Interdisciplinary Studies in Musicology 10, 2011

JUSTYNA HUMIĘCKA-JAKUBOWSKA (Poznań)

On certain 'tools' for research into the perception and creation of music and the complex ways in which they affect one another

ABSTRACT: Perception is a constructive mental process, which cannot be considered impersonally. Similarly, music cannot be cognised solely on the basis of its score, since its coming into being is strictly connected to the activation of human memory and sound imagination. The patterns that emerge from the sounds of heard music enable the listener to draw conclusions regarding the structures those sounds embody. However, such conclusions are accompanied by a degree of uncertainty, which concerns not just the perceived moment of the heard music, but also the way in which it is represented in the listener's memory. Perception is an inferential, multi-layered, uncertain process, in which particular patterns seem more likely than others. Mental representations of those probabilities lie behind such essential musical phenomena as surprise, tension, expectation and pitch identification, which are fixed elements of the perception of music.

The aim of the present article is to describe the essence of three selected types of music modelling, based on spectral anticipation (Shlomo Dubnov), based on memory (Rens Bod), and exploiting the dynamic character of music to obtain information (Samer Abdallah and Mark Plumbley). All these models take account of the element of uncertainty that accompanies the perception of music; hence they make use the foundations of information theory and statistical analysis as measurement 'tools'. The use of these tools makes it possible to obtain numerical rates, which inform us of the degree of predictability of the musical structures being analysed. One crucial advantage of these methods is the possibility of evaluating them in respect to the use of real musical structures, deriving from actual music, and not abstract structures formed for the purposes of research. We obtain cognitive insight into the analysed music by employing methods of a mathematical provenance, and so we have the possibility of examining music whilst taking account of the role of the listener, but with the use of objectivised methods.

KEYWORDS: perception, musical structure, modelling, entropy, probability, Markov process, mutual information, parse, expectation, surprise

Introduction

Insight into the mechanisms responsible for the perception of music not only provides us with knowledge about perception itself, as one of the elements – alongside memory and thought – in the process of cognising music, but is also a source of knowledge about music itself. Perception is a constructive mental process, which cannot be considered impersonally, be it only for the reason that the perception of music is dependent to a substantial degree on the listener's level of perceptual 'training'. Similarly, music cannot be cognised solely on the basis of its score, since its coming into being is strictly connected to the activation of human memory and sound imagination.

Although the notation of music in a given musical style is unchanging, the perception of that music alters over the course of time. This is because, in our everyday lives, we are subjected to a huge amount of mental stimuli, and nature has equipped us for gaining control over such information. A listener's musical behaviour proceeds according to a specific standard. Listening to an unfamiliar piece of music for the first time, a listener attempts to associate/match its material with his previous perceptual experiences, the effects of which he has assembled in his long-term memory in the form of cognitive schemata. If such association/matching fails, then the listener seeks new patterns for the information that is reaching him. Since the capacity of short-term memory is limited to a more or less constant quantity of perceptual units, and that limitation is independent of the quantity of information contained in each perceptual unit,[1] the quantity of information that can be stored in that memory is determined by the way in which the listener forms a perceptual pattern from that information. If, while consciously listening to music, a listener recognises a pattern, then the quantity of perceptual units requiring further consideration decreases. As a listener acquires auditory experience, information that originally occupied several perceptual units is patterned on a lesser quantity of units, and space in the short-term memory is freed for additional information. In this way, over time, the perception of music experienced many times over changes.

