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

***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.

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 $\mathcal{X}$ 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 $p : \mathcal{X} \to [0, 1]$, this is

$$H(X) = \sum_{x \in \mathcal{X}} -p(x) \log p(x), \tag{1}$$

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 expectation of the surprisingness $\mathrm{E}\, \ell(X)$.

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:

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

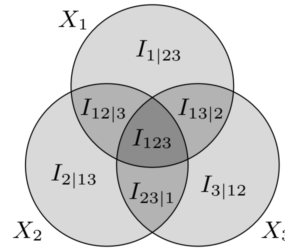
When there are multiple variables X_1, X_2 etc. which the observer believes to be dependent, then the observation of one may change its beliefs and hence yield information about the others. The joint and conditional entropies as described in any textbook on information theory (e.g. [16]) then quantify the observer's expected uncertainty about groups of variables given the values of others. In particular, the *mutual information* $I(X_1; X_2)$ is both the expected information in an observation of X_2 about X_1 and the expected reduction in uncertainty about X_1 after observing X_2:

$$I(X_1; X_2) = H(X_1) - H(X_1|X_2), \tag{3}$$

where $H(X_1|X_2) = H(X_1, X_2) - H(X_2)$ is the conditional entropy of X_2 given X_1. A little algebra shows that $I(X_1; X_2) = I(X_2; X_1)$ and so the mutual information is symmetric in its arguments. A conditional form of the mutual information can be formulated analogously:

$$I(X_1; X_2|X_3) = H(X_1|X_3) - H(X_1|X_2, X_3). \tag{4}$$

These relationships between the various entropies and mutual informations are conveniently visualised in Venn diagram-like *information diagrams* or I-diagrams [17] such as the one in fig. 1.



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

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 [18]. Some other information measures are indicated in the legend.

B. Surprise and information in sequences

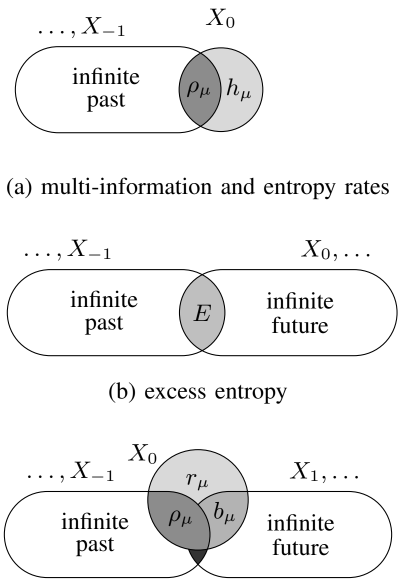
Suppose that $(\ldots, X_{-1}, X_0, X_1, \ldots)$ is a sequence of random variables, infinite in both directions, and that μ is the associated probability measure over all realisations of the sequence—in the following, μ will simply serve as a label for the process. We can indentify a number of information-theoretic measures meaningful in the context of a sequential observation of the sequence, during which, at any time t, the sequence of variables can be divided into a 'present' X_t, a 'past' $\overleftarrow{X}_t \equiv (\ldots, X_{t-2}, X_{t-1})$, and a 'future' $\overrightarrow{X}_t \equiv (X_{t+1}, X_{t+2}, \ldots)$. We will write the actually observed value of X_t as x_t, and the sequence of observations up to but not including x_t as $\overleftarrow{x}_t$.

The in-context surprisingness of the observation $X_t = x_t$ depends on both x_t and the context $\overleftarrow{x}_t$:

$$\ell_t = -\log p(x_t|\overleftarrow{x}_t). \tag{5}$$

However, before X_t is observed to be x_t, the observer can compute its *expected* surprisingness as a measure of its uncertainty about the very next event; this may be written as an entropy $H(X_t|\overleftarrow{X}_t = \overleftarrow{x}_t)$, but note that this is conditional on the *event* $\overleftarrow{X}_t = \overleftarrow{x}_t$, not *variables* $\overleftarrow{X}_t$ as in the conventional conditional entropy.

