Cue Abstraction, Paradigmatic Analysis and Information Dynamics: Towards Music Analysis by Cognitive Model

Geraint A. Wiggins
Centre for Cognition, Computation and Culture
Goldsmiths, University of London

Abstract

I present a preliminary study intended to explore the explication of the mechanisms of *cue abstraction* (Deliège, 1987, ff.) in terms of the information dynamics of music. I build on the apparent relation between cue abstraction and the *paradigmatic analysis* technique of Ruwet (1972). I argue that the introduction of the IDyOM information-theoretic model of musical listening (Pearce, 2005; Wiggins et al., 2009) into the study begins to account for the appearance of *cues* and *imprint formation*, in Deliège's terms. I illustrate how the model may in future contribute to systems which are capable of automating the process of music analysis.

Introduction

In this paper, I examine the ideas of *cue abstraction*, proposed by Deliège (1987, and subsequent papers) from the viewpoint of information-dynamic modelling of musical behaviour (Wiggins et al., 2009), with the aim of moving towards a method of music analysis in which a cognitive model is a strong guiding influence. To do so, I draw on the work of Nattiez (1975), who applied the *paradigmatic analysis* method of Ruwet (1972) to Debussy's *Syrinx*. Taking this approach allows me to draw a comparison between the ideas in cue abstraction and those in paradigmatic analysis, while supplying some supporting evidence for the importance of information dynamics in the human experience of music, while also building on the foundation laid by Potter et al. (2007), who examined the relationship between information dynamics and a music analyst's description of minimalist music. I use *Syrinx* as my example, since it is readily amenable to discussion in these terms, and since Nattiez's existing analysis provides some strong grounding.

I emphasise that the aim of this paper is not to present the results of a particular experiment, but rather to permit a preliminary exploration of the relationship between three different approaches to the same problem, especially in context of information-theoretic models of cognition. Therefore, I present the thinking "warts and all", rather than merely producing a "correct" answer and claiming victory. I believe that detailed argument is a more useful contribution to the nascent literature in information-theoretic modelling of music cognition than mere numbers.

Bringing computational modelling to this kind of study can do more than merely provide an alternative way of thinking about an observed phenomenon. If the model is an *explanatory* model, in the sense that I have discussed elsewhere (Wiggins, 2007), then it can serve not only to predict the phenomenon itself, but also to provide possible mechanistic explanations (amenable to subsequent empirical test) as to the mechanism underlying it. Further, these models may be used to generate predictions in an entirely rigorous way, which may not be so easy in purely theoretical models.

Statistical modelling of sequence learning (on which the model in question is based) is so *obviously* related to both paradigmatic analysis and cue abstraction, that it behoves one to ask the question "Is that obvious relation real?" In examining that question, one can also, in principle, supply a candidate method by which cue abstraction and the related processes take place, which is, in turn, an account of why paradigmatic analyses gives accounts of the structure of a piece that are musically reasonable.

Methodologically speaking, it is important to understand that, in this context, statistics of the kind conventionally supplied in psychological studies are not always meaningful: each piece needs to be studied individually, and the statistics across a corpus of pieces are not really at issue because composers are generally trying to do different things in different pieces—the issue of interest is the human reaction to those stimuli. Empirical support for the statistical model I am using is given elsewhere (Pearce and Wiggins, 2006; Wiggins et al., 2009), and it is not the purpose of this paper to supply more; rather, the questions asked here are, "Can that model give an account that corresponds with both cue abstraction and paradigmatic analysis?" and "Can it therefore be used for music analysis?" To investigate these questions properly, multiple pieces do need to be examined in the longer term. However, summary statistics are too blunt an instrument, in general, for these purposes, because this study is on the very boundary of what can be made objective (Potter et al., 2007). The reader can decide, in due course, whether the case has been made.

Background

Cue abstraction

It is scarcely necessary to summarise cue abstraction in a volume of this nature: Deliège (2001) has already laid out the theory as clearly as can be. However, it is appropriate to identify the primary aspects of the theory that will be relevant to the study, and to make precise my interpretation of their meaning.