The patterns that emerge from the sounds of heard music enable the listener to draw conclusions regarding the structures those sounds embody. However, such conclusions are accompanied by a degree of uncertainty, which concerns not just the perceived moment of the heard music, but also the way in which it is represented in the listener's memory. Perception is an inferential, multi-layered, uncertain process, in which particular patterns

[1] George A. Miller, 'The Magic Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', in *The Psychology of Communication: Seven Essays*, ed. George A. Miller (Baltimore, 1969), 14–44, at 36.

seem more likely than others. We know today that mental representations of those probabilities lie behind such essential musical phenomena as surprise, tension, expectation and pitch identification, which are fixed elements of the perception of music.[2] The variability of the perception that is triggered by human musical experience affects the size of the probabilities ascribed to the sound patterns formed while listening to music, which shape the surface of that music, and to the structures on which they are based. Human knowledge of probabilities also derives, to a large extent, from the regularity of the aural environment. The awareness of these relationships influences the creation of music, with the effect that the composer, employing a particular strategy, determines the perceptual processes. The perceiving of music is linked to the communicating of particular structures from the mind of the composer to the mind of the listener by means of a particular representation of the surface of music. The success of such communication depends on a common knowledge of the style of a given piece of music, and also on the character of that style. Every musical style is characterised by means of probabilistic limitations imposed on the structures and on the mutual relations between them and the representations of the musical surface that are formed on their basis. However, in order for the communication between the composer and the listener to proceed successfully, there must also occur between them an understanding in respect to those limitations. Otherwise, the communication may be hampered or even precluded, as a result of which even a listener with complete knowledge of the style of the given piece of music may not be able to divine the structures intended by the composer from the perceived surface of the music.

The purpose of interdisciplinary research into music is to discover the mental processes and representations involved in such musical behaviour as listening, performing and composing music. Thanks to elaborated models enabling us to analyse the musical structures that underpin the shaping of perceived representations of musical surface, we can gain insight into many aspects of the perception of music, and thereby point to the solutions employed in the creative process through which the listener can discover those same structures intended by the composer.

Theoretical foundations

The starting point for applying the foundations of information theory in the modelling of musical perception is to treat a musical structure as a source of information about its surface, formed while listening to music. This information is transmitted to the listener – the information sink – at a

[2] David Temperley, *Music and Probability* (Cambridge, 2007), 3.

specific time, which is why the process of communication is dynamic and depends both on the music itself and on the listener, who – in perceiving the music – makes predications of a sort regarding the information source and forms particular expectations on the basis of previous musical experiences. So if the musical structure, as an information source, is responsible for creating particular data about the heard music, this also allows us to describe many different sequences by means of a single statistical model with a specific probability distribution. A research approach that takes account of the foundations of information theory makes it possible to evaluate which data are more likely to appear, and also, further down the line, to state which structures of music are more typical or appear extremely rarely.

One of the fundamental notions employed in the modelling of perception is *entropy*, which defines the characteristic size of uncertainty proper to a given source and calculated as the quotient of the logarithm from the number of typical sequences and the logarithm from all possible sequences of the same length.[3] Thus entropy is a logarithm from the relative size of a typical set. This means that the uncertainty is greater with a larger typical set: a high entropy signifies a low certainty, and so a high uncertainty. At this point, it is worth also emphasising the difference between determinism and predictability. Randomness – triggering uncertainty with a low level of predictability – usually applies to a set that contains an element of variability or surprise, whilst structure is understood as something more predictable, based on rules, or even deterministic. Structures, which possess deterministic dynamics, can have different degrees of predictability, depending on the precision of the measurement or accurate knowledge about their past.

One important aspect of information theory is the information channel, or the mutual link between the information source and the information sink. The channel is characterised by uncertainty as to what has been transmitted. In mathematical terms, this aspect is described by means of so-called mutual information. This can be defined as the difference between the entropy of a source and the conditional entropy between the source and the sink. If we treat the source information as a discrete random variable x, with a probability of $P(x)$, then its entropy is expressed by the equation $H(x) = \Sigma P(x)\log P(x)$. When the information reaching the sink is the discrete random variable y, with a probability of $P(y)$, then the entropy of this information is $H(y) = \Sigma P(y)\log P(y)$. The joint entropy for the two discrete random variables x and y is expressed by the equation $H(x, y) = -\Sigma\Sigma P(x, y)\log P(x, y)$. Finally, the conditional entropy, indicating the uncertainty as to the random variable y, on condition that we know random variable x, equates to $H(y|x) = \Sigma P(x)H(y|x)$ (the conditional entropy for random variable x may be indicated in a similar