The surprisingness ℓ_t and expected surprisingness $H(X_t|\overleftarrow{X}_t = \overleftarrow{x}_t)$ can be understood as *subjective* information dynamic measures, since they are based on the observer's probability model in the context of the actually observed sequence $\overleftarrow{x}_t$—they characterise what it is like to 'be in the observer's shoes'. If we view the observer as a purely passive or reactive agent, this would probably be sufficient, but for active agents such as humans or animals, it is often necessary to *aniticipate* future events in order, for example, to plan the most effective course of action. It makes sense for such observers to be concerned about the predictive probability distribution over future events, $p(\overrightarrow{x}_t|\overleftarrow{x}_t)$. When an observation $X_t = x_t$ is made in this context, the *instantaneous predictive information* (IPI) $\mathcal{I}_t$ at time t is the information in the event $X_t = x_t$ about the entire future of the sequence $\overrightarrow{X}_t$, *given* the observed past $\overleftarrow{X}_t = \overleftarrow{x}_t$. Referring to the definition of information (2), this is the KL divergence between prior



(a) multi-information and entropy rates

(b) excess entropy

(c) 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 (b), the circle represents the 'present'. Its total area is $H(X_0) = \rho_\mu + r_\mu + b_\mu$, 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 $h_\mu = r_\mu + b_\mu$. The small dark region below X_0 in (c) is $\sigma_\mu = E - \rho_\mu$.

and posterior distributions over possible futures, which written out in full, is

$$\mathcal{I}_t = \sum_{\overrightarrow{x}_t \in \mathcal{X}^*} p(\overrightarrow{x}_t | x_t, \overleftarrow{x}_t) \log \frac{p(\overrightarrow{x}_t | x_t, \overleftarrow{x}_t)}{p(\overrightarrow{x}_t | \overleftarrow{x}_t)}, \tag{6}$$

where the sum is to be taken over the set of infinite sequences $\mathcal{X}^*$. As with the surprisingness, the observer can compute its *expected* IPI at time t, which reduces to a mutual information $I(X_t; \overrightarrow{X}_t | \overleftarrow{X}_t = \overleftarrow{x}_t)$ conditioned on the observed past. This could be used, for example, as an estimate of attentional resources which should be directed at this stream of data, which may be in competition with other sensory streams.

C. Information measures for stationary random processes

If we step back, out of the observer's shoes as it were, and consider the random process $(\ldots, X_{-1}, X_0, X_1, \ldots)$ as a statistical ensemble of possible realisations, and furthermore assume that it is stationary, then it becomes possible to define a number of information-theoretic measures, closely related to those described above, but which characterise the process as a whole, rather than on a moment-by-moment basis. Some of these, such as the entropy rate, are well-known, but others are only recently being investigated. (In the following, the assumption of stationarity means that the measures defined below are independent of t.)

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

$$h_\mu = H(X_t | \overleftarrow{X}_t). \tag{7}$$

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':

$$\rho_\mu = I(\overleftarrow{X}_t; X_t) = H(X_t) - h_\mu. \tag{8}$$

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* [20] is the mutual information between the entire 'past' and the entire 'future':

$$E = I(\overleftarrow{X}_t; X_t, \overrightarrow{X}_t). \tag{9}$$

Both the excess entropy and the multi-information rate can be thought of as measures of *redundancy*, quantifying the extent to which the same information is to be found in all parts of the sequence.

The *predictive information rate* (or PIR) [21] is the average information in one observation about the infinite future given the infinite past, and is defined as a conditional mutual information:

$$b_\mu = I(X_t; \overrightarrow{X}_t | \overleftarrow{X}_t) = H(\overrightarrow{X}_t | \overleftarrow{X}_t) - H(\overrightarrow{X}_t | X_t, \overleftarrow{X}_t). \tag{10}$$

Equation (10) 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

$$b_\mu = H(X_t | \overleftarrow{X}_t) - H(X_t | \overleftarrow{X}_t, \overrightarrow{X}_t) = h_\mu - r_\mu, \tag{11}$$

where $r_\mu = H(X_t | \overrightarrow{X}_t, \overleftarrow{X}_t)$, is the *residual* [22], or *erasure* [23] entropy rate. These relationships are illustrated in Fig. 2, along with several of the information measures we have discussed so far.