The key idea, that of the *cue*, refers to "a salient element that is prominent at the musical surface" Deliège (2001, p. 237). As with many such definitions, the use of the word "salient" is problematised: one of the most interesting questions arising from the theory is how to quantify or formalise this elusive notion, so often used—and so rarely defined—in music psychology. In fact, it is part of the aim of the current work to begin to identify an objective definition, and I return to this later. Cues are, essentially, motifs (but not necessarily short ones) which are somehow distinctive, and which supply "a basic point of reference for the comparisons between musical structures that occur throughout the listening process" (Deliège, 2001, p. 238).

Processing of cues which are grouped together by virtue of some similar properties is termed *imprint formation*: "the cognitive processes at work when there is an insistence upon using the same cue by a composer—through literal repetitions or more or less varied elaborations" (Deliège, 2001, p. 238). It follows, therefore, that an *imprint* is an abstract summary of what was salient about the cues as they were identified as being within a particular group; it is important to note that, because elaborations are allowed, something more subtle than simple literal memory is supposed here.

The cues (creating and supported by their respective imprints) provide milestones along the timeline of a piece, marking section boundaries, and helping to group sections together by similarity of one kind or another. Each one of these grouped sections is viewed as a structural marker, which then allows elucidation of the structure of the musical narrative in terms of similarity between inhabitants of the groups. It is this process of structuring that interests me the most here, in context of my comparison with paradigmatic analysis.

Deliège (1987, 1996, 2001) and many others supply supporting evidence for the theory as a description of processing, and Deliège (2001) provides a hypothetical mechanism by which it might work. Here, I aim to assist two aspects of that mechanism: the ability to detect a cue *qua se* and then to identify references to cues, in such a way that an imprint is formed.

Paradigmatic analysis

It is often interesting to consider the different sides of the music/psychology divide in music psychology, and to observe the parallel development of musicological approaches to problems which are also considered in psychology. Paradigmatic analysis, proposed by Ruwet (1972) and famously applied by Nattiez (1975), may be one such example, in its relationship with cue abstraction.

In paradigmatic analysis, collections of *salient elements* of music, *paradigms*, are identified by the expert musicologist on the grounds that they are repeated, *literally or with more or less varied elaborations*, throughout the piece, and/or that they have contextually distinctive musical features in common. *Each one of these units is viewed as a structural marker, which then allows elucidation of the structure of the musical narrative in terms of similarity between inhabitants of the paradigms.* So there seems to be a strong relationship between Ruwet's paradigms and Deliège's cues, even though the former come explicitly from the expert analyst and the latter implicitly from the ordinary listener[1]. While Deliège evidently seeks a rigorous characterisation of what constitutes a cue and what defines similarity relations between cues, Ruwet does not, being content with the informed, subjective judgement of the expert analyst. In this sense, Deliège's theory may be seen as (among other things) an explication of paradigmatic analysis, in terms of an "archetypal listener".

Nattiez (1975) applies paradigmatic analysis to a well-known piece by Claude Debussy: *Syrinx*, for solo flute (see the Appendix). It does not follow that paradigmatic analysis will only apply to monodic music (Smaill et al., 1993, apply it to piano music by Charles Ives), but doing so helps to make the example clear. Nattiez (1975, pp. 332ff.) gives his analysis on two levels, described by two tables, Tableaux[2] I and II, working at a larger and smaller scale, respectively. There is what seems to be an error in Nattiez's analysis: in Tableau I, the letter X is used to identify two apparently distinct paradigms, which are treated as though they are unrelated (though I must acknowledge that the second application is abstract enough to render this judgement difficult), and, later (p. 348), the first of these paradigms is referred to as B_2; to clarify, I use this latter notation, while retaining Nattiez's layout.

The opening phrase of *Syrinx* and the corresponding Tableau I analysis are reproduced in Figure 1, and the Appendix gives the unbroken score, annotated with Nattiez's paradigm labelling; Fig. 1 also includes graphs of the output of the model, corresponding with the musical fragments, details of which are explained below. Tableau I specifies 11 large-scale paradigms in *Syrinx*, three of which (V, W and Z in Nattiez's notation) do not recur, and which, therefore, are not of interest here, as paradigms in themselves. Of the paradigms which do recur, one (Y) is sufficiently tenuous in its "repetition" to be omitted (see Fig. 1), and three (B_2—X in Nattiez's Tableau I—, D, E) are arguably the same paradigm, depending on the exact notions of similarity being used; this is made explicit in the analysis result. Nattiez's analysis clearly shows how the structure of the work is defined by repetition and modification of paradigms, and is unusually explicit about the nature of the similarities used. This is visible in Fig. 1 in the repetitions of letters (e.g., A) and modifications of letters (e.g., B

[1] One might argue that music analysis, from a listener perspective, is precisely about identifying what structure an archetypal listener will hear. Even if one does not accept this view in general, even the most cursory reading of Nattiez's analysis text clearly demonstrates thinking which is similar to that behind cue abstraction, so it still applies here.