[3] Henryk Górecki, *Teoria informacji* [Information theory] (Łódź, 2006).

way, on condition that we know random variable *y*). In this context, the function I(x, y), representing the value of the mutual information, shows how much information one variable provides about the other. One can also point to two extreme situations: the two variables are independent of one another, in which case the mutual information I(x, y) is zero; the variables are equivalent to one another, in which case the mutual information I(x, y) = H(x) = H(y) = H(x, y), since the conditional entropy is zero, because knowledge of one variable wholly describes the other variable, leaving no conditional uncertainty.[4] Ultimately, in the context of the terms and equations given above, we may represent the mutual information in the form of the following expressions: $I(x, y) = H(x) - H(x|y) = H(y) - H(y|x) = H(x) + H(y) - H(x, y) = \Sigma P(x, y)\log P(x, y)/P(x)P(y)$.

The models:
1. The model of spectral anticipation

The model of spectral anticipation[5] makes use of mutual information for the theoretical characterisation of the size of the information transmitted through the communication channel, which is a time channel. This means that the entry data for the channel is the history of the signal up to the current point in time, and the exit data is its next (present) sample. In addition, the receiver must employ certain algorithms for predicting the current sample on the basis of samples from the past. Accordingly, the information at the sink *y* consists of the history of the signal $x_1, x_2,..., x_{n-1}$ available to the receiver prior to its hearing x_n. As Dubnov indicates, the process of transmission is interpreted here in terms of the execution of prediction/anticipation in time. The time channel brings to the next sample the element of 'surprise'; the size of the mutual information depends on that surprise and on the listener's ability to predict it. The information transmission via time channel outlined above is registered by means of the information rate, IR, often defined as the scalar-IR. This is often interpreted as the size of the information which the signal takes into its future. It may be defined as the relative reduction of the uncertainty of the present, in a situation of consideration of the past, which equates to the size of the mutual information transmitted between the past $x_{past}=\{x_1, x_2,..., x_{n-1}\}$ and the present x_n. In the case of information for multiple variables (defined as multi-information), IR corresponds to the difference between the multi-information contained in

[4] Thomas M. Cover and Thomas A. Joy, *Elements of Information Theory* (New York, 1991); Monson H. Hayes, *Statistical Signal Processing and Modeling* (New York, 1996).

[5] Shlomo Dubnov, 'Spectral Anticipations', *Computer Music Journal* 30/2 (2006), 63–83.

the two sets of variables $x_1, x_2,..., x_n$ and $x_1, x_2,..., x_{n-1}$ (that is, the size of the extra information that is added when one more sample is observed during the transmission process): $\rho(x_1, x_2,..., x_n) = H(x_n) - H(x_n| x_{past}) = I(x_n, x_{past}) = I(x_1, x_2,..., x_n) - I(x_1, x_2,..., x_{n-1})$.[6]

Dubnov proposes the expansion of IR into a multi-dimensional process. In this case, he considers a new type of IR, which may apply to sequences with multiple variables described as vectors in higher-dimensional space. Marking the sequence of vectors $X_1, X_2,..., X_L$ and generalising the definition of IR (here called the vector-IR), we obtain the expression $\rho(X_1, X_2,..., X_L) = I(X_1, X_2,..., X_L) - \{I(X_1, X_2,..., X_{L-1}) + I(X_L)\}$, where the new definition for the multi-dimensional information rate states that this is the difference in the information following L successive vectors minus the sum of the information in the first $L-1$ vectors and the multi-information between elements within the last vector X_L. It also assumes the existence of a certain transformation T, such that $S = TX$ and the elements $S_1, S_2,..., S_L$ after transformation are statistically independent. Making use of the links between the entropies of the linear transformation of random vectors, IR may be calculated as the sum of the IRs of the individual elements, $s_i(n)$, $i = 1...n$, that is, $\rho_L(X_1, X_2,..., X_L) = \Sigma\ \rho(s_i(i),..., s_i(L))$. The crucial significance of the generalisation of the vector-IR for research into the perception of music is the fact that it allows us to identify the structural elements of the signal in the form of elements with a high scalar-IR.