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).

D. First and higher order Markov chains

First order Markov chains are the simplest non-trivial models to which information dynamics methods can be applied. In [21] we derived expressions for all the information measures described in § II-B for irreducible stationary Markov chains (i.e. that have a unique stationary distribution). The derivation is greatly simplified by the dependency structure of the Markov chain: for the purpose of the analysis, the 'past' and 'future' segments $\overleftarrow{X}_t$ and $\overrightarrow{X}_t$ can be collapsed to just the previous and next variables X_{t-1} and X_{t-1} respectively. We also showed that the predictive information rate can be expressed simply in terms of entropy rates: if we let a denote the $K \times K$ transition matrix of a Markov chain over an alphabet of $\{1, \ldots, K\}$, such that $a_{ij} = \Pr(X_t{=}i | X_{t-1} = j)$, and let $h : \mathbb{R}^{K \times K} \to \mathbb{R}$ be the entropy rate function such that $h(a)$ is the entropy rate of a Markov chain with transition matrix a, then the predictive information rate $b(a)$ is

$$b(a) = h(a^2) - h(a), \tag{12}$$

where a^2, the transition matrix squared, is the transition matrix of the 'skip one' Markov chain obtained by jumping two steps at a time along the original chain.

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 size K^N consisting of all possible N-tuples of symbols from the base alphabet. An observation $\hat{x}_t$ in this new model encodes a block of N observations $(x_{t+1}, \ldots, x_{t+N})$ from the base model. The next observation $\hat{x}_{t+1}$ encodes the block of N obtained by shifting the previous block along by one step. The new Markov of chain is parameterised by a sparse $K^N \times K^N$ transition matrix $\hat{a}$. Adopting the label μ for the order N system, we obtain:

$$h_\mu = h(\hat{a}), \qquad b_\mu = h(\hat{a}^{N+1}) - N h(\hat{a}), \tag{13}$$

where $\hat{a}^{N+1}$ is the $(N{+}1)$th power of the first order transition matrix. Other information measures can also be computed for the high-order Markov chain, including the multi-information rate ρ_μ and the excess entropy E. These are identical for first order Markov chains, but for order N chains, E can be up to N times larger than ρ_μ.

[Something about what kinds of Markov chain maximise h_μ (uncorrelated 'white' sequences, no temporal structure), ρ_μ and E (periodic) and b_μ. We return this in § IV.]

III. Information Dynamics in Analysis

A. Musicological Analysis

In [21], methods based on the theory described above were used to analysis two pieces of music in the minimalist style by Philip Glass: *Two Pages* (1969) and *Gradus* (1968). The analysis was done using a first-order Markov chain model, with the enhancement that the transition matrix of the model was allowed to evolve dynamically as the notes were processed, and was tracked (in a Bayesian way) as a *distribution* over possible transition matrices, rather than a point estimate. The results are summarised in fig. 3: the upper four plots show the dynamically evolving subjective information measures as described in § II-B computed using a point estimate of the current transition matrix, but the fifth plot (the 'model information rate') measures the information in each observation about the transition matrix. In [25], we showed that this 'model information rate' is actually a component of the true IPI in a