[2] *Tableau I* and *II*, respectively, here refer to Nattiez's original; *Table 1, 2, ...* refer to tables in the current paper.

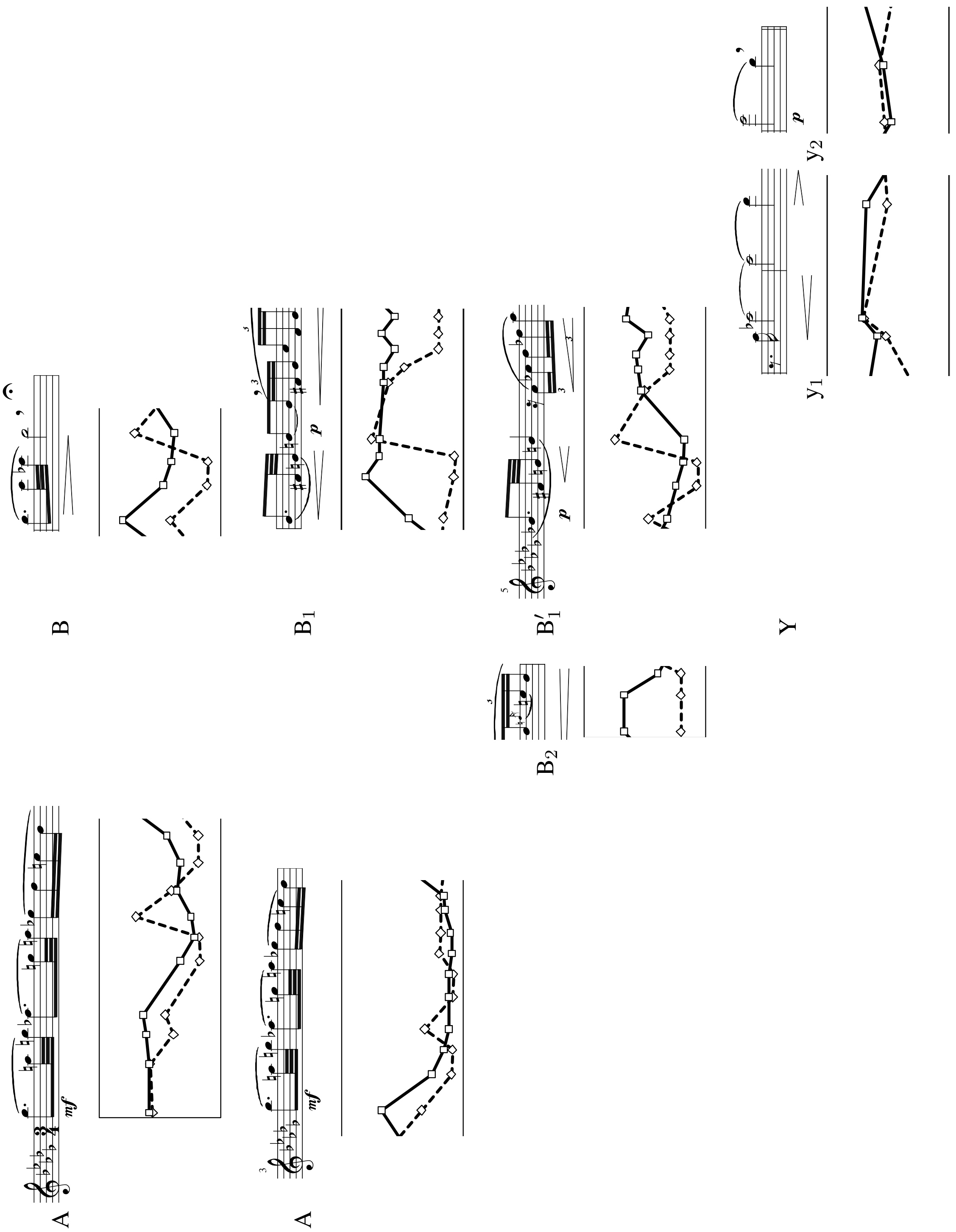


Figure 1: Nattiez's Tableau I (large-scale) paradigmatic analysis of the introduction to *Syrinx*, with the two information-theoretic signals interposed below the relevant notes.