In the Vector-IR Anticipation algorithm, Dubnov draws on representations of the Audio Basis. Such representations are obtained by transforming the time signal into a spectral domain by means of the Short-Time Fourier Transform (STFT). This is attained by employing Fourier transformation for blocks of audio samples, with the use of a so-called sliding-window, or 'windowing', which extracts short segments of the signal, known as 'frames', from the audio stream. Each frame can be mathematically considered as a vector in higher-dimensional space. In this context, the first stage in describing the modelling is to introduce a suitable geometrical representation of the music being studied, in the form of frames of audio samples or certain features obtained from those frames, through the use of various methods, such as the STFT or Filter Banks. The next stage in the research procedure is the decomposition of the base, combined with the reduction of the data through its mapping onto a lower-dimensional subspace. Ultimately, a separate IR assessment of the individual elements is carried out, in accordance with the principles of IR evaluation for a multi-dimensional/ vectorial process.

The use of this algorithm to analyse music makes it necessary to create an anticipation profile. This is because the properties of music, as a signal evolving over time, cannot be summarised as a single 'anticipation number'. The use of

[6] Dubnov, 'Spectral', 66.

this method of analysis is extended to a non-stationary case through the use of IR in a time-varying fashion.[7] In this phase in the research, the music is presented by means of a sequence of spectral envelopes, shown as spectral or cepstral coefficients. These vectors are then grouped into macro-frames and subjected to separate IR analyses, through which we obtain a single value for each macro-frame. Dubnov employs the term 'anticipation profile' to describe the graph of the evolution of the IR over a period of time corresponding to the duration of the music, and so an IR time graph. It should be added that cepstral coefficients are a representation of the spectral composition of the sound signal.[8] The cepstrum has come to be defined as the reverse of the Fourier transform from the logarithm of the absolute value of the Fourier transform of the signal, which is shown by the expression $C = F^{-1}\{\log(|F\{x(n)\}|)\}$. The advantage of this approach is the ability to capture various details of the signal's spectrum in a single representation. Hence, for example, the first cepstral coefficient corresponds to the energy of the signal, and lower coefficients capture the shape of the spectral envelope, whilst higher cepstral coefficients show spectral peaks corresponding to pitch or other long-term correlations of the signal. The choice of the cepstrum part makes it easy to control the type of spectral information, which we would like to regard as IR analysis.

By means of various parameters of analysis, anticipation profiles can be used to examine many aspects of music. The length of the frames required depends on the music: for example, approximately 3 seconds for solo and chamber music, and approximately 30 seconds for a complex orchestral texture. A crucial insight into the structural features of music is obtained by comparing the course of IR time graphs – observing the ridges and valleys and the falling and rising of the curves – with the surface of the music connected with the principal sections marked in the score and analysed from the perspective of music theory. As Dubnov stresses, the method of spectral anticipation, based on vector-IR analysis and anticipation profile, captures essential aspects of a musical structure.

2. Memory-based models

Many models that are designed to create the possibility of gaining insight into the mechanisms governing the perception of music employ abstract structures, built for the needs of specific research and applied to those models, as the analysed material. In a memory-based model, it is proposed that structures deriving from previously heard music be used. This procedure is based on

[7] Dubnov, 'Spectral', 75.

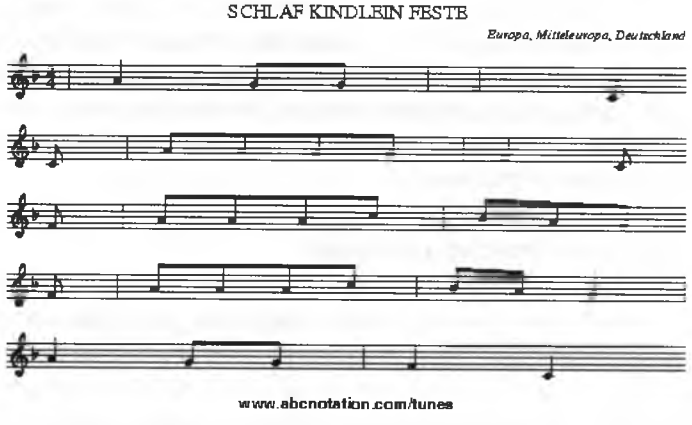
[8] Alan V. Oppenheim and Ronald W. Schafer, *Discrete Time Signal Processing* (New Jersey, 1989).

