hat{a}$.

$$b_{\hat{a}} = h_{\hat{a}^{N+1}} - N h_{\hat{a}}, \quad (9)$$

where $\hat{a}^{N+1}$ is the $N+1$th power of the transition matrix.

III. Information Dynamics in Analysis

A. Musicological Analysis

refer to the work with the analysis of minimalist pieces

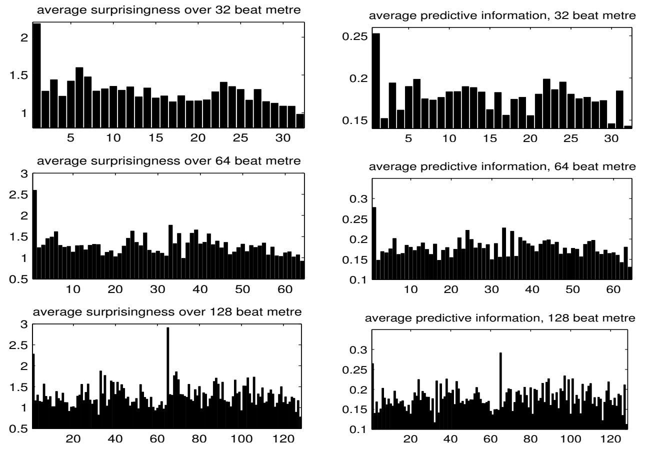


Fig. 5. Metrical analysis by computing average surprisingness and informative of notes at different periodicities (i.e. hypothetical bar lengths) and phases (i.e. positions within a bar).

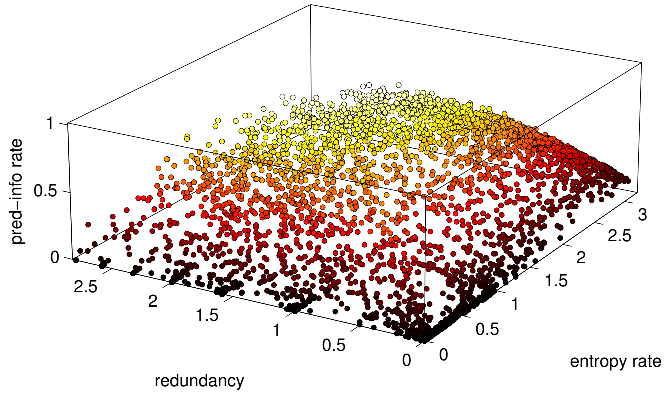


Fig. 6. The population of transition matrices distributed along three axes of redundancy, entropy rate and predictive information rate (all measured in bits). The concentrations of points along the redundancy axis correspond to Markov chains which are roughly periodic with periods of 2 (redundancy 1 bit), 3, 4, etc. all the way to period 8 (redundancy 3 bits). The colour of each point represents its PIR—note that the highest values are found at intermediate entropy and redundancy, and that the distribution as a whole makes a curved triangle. Although not visible in this plot, it is largely hollow in the middle.

B. *Content analysis/Sound Categorisation*

. Using Information Dynamics it is possible to segment music. From there we can then use this to search large data sets. Determine musical structure for the purpose of playlist navigation and search. *Peter*

C. *Beat Tracking*

Andrew

IV. Information dynamics as compositional aid

In addition to applying information dynamics to analysis, it is also possible use this approach in design, such as the composition of musical materials. By providing a framework for linking information theoretic measures to the control of generative processes, it becomes possible to steer the output of these processes to match a criteria defined by these measures. For instance outputs of a stochastic musical process could be filtered to match constraints defined by a set of information theoretic measures.

The use of stochastic processes for the generation of musical material has been widespread for decades – Iannis Xenakis applied probabilistic mathematical models to the creation of musical materials, including to the formulation of a theory of Markovian Stochastic Music. However we can use information dynamics measures to explore and interface with such processes at the high and abstract level of expectation, randomness and predictability. The Melody Triangle is such a system.

A. *The Melody Triangle*

The Melody Triangle is an exploratory interface for the discovery of melodic content, where the input – positions within a triangle – directly map to information theoretic measures associated with the output. The measures are the entropy rate, redundancy and predictive information rate of the random process used to generate the sequence of notes. These are all related to the predictability of the the sequence and as such address the notions of expectation and surprise in the perception of music.*self-plagiarised*

Before the Melody Triangle can used, it has to be 'populated' with possible parameter values for the melody generators. These are then plotted in a 3d statistical space of redundancy, entropy rate and predictive information rate. In our case we generated thousands of transition matrixes, representing first-order Markov chains, by a random sampling method. In figure 6 we see a representation of how these matrixes are distributed in the 3d statistical space; each one of these points corresponds to a transition matrix.*self-plagiarised*