and B_1): the letter identifies the paradigm as introduced, and the modifications (subscripts and primes) indicate (not entirely systematically) the modifications of its subsequent occurrences.

Tableau II is a much more granular analysis, mostly at the level of individual beats (though a notated beat is long in *Syrinx*, and therefore can contain quite a lot of information). It describes the basic building blocks of the piece and how they fit together to underpin the structures in Tableau II. However, although it is at a different scale from Tableau I, it is very easy to see operations akin to cue abstraction in the analysis: similarity, repetition, elaboration. I return to this in more detail below.

Since there is an evident consonance of intent between the theory of cue abstraction and the paradigmatic analysis technique, Nattiez's analysis forms a useful reference against which to evaluate attempts to automate aspects of cue abstraction. In the following sections, I examine the information dynamics of *Syrinx*, estimated by the Information Dynamics of Music (IDyOM) model (Pearce and Wiggins, 2006; Wiggins et al., 2009), in context of the analysis, with the aim of seeing how the information theoretic methods might serve as a mechanism to underpin and help account for cue abstraction in the context of (probably partially) automated music analysis.

Information-theoretic analysis of *Syrinx*

The IDyOM model

I am using a model of the information dynamics of music that is based on a statistical model of pitch expectation, invented by Pearce (2005). The model is a complex one, involving a short-term and a long-term memory, and a complicated process for matching variable-length, monophonic sequences of notes with structures recorded in those memories. A compact description of the model is given by Pearce and Wiggins (2006). The IDyOM (Information Dynamics of Music) project has extended this model to use the information theory of Shannon (1948) to predict not only expected pitches, but also the strength of expectation itself, in doing which it correlates well with human responses ($r = .91$; Pearce et al., 2010). Since strength of expectation is the converse of unexpectedness, or perhaps surprisingness, my colleagues and I have proposed this signal as a component of musical comprehension, following Narmour (1990); indeed, we have used the model to predict melodic segmentation, using the method Narmour proposed, in various styles of music, with some success (Pearce et al., 2008). More recent work has demonstrated that the model also reliably predicts particular kinds of neural activity in the centro-parietal area (Pearce et al., 2010).

In the present study, I use a simplified version of the model, featuring only the so-called short-term model[3] (STM), which is concerned with simulating the on-going experience of listening to a piece of music, but not in context of background enculturated knowledge of other music. Thus, I focus on local context only, and omit factors concerned with enculturation into particular musical styles; these wider issues are deferred for later study.

The representation used in IDyOM presupposes note-like percepts, with discrete properties of pitch and (metrical) time; nuances of tuning and expressive timing are not represented. Rests are represented as gaps between notes, rather than as events themselves. These notes are placed in time-sequence, and then the model extracts sequences of basic properties, such as pitch, duration, etc., and of derived properties such as pitch interval and inter-onset interval, which it uses to make its predictions. These sequences are known as *viewpoints* (Conklin and Witten, 1995); each one notionally includes all the notes from the start of the piece to the current point of "listening". The data thus represented are used to build a Markov model of the corpus. IDyOM is unusual as a Markov model in two

[3]Not to be confused with short-term *memory*.

ways. First, it uses variable-length contexts in its predictions, the complicated detail of which process is described by Pearce and Wiggins (2006); the key point is that the computed probability mass is distributed between multiple contexts of different lengths in a way which has been empirically validated (Pearce and Wiggins, 2006; Pearce et al., 2010). Second, it maintains multiple Markov chains, modelling different features of the sequence data in parallel; each chain makes its predictions and they are combined in a weighted sum, where the weights are determined by the note-wise entropy of the feature. Thus, there is an on-going interplay between the different features as the music proceeds.