the results of psychological tests,[9] indicating the effect of the storing of the cognitive schemata of the heard music in the memory and a greater ease in activating them when the music that forms those schemata is more often available to the listener. The research material used in the memory-based models described by Rens Bod[10] is the Essen Folksong Collection, the musical parameters of which – such as pitch, duration, time signatures and explicit phrase markers – are specially encoded, which favours the use of this collection for research employing computational technique. However, the format of this collection contains no indications as to the hierarchic dimension of the structures, for instance phrase-internal structures, such as subphrases or motives, and phrase-external structures, like periods and segments. To illustrate this procedure, reproduced below is the notation of the song marked in the Essen Folksong Collection as K0029, 'Schlaf Kindlein Feste', and the encoded representation of five phrases from that song, put forward by Bod:[11]

S(P(3_221_-5) P(-533221_-5) P(13335432) P(13335432_) P(3_221_-5_))

S → PPPPP
P → 3_221_-5
P → -533221_-5
P → 13335432
P → 13335432_
P → 3_221_-5_

The essence of this research proposal is the modelling of the musical seg-



[9] Jenny R. Saffran, Michelle M. Loman and Rachel R. Robertson, 'Infant Memory for Musical Experiences', *Cognition* 77 (2000), B16-23.

[10] Rens Bod, 'Memory-Based Models of Melodic Analysis: Challenging the Gestalt Principles', *Journal of New Music Research* 31/1 (2002), 27-36.

[11] Bod, 'Memory-Based', 29 and 30.

mentation of the sound material that occurs while listening to music. The observed ambiguity of this mechanism concerns the divergence that frequently arises between several differently grouped sound structures that may be compatible with the sequence of notes recorded in the score, on one hand, and the structure perceived by the listener, apprehended in the form of just a single detailed specific structure, on the other. The modelling of the musical segmentation of a given work of music is based here on a linguistic approach to the musical work, with the consequent use of research methods/techniques that have come to be called 'grammar techniques'. One crucial feature of this type of modelling is the twofold use of the body of research material Part of the collection serves as a training set, and the rest as a test set. In the training set, all the phrases of a song are notated in the form of annotation strings, with the phrase ends marked by setting the representation of the phrase in brackets. In this way, the phrases are legible for a memory-based parser.[12] The training set is used here as the basis for machine learning, in which the process of learning a system is designed to attain certain results based on fragmentary knowledge. In this way, the process itself is improved; as a result, new notions may arise or an inductive inference be obtained. It should also be mentioned that the grammars used here are 'context-free grammars', characterised by the fact that all the rules for deriving expressions are given the form $A \rightarrow \Gamma$, where A is any nonterminal symbol and its signification does not depend on the context in which is occurs, and Γ is any sequence (even empty) of terminal and nonterminal symbols. In this notation, the symbol $\rightarrow$ denotes the act of derivation, that is, replacing a variable with the right side of the production for that variable, where productions are the rules linking variables to one another.

The memory-based models described by Bod, analysing the musical segmentation of the sound material, are based on the Treebank grammar technique, the Markov grammar technique and a technique combining the Markov grammar technique proposed by Collins[13] with Bod's Data-Oriented Parsing technique.[14]

The Treebank grammar technique essentially involves reading all the context-free rewrite rules from the structures of the training set, and then indicating for each rule the probability proportional to the frequency at which that rule occurs in the training set. Determined for each rewrite rule is a probability, the result of dividing the number of occurrences of a particular rule in the training set by the total number of the occurrences of the rules

[12] Thanks to the parser, computers are capable of transforming a human legible text into a data structure suitable for further processing.