When we look at the distribution of transition matrixes plotted in this space, we see that it forms an arch shape that is fairly thin. It thus becomes a reasonable approximation to pretend that it is just a sheet in two dimensions; and so we stretch out this curved arc into a flat triangle. It is this triangular sheet that is our 'Melody Triangle' and forms the interface by which the system is controlled. *self-plagiarised*

When the Melody Triangle is used, regardless of whether it is as a screen based system, or as an interactive installation, it involves a mapping to this statistical space. When the user, through the interface, selects a position within the triangle, the corresponding transition matrix is returned. Figure 7 shows how the triangle maps to different measures of redundancy, entropy rate and predictive information rate.*self-plagiarised* Each corner corresponds to three different extremes of predictability and unpredictability, which could be loosely characterised as 'periodicity', 'noise' and 'repetition'. Melodies from the 'noise' corner have no discernible pattern; they have high entropy rate, low predictive information rate and low redundancy. These melodies are essentially totally random. A melody along the 'periodicity' to 'repetition' edge are all deterministic loops that get shorter as we approach the 'repetition' corner, until it becomes just one repeating note. It is the areas in between the extremes that provide the more 'interesting' melodies. That

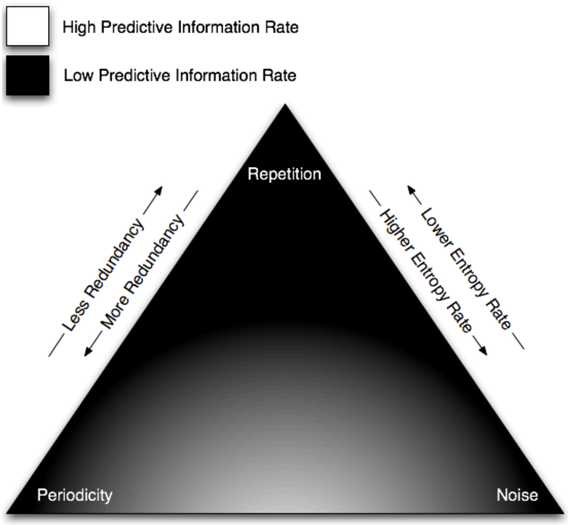


Fig. 7. The Melody Triangle

is, those that have some level of unpredictability, but are not completely ran- dom. Or, conversely, that are predictable, but not entirely so. This triangular space allows for an intuitive explorationof expectation and surprise in temporal sequences based on a simple model of how one might guess the next event given the previous one.*self-plagiarised*

Any number of interfaces could be developed for the Melody Triangle. We have developed two; a standard screen based interface where a user moves tokens with a mouse in and around a triangle on screen, and a multi-user interactive installation where a Kinect camera tracks individuals in a space and maps their positions in the space to the triangle. Each visitor would generate a melody, and could collaborate with their co-visitors to generate musical textures – a playful yet informative way to explore expectation and surprise in music.

As a screen based interface the Melody Triangle can serve as composition tool. A triangle is drawn on the screen, screen space thus mapped to the statistical space of the Melody Triangle. A number of round tokens, each representing a melody can be dragged in and around the triangle. When a token is dragged into the triangle, the system will start generating the sequence of notes with statistical properties that correspond to its position in the triangle.*self-plagiarised*

In this mode, the Melody Triangle can be used as a kind of composition assistant for the generation of interesting musical textures and melodies. However unlike other computer aided composition tools or programming environments, here the composer engages with music on the high and abstract level of expectation, randomness and predictability.*self-plagiarised*

Additionally the Melody Triangle serves as an effective tool for experimental investigations into musical preference and their relationship to the information dynamics models.

V. Musical Preference and Information Dynamics

We carried out a preliminary study that sought to identify any correlation between aesthetic preference and the information theoretical measures of the Melody Triangle. In this study participants were asked to use the screen based interface but it was simplified so that all they could do was move tokens around. To help discount visual biases, the axes of the triangle would be randomly rearranged for each participant.*self-plagiarised*

The study was divided in to two parts, the first investigated musical preference with respect to single melodies at different tempos. In the second part of the study, a background melody is playing and the participants are asked to continue playing with the system under the implicit assumption that they will try to find a second melody that works well with the background melody. For each participant this was done four times, each with a different background melody from four different areas of the Melody Triangle. For all parts of the study the participants were asked to signal, by pressing the space bar, whenever they liked what they were hearing.*self-plagiarised*

todo - results

VI. Information Dynamics as Evaluative Feedback Mechanism

todo - code the info dyn evaluator :)