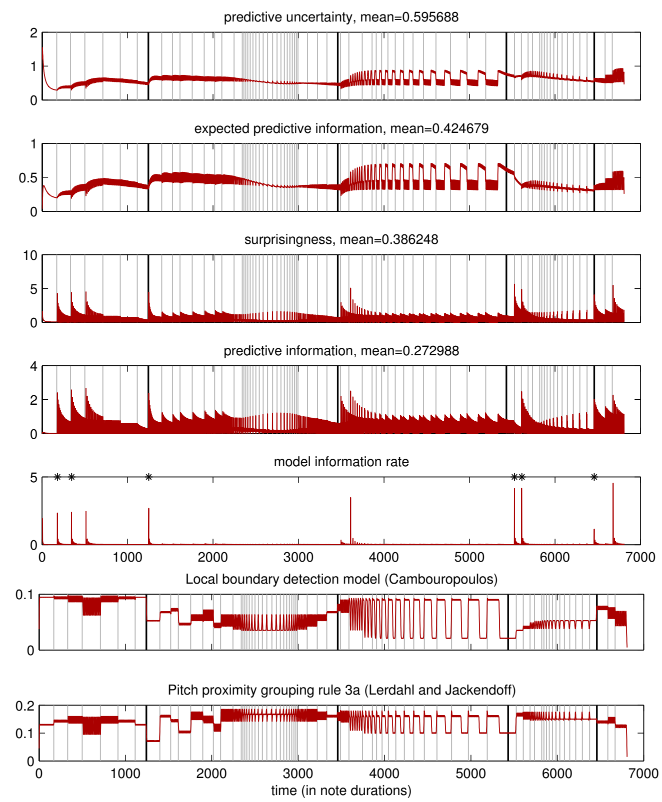


Fig. 3. Analysis of *Two Pages*. The thick vertical lines are the part boundaries as indicated in the score by the composer. The thin grey lines indicate changes in the melodic 'figures' of which the piece is constructed. In the 'model information rate' panel, the black asterisks mark the six most surprising 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.

time-varying Markov chain, which was neglected when we computed the IPI from point estimates of the transition matrix as if the transition probabilities were constant.

The peaks of the surprisingness and both components of the predictive information show good correspondence with structure of the piece both as marked in the score and as analysed by musicologist Keith Potter, who was asked to mark the six 'most surprising moments' of the piece (shown as asterisks in the fifth plot)[1].

In contrast, the analyses shown in the lower two plots of fig. 3, obtained using two rule-based music segmentation algorithms, while clearly *reflecting* the structure of the piece, do not *segment* the piece, with no tendency to peaking of the boundary strength function at the boundaries in the piece.

B. Content analysis/Sound Categorisation

. Using analogous definitions of differential entropy, the methods outlined in the previous section are equally appli-

[1]Note that the boundary marked in the score at around note 5,400 is known to be anomalous; on the basis of a listening analysis, some musicologists [ref] have placed the boundary a few bars later, in agreement with our analysis.

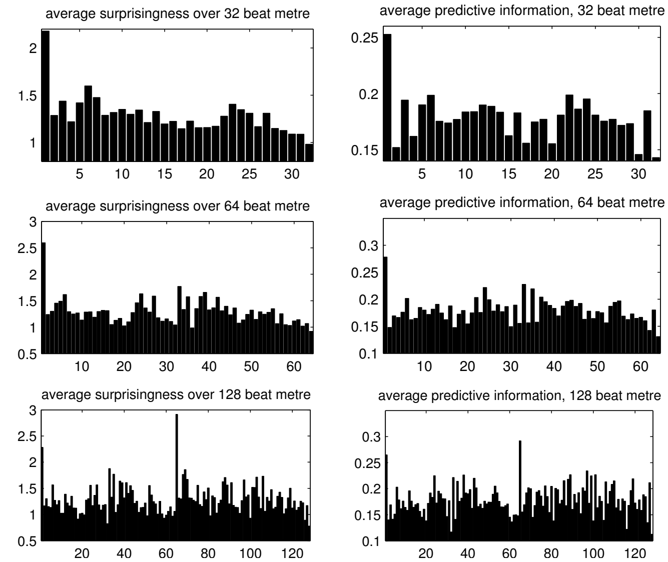


Fig. 4. Metrical analysis by computing average surprisingness and informative of notes at different periodicities (i.e. hypothetical bar lengths) and phases (i.e. positions within a bar).

cable to continuous random variables. In the case of music, where expressive properties such as dynamics, tempo, timing and timbre are readily quantified on a continuous scale, the information dynamic framework thus may also be considered.