[13] Michael Collins, 'Head-Driven Statistical Models for Natural Language Parsing', Ph.D. thesis (University of Pennsylvania, 1999).

[14] Rens Bod, *Beyond Grammar: An Experience-Based Theory of Language* (Cambridge, 1998).

which develop the same nonterminal as the rule in question. Thus, in the example given above, the value of the probability of the rule, for example, corresponding to the second phrase, that is, P → -533221_-5, is 1/5, since of all the five rules that develop this nonterminal P, this is the rule which occurs only once. As Bod emphasises, the Treebank grammar obtained from the training set in this way corresponds, in turn, to the Probabilistic Context-Free Grammar (PCFG),[15] in which it is assumed, above all, that context-free rules are statistically independent. Hence, taking into account the probabilities of the individual rules, one can calculate the probability of a parse tree as a product of the probabilities of each rule applied in that tree.

In this kind of research approach, there exists the problem of data sparseness, which is linked to the sporadic occurrence of some of the rules in the training set. This makes it more difficult to estimate their observed probabilities in their actual population of probabilities. In such instances, one may employ the Good-Turing method, involving the estimation of the expected population frequency of a particular type f * by adjusting frequency f to its observed sample frequency. The use of this method brings a new parameter to the research, namely n_f, the frequency of frequency f, denoting the number of types that occur f times in the observed sample. In such cases, the adjusted frequency f * is calculated from the following formula:

$$f^*=(f+1)\frac{n_{f+1}}{n_f}$$

This enables us to calculate the probabilities of context-free rules in the Treebank grammar from their adjusted, but not observed, frequencies.[16]

The Markov grammar technique seems a more solid model, since it enables probabilities to be calculated for every possible context-free rule, and not just for those rules which are seen in the training set. The essence of this technique is the decomposition of the rule and its probability by means of the Markov process, the starting point of which is the assumption of the presence of a sequence of events, with the probability of each event dependent solely on the result of the previous event. This means that a third-order Markov process will estimate probability *p* of rule P→12345 according to the following equation:

$$p(\mathrm{P}\rightarrow 12345)=p(1)\times p(2|1)\times p(3|1,2)\times p(4|1,2,3)\times p(5|2,3,4)\times p(\mathrm{END}|3,4,5),$$[17]

[15] Taylor L. Booth, T. 'Probabilistic Representation of Formal Languages', *Proceedings of the Tenth Annual IEEE Symposium on Switching and Automata Theory* (Canada, 1969) http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4569592, accessed 3 Novemeber 2011; Eugene Charniak, *Statistical Language Learning* (Cambridge, 1993).

[16] Bod, 'Memory-Based', 30.

[17] Ibid.

with the conditional probability $p(\text{END}|3,4,5)$ encoding the probability that the rule ends after notes 3, 4 and 5. This enables us to assess the probability of rule P→12345, even if it does not appear literally in the training set. The occurrence of data dispersal might mean that some Markov histories attain a particularly low, or even zero, value. Since the example equation given above is of a product form, even a single zero factor brings with it a zero probability of the whole rule. Adopted as a solution to the problem of the presence of data-sparseness was the linear interpolation technique, which interpolates the Markov history by including shorter histories in the analysis. Making use of Powell's algorithm,[18] enabling us to attain the optimal weights employed in linear interpolation, we can assess conditional probability in the Markov grammar technique. Ultimately, the probability of a parse tree of a particular musical work is calculated as a product of the probabilities of the rules that participate in the parse tree, like in the Treebank grammar.