It is possible to use information dynamics measures to develop a kind of ‘critic’ that would evaluate a stream of symbols. For instance we could develop a system to notify us if a stream of symbols is too boring, either because they are too repetitive or too chaotic. This could be used to evaluate both pre-composed streams of symbols, or could even be used to provide real-time feedback in an improvisatory setup.

comparable system Gordon Pask’s Musicolor (1953) applied a similar notion of boredom in its design. The Musicolour would react to audio input through a microphone by flashing coloured lights. Rather than a direct mapping of sound to light, Pask designed the device to be a partner to a performing musician. It would adapt its lighting pattern based on the rhythms and frequencies it would hear, quickly ‘learning’ to flash in time with the music. However Pask endowed the device with the ability to ‘be bored’; if the rhythmic and frequency content of the input remained the same for too long it would listen for other rhythms and frequencies, only lighting when it heard these. As the Musicolour would ‘get bored’, the musician would have to change and vary their playing, eliciting new and unexpected outputs in trying to keep the Musicolour interested.

In a similar vein, our *Information Dynamics Critic*(name?) allows for an evaluative measure of an input stream, however containing a more sophisticated notion of boredom that …

VII. Conclusion

References

[1] Leonard B. Meyer. *Music, the arts and ideas: Patterns and Predictions in Twentieth-century culture*. University of Chicago Press, 1967.

[2] Eugene Narmour. *Beyond Schenkerism*. University of Chicago Press, 1977.

[3] Eduard Hanslick. *On the musically beautiful: A contribution towards the revision of the aesthetics of music*. Hackett, Indianapolis, IN, 1854/1986.

[4] J. R. Saffran, E. K. Johnson, R. N. Aslin, and E. L. Newport. Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1):27–52, 1999.

[5] T. Eerola, P. Toiviainen, and C. L. Krumhansl. Real-time prediction of melodies: Continous predictability judgments and dynamic models. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, and J. Renwick, editors, *Proceedings of the 7th International Conference on Music Perception and Cognition (ICMPC7)*, Syndney, Australia, 2002. Causal Productions.

[6] Darrell Conklin and Ian H. Witten. Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73, 1995.

[7] D. Ponsford, G. A. Wiggins, and C. S. Mellish. Statistical learning of harmonic movement. *Journal of New Music Research*, 28(2):150–177, 1999. Also available as Research Paper 874, from the Division of Informatics, University of Edinburgh.

[8] Marcus T. Pearce. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, Department of Computing, City University, London, 2005.

[9] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656, 1948.

[10] J. E. Youngblood. Style as information. *Journal of Music Theory*, 2:24–35, 1958.

[11] E. Coons and D. Kraehenbuehl. Information as a measure of structure in music. *Journal of Music Theory*, 2(2):127–161, 1958.

[12] Lejaren Hiller and Calvert Bean. Information theory analyses of four sonata expositions. *Journal of Music Theory*, 10(1):96–137, 1966.

[13] Abraham Moles. *Information Theory and Esthetic Perception*. University of Illinois Press, 1966.

[14] J. E. Cohen. Information theory and music. *Behavioral Science*, 7(2):137–163, 1962.

[15] D. E. Berlyne. *Aesthetics and Psychobiology*. Appleton Century Crofts, New York, 1971.

[16] R.W. Yeung. A new outlook on Shannon's information measures. *Information Theory, IEEE Transactions on*, 37(3):466–474, 1991.

[17] W. McGill. Multivariate information transmission. *Information Theory, IRE Professional Group on*, 4(4):93–111, 1954.

[18] JP Crutchfield and NH Packard. Symbolic dynamics of noisy chaos. *Physica D: Nonlinear Phenomena*, 7(1-3):201–223, 1983.

[19] Shlomo Dubnov. Spectral anticipations. *Computer Music Journal*, 30(2):63–83, 2006.

[20] Ionas Erb and Nihat Ay. Multi-information in the thermodynamic limit. *Journal of Statistical Physics*, 115:949–976, 2004.

[21] Samer A. Abdallah and Mark D. Plumbley. Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, 21(2):89–117, 2009.

[22] Samer A. Abdallah and Mark D. Plumbley. Predictive information, multi-information and binding information. Technical Report C4DM-TR-10-10, Queen Mary University of London, 2010.

[23] S. Verdú and T Weissman. Erasure entropy. In *IEEE International Symposium on Information Theory (ISIT 2006)*, pages 98–102, 2006.

[24] Ryan G. James, Christopher J. Ellison, and James P. Crutchfield. Anatomy of a bit: Information in a time series observation. *Chaos*, 21(3):037109, 2011.