In [19], Dubnov considers the class of stationary Gaussian processes. For such processes, the entropy rate may be obtained analytically from the power spectral density of the signal, allowing the multi-information rate to be subsequently obtained. Local stationarity is assumed, which may be achieved by windowing or change point detection [26]. mention non-gaussian processes extension Similarly, the predictive information rate may be computed using a Gaussian linear formulation CITE. In this view, the PIR is a function of the correlation between random innovations supplied to the stochastic process. Dean (2009)

- Continuous domain information
- Audio based music expectation modelling
- Proposed model for Gaussian processes

C. Beat Tracking

A probabilistic method for drum tracking was presented by Robertson [**?**]. The algorithm is used to synchronise a music sequencer to a live drummer. The expected beat time of the sequencer is represented by a click track, and the algorithm takes as input event times for discrete kick and snare drum events relative to this click track. These are obtained using dedicated microphones for each drum and using a percussive onset detector (Puckette 1998). The drum tracker continually updates distributions for tempo and phase on receiving a new event time. We can thus quantify the information contributed of an event by measuring the difference between the system's prior distribution and the posterior distribution using the Kullback-Leiber divergence.

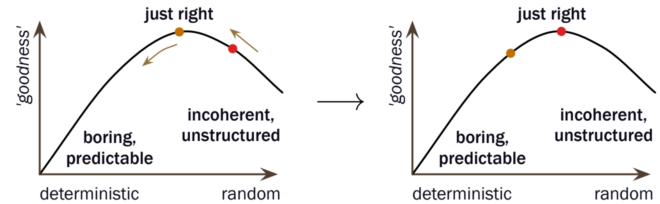


Fig. 5. The Wundt curve relating randomness/complexity with perceived value. Repeated exposure sometimes results in a move to the left along the curve [15].

Here, we have calculated the KL divergence and entropy for kick and snare events in sixteen files. The analysis of information rates can be considered *subjective*, in that it measures how the drum tracker's probability distributions change, and these are contingent upon the model used as well as external properties in the signal. We expect, however, that following periods of increased uncertainty, such as fills or expressive timing, the information contained in an individual event increases. We also examine whether the information is dependent upon metrical position.

IV. Information dynamics as compositional aid

In addition to applying information dynamics to analysis, it is also possible to apply it to the generation of content, such as to the composition of musical materials. The outputs of algorithmic or stochastic processes can be filtered to match a set of criteria defined in terms of the information dynamics model, this criteria thus becoming a means of interfacing with the generative process. For instance a stochastic music generating process could be controlled by modifying constraints on its output in terms of predictive information rate or entropy rate.

The use of stochastic processes for the composition of musical material has been widespread for decades—for instance Iannis Xenakis applied probabilistic mathematical models to the creation of musical materials [27]. Information dynamics can serve as a novel framework for the exploration of the possibilities of such processes at the high and abstract level of expectation, randomness and predictability.

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 of the output. The measures—entropy rate, redundancy and predictive information rate—form a criteria with which to filter the output of the stochastic processes used to generate sequences of notes. These measures address notions of expectation and surprise in music, and as such the Melody Triangle is a means of interfacing with a generative process in terms of the predictability of its output.

The triangle is 'populated' with possible parameter values for melody generators. These are plotted in a 3D information space of ρ_μ (redundancy), h_μ (entropy rate) and b_μ (predictive

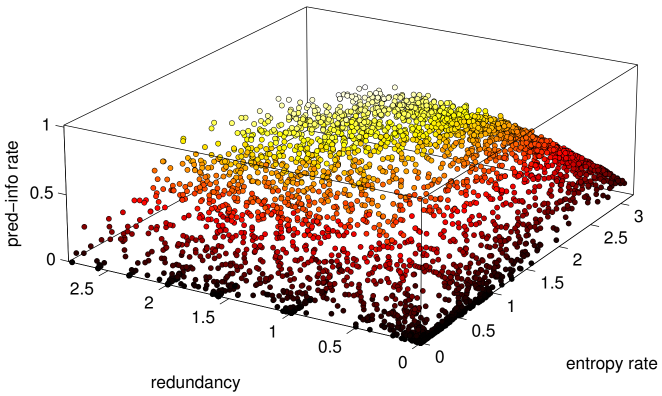


Fig. 6. The population of transition matrices distributed along three axes of redundancy, entropy rate and predictive information rate (all measured in bits). The concentrations of points along the redundancy axis correspond to Markov chains which are roughly periodic with periods of 2 (redundancy 1 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.