It is also crucial to modelling the expectation of the correct segmentation of music to take account of the contexts. In the case of the Essen Folksong Collection, the form of contextual knowledge might be knowledge of the number of phrases appearing in an analysed song. A model combining the Markov grammar technique with Data-Oriented Parsing (DOP) makes it possible to take account of knowledge about the number of phrases when studying folk songs, by making use of one fundamental characteristic of the original DOP technique, namely the use of every musical fragment, of any length, that is visible in the training set as a productive unit. If the structure of the whole of a given song is designated by rule S, and that song comprises, for example, four phrases described by rule P (with that rule taking the form P→12345), the rule of the song takes the form S→PPPP, and then the DOP-Markov model, based on the history of three notes, determines the conditional probability of that rule on the basis of the following relationship:

$$\begin{aligned} p(\text{P}\rightarrow 12345|\text{S}\rightarrow\text{PPPP}) = {} & p(1|\text{S}\rightarrow\text{PPPP}) \\ & \times p(2|\text{S}\rightarrow\text{PPPP},1) \\ & \times p(3|\text{S}\rightarrow\text{PPPP},1,2) \\ & \times p(4|\text{S}\rightarrow\text{PPPP},1,2,3) \\ & \times p(5|\text{S}\rightarrow\text{PPPP},2,3,4) \\ & \times p(\text{END}|\text{S}\rightarrow\text{PPPP},3,4,5).[19] \end{aligned}$$

In this modelling technique, as well, the most probable parse of a folk song is calculated by maximising the product of the rule probabilities that generate the analysed song.

[18] Aleksander Ostanin, *Metody i algorytmy optymalizacji* [Methods and algorithms of optimalisation] (Białystok 2003).

[19] Bod, 'Memory-Based', 32.

As the results of research show, the use of probabilistic memory-based models allows us to predict musical segmentation more accurately, especially in the case of the analysis of phrases containing intervallic leaps. The grouping boundaries of such phrases occur before or after large pitch intervals, and they can even appear between identical notes which are preceded or followed by such large intervals. The modelling techniques outlined above make it possible to capture the whole continuum of an analysed musical work between leap-phrases and phrases not displaying such a feature to the pitch organisation of the analysed work.

3. The information-dynamic approach to modelling

This type of modelling takes account of one crucial characteristic of the perception of music. Music is not a static object of observation, but a dynamic one; consequently, as it unfolds, it modifies the listener's expectations and surprises, which arise from his previous experiences, current observations and future predictions in respect to the heard music. Much experimental research has reinforced the conviction that listeners are capable of internalising statistically collected knowledge of musical structures. This ability allows statistical models to be used to create an effective foundation for computational methods of analysing music. The chief thesis of such an approach is the assumption that perceived qualities and subjective states, such as uncertainty, surprise, complexity, tension and curiosity, are closely linked to certain quantities of information theory, and especially with entropy, relative entropy and mutual information.[20] On one hand, listening to music is connected with the behaviour of a dynamically evolving statistical model of music, which helps to form predictions as to the further course of a heard piece of music; on the other, it makes it possible to revise a model thus formed and thereby also the state of our probabilistic convictions in respect to present and future sound events. By following the evolution of these quantities, we can obtain a representation that captures many significant structures of music.

A crucial stage in modelling that is based on the observation of random processes is the defining of certain information measures, calculated with account taken of the realisation of a random process and a statistical model, which may be dynamically updated by analogy to the unfolding of the process.

[20] Samer Abdullah and Mark Plumbley, 'Information dynamics', *Centre for Digital Music, Queen Mary, University of London, Technical Report C4DM-TR07-01, Version 1.0 – July 18, 2007*, http://www.elec.qmul.ac.uk/people/markp/2007/AbdallahPlumbley07-tr07-01.pdf, accessed 3 November 2011.

By dividing the time axis into intervals of time indicating infinite past and future and hypothetical finite present, we can group observations of the analysed process into three random variables (*Z*, *Y* and *X*), corresponding respectively to three intervals of time. In effect, we obtain the observer's probability distribution $p_{XY|Z}$. If the random variables are discrete, then $p_{XY|Z}(x,y|z)$ denotes the probability of a situation in which the observer expects to notice *x* and then *y*, taking into account that he has already noticed *z*.[21]

In the case of measures relating to the phenomenon of perceptual surprise, the negative probability logarithm

$$\mathcal{L}(x|z) \triangleq -\log p_{x|z}(x|z)$$

can be treated as the 'surprisingness' of the occurrence of *x* on condition that *z* occurs.[22] As indicated by Abdullah and Plumbley, the entropy of the predictive distribution H(X|Z=z) is a function of the observed past *z*, and so it is a measure of the observer's uncertainty as to *X* before the observation ensued. Averaging over the two variables the surprisingness expressed by the above equation, we obtain the entropy rate of the process, H(X|Z), according to the observer's current model. In the context of the above considerations, we can obtain four information measures, which are surprisingness and its three averages over (X|Z=z), (Z|X=x) and (X, Z) jointly.