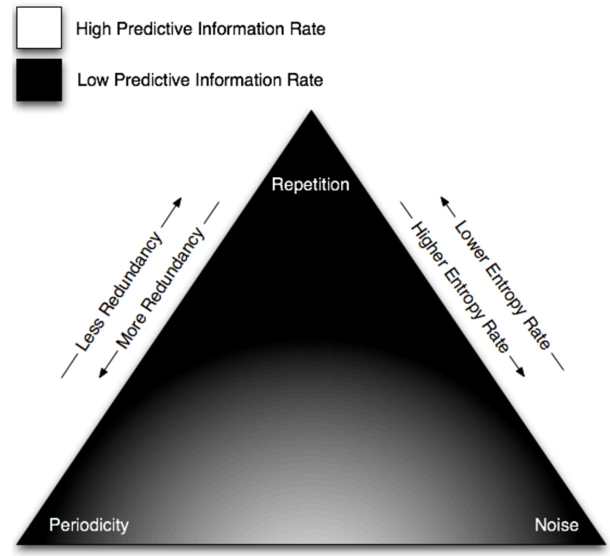


Fig. 7. The Melody Triangle

information rate), as defined in § II-C. In our case we generated thousands of transition matrices, representing first-order Markov chains, by a random sampling method. In figure 6 we see a representation of how these matrices are distributed in the 3d statistical space; each one of these points corresponds to a transition matrix.

The distribution of transition matrices plotted in this space 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. Using this interface thus involves a mapping to statistical space; a user selects a position within the triangle, and a corresponding transition matrix is returned. Figure 7 shows how the triangle maps to different measures of redundancy, entropy rate and predictive information rate.

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. These melodies have some level of unpredictability, but are not completely random. Or, conversely, are predictable, but not entirely so.

The Melody Triangle exists in two incarnations; a standard screen based interface where a user moves tokens 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 physical space to the triangle. In the latter visitors entering the installation generates 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. Additionally different gestures could be detected to change the tempo, register, instrumentation and periodicity of the output melody.

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 symbols with statistical properties that correspond to the position of the token. These symbols are then mapped to notes of a scale. Keyboard input allow for control over additionally parameters.

The Melody Triangle is can assist a composer in the creation not only of melodies, but, by placing multiple tokens in the triangle, can generate intricate musical textures. 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.

B. Information Dynamics as Evaluative Feedback Mechanism

Information measures on a stream of symbols can form a feedback mechanism; a rudamentary 'critic' of sorts. For instance symbol by symbol measure of predictive information rate, entropy rate and redundancy could tell us if a stream of symbols is currently 'boring', either because it is too repetitive, or because it is too chaotic. Such feedback would be oblivious to more long term and large scale structures, but it nonetheless could be provide a composer valuable insight on the short term properties of a work. This could not only be used for the evaluation of pre-composed streams of symbols, but could also provide real-time feedback in an improvisatory setup.

V. Musical Preference and Information Dynamics

We are carrying out a study to investigate the relationship between musical preference and the information dynamics models, the experimental interface a simplified version of the screen-based Melody Triangle. Participants are asked to use this music pattern generator under various experimental conditions in a composition task. The data collected includes usage statistics of the system: where in the triangle they place the tokens, how long they leave them there and the state of the system when users, by pressing a key, indicate that they like what they are hearing. As such the experiments will help us identify any correlation between the information theoretic properties of a stream and its perceived aesthetic worth.

VI. Conclusion

VII. acknowledgments

This work is supported by EPSRC Doctoral Training Centre EP/G03723X/1 (HE), GR/S82213/01 and EP/E045235/1(SA), an EPSRC Leadership Fellowship, EP/G007144/1 (MDP) and EPSRC IDyOM2 EP/H013059/1.

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] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, New York, 1991.

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

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

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

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

[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.

[25] Samer A. Abdallah and Mark D. Plumbley. Predictive information and Bayesian surprise in exchangeable random processes. Technical Report C4DM-TR10-09, Queen Mary University of London, 2010.

[26] Shlomo Dubnov. Unified view of prediction and repetition structure in audio signals with application to interest point detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):327–337, 2008.

[27] Iannis Xenakis. *Formalized music : thought and mathematics in composition*. Pendragon Press, Stuyvesant, NY, 1992.