In the case of information-based measures of prediction, a crucial quantity obtained in a model based on the observation of random processes is the so-called *instantaneous predictive information rate* (IPIR). To this end, the information about *Y* should be considered through the observation that X=x, taking into account that we already know Z=z, defined as the Kullback-Leibler divergence (KL) between the predictive distribution over Y before and after the occurrence of X=x, and so

$$\mathcal{I}(x|z) \triangleq I(X{=}x,\, Y|Z{=}\,z) = D(p_{Y|X=x,\,Z=z}||p_{Y|Z=z})$$

where $p_{Y|Z=z}(y) = \int p_{XY|Z=z}(x,y)\,dx$, and the expression $D(\cdot||\cdot)$ is the divergence KL between two distributions. As in the previous situation, this is a function of the observations *z* and *x*. Averaging over the prediction X|Z=z, that is, the calculation $E_{X|Z=z}\mathcal{I}(X|z)$, tells us about the quantity of new information that we expect to receive from the next observation about the future. This is a useful indication of how much attention needs to be directed at the next event before it occurs. 'The average of the IPIR over the preceding con-

[21] Ibid.
[22] Ibid.

texts Z|X=x, that is, the expectation $E_{Z|X=x}\, \mathcal{I}(X|z)$, is the amount of information about the future, on average, by each value in the state space of *X*.'[23] As Abdullah and Plumbley indicate, averaging over both *X* and *Z* gives us the *average predictive information rate* (APIR) for the given random process model which is the average rate at which new information about the future arrives.

Hence the expression is reduced to the so-called conditional mutual information: $I(X,Y|Z)=H(Y|Z) - H(Y|X,Z)$. In this way, we again obtain a set of four measures, which are $\mathcal{I}(x|z)$ and its expectations over *X*, *Z* and (*X*, *Z*) jointly. It must be added that these measures are calculated in terms of the KL divergence, and so they are invariant for the reversible transformations of the considered spaces of observation.

One further research proposal is to obtain an information measure in a situation where an observer is making use of an openly parameterised model. This is a case in which the state of the observer's convictions would also cover the probability of distribution for the parameters Θ. It is emphasised that every observation could contribute to altering that state of convictions, that is, inform us about the parameters, the quantitative estimation of which would be made as a KL divergence between a prior and a posterior distribution $D(p_\Theta|X=x, Z=z||\ p_\Theta|Z=z)$, which can be defined as a 'model information rate'.[24]

Conclusion

In the case of information triggering a predictability of sound events, it is important for retaining a sort of perceptual curiosity that the musical structure display a certain balance in the revealing of predictable and unpredictable events. A structure that is too predictable and ordered arouses perceptual tedium, whilst a structure that is too unpredictable might cause the listener to treat music as unstructured, or 'random', devoid of characteristic features, and by the same stroke overly monotonous.

The selected methods of modelling music presented here are designed to assess the value of the predictability of perceived music. By referring to prior events (from the past), observing current events and creating expectations in respect to future events, and also assessing the information rate, these models take account of the real course of the listener's perception of music. We obtain cognitive insight into the analysed music by employing methods of a mathematical provenance, and so we have the possibility of examining music whilst taking account of the role of the listener, but with the use of objectivised methods.

[23] Ibid., 5.
[24] Ibid.

Acknowledgment

This work has been funded by the Ministry of Science and Higher Education appropriations for science in the years 2009 – 2011 as a research project Grant number: NN 105 134 737.

Translated by John Comber