A model is built of the piece being studied, note by note. As the piece proceeds, it is possible to estimate the likelihood of the next, "unheard" note at each point, on the basis of the sequences already recorded; the prediction is made in the form of a likelihood distribution, which is initially flat, across the symbols used in representing the piece. Since I am using no background memory of other pieces, but only memory of the current one as it proceeds (as though hearing the piece for the first time), it follows that the predictions made are based entirely on the motivic structure of the piece being studied, and there is no extrinsic contextual data (e.g., tonal theory) in play. Once a note value is "heard" it is possible to say, from the distribution, what its likelihood was. Following Shannon (1948), the *information content (IC)* of the note is estimated, in context of the piece so far, by taking the negative log, base 2, of its likelihood, p:

$$IC = -\log_2 p.$$

This can be explained as an estimate of the number of binary digits required to transmit the information in the note, in context of the memory model developed so far in the piece.

There is one aspect of the data which is not modelled directly from the score: the *accacciature*. There are two ways in which one might represent these musical features: as separate notes in sequence, or, following score notation, as an ornament of the following note with which they are associated. This is important because of two factors. First, which is the perceptually more valid representation? The easily-observed behaviour of performers, who usually vocalise such things as a single "ty-aaa", rather than by singing two separate notes, may perhaps give a hint here; but it is important to make a clear choice, because this issue straddles the boundary of what constitutes a "note" in music perception, and that is exactly where the abstraction boundary of the IDyOM representation lies. Second, on a more practical note, if one represents the *accacciature* as separate notes, what is their duration? Because the duration is very small, much smaller than values elsewhere in the piece, it attracts disproportionate values in the information-theoretic analysis. Fortunately, there is a heuristic that can help to answer the question: since the whole information-dynamic approach is motivated by the idea that the perceptual inputs are being *compressed*, the model can determine which representation is more readily compressible. This can be explained as choosing the representation which is most compact, in the sense that it requires the fewest bits to describe the musical content. I base this approach on the hypothesis that brains are essentially information-compressing devices; it was successful in the studies referred to in my brief introduction to the model, above.

The representation using separate very short notes for *accacciature* requires a mean 2.71 bits per note to represent the piece, whereas one using a separate dimension to denote the difference between a note with or without an *accacciatura* requires only 2.49 bits per note, despite the use of two dimensions where one previously sufficed—a substantial difference. I have therefore used the two-dimensional representation in the rest of the paper. The selected representation is composed of two viewpoints: duration is predicted from sequences of durations, and also from sequences of $\langle$duration,*accacciatura*$\rangle$ pairs, where the second element is either a 1 or a 0, indicating whether or not the note whose duration is given is preceded by an *accacciatura* or not.

In this study, I use two sub-models learned using the above approach, one of which is based on the note duration and accacciatura viewpoints only, explained first, and one of which is based on a complex predictor of the pitch viewpoint; this I explain later in the paper. The two time-variant information content signals produced by the sub-models are labelled $IC_{duration}$ and IC_{pitch}, respectively.

Hypothesis: the relationship between information content and imprint formation

I begin with the hypothesis that a major factor in the formation of cues (and, subsequently, imprints) is the attraction of attention (conscious or not) of the listener to the cue, in context; this is in line with Deliege's qualitative description, above, and with Nattiez's argument for the identification of paradigms in context: for example, units that "constituent un paradigm en raison de l'importance *accordée* au triolet et à l'appoggiature"[4] (Nattiez, 1975, pp. 330–331, original emphasis). Further, following Narmour (1990), I hypothesise that such attraction may be due to relatively high information content in context, which is associated with perceptual surprisingness, especially a sharp rise in that value. Note that the contextual aspect means that the same motif/cue may function in different ways, or, indeed, not at all, depending on where it is found. Conversely, if a cue is repeated, literally or approximately, the information content of subsequent occurrences should be lower than the surrounding non-cue context, and this would perhaps be a predictor of imprint formation. To test this hypothesis, I will compute the information content of the successive notes of *Syrinx*, using the two IDyOM sub-models (of duration and pitch, respectively, introduced above), and compare the relative values as the piece proceeds with Nattiez's analysis, using the analysis as a proxy for a detailed cue abstraction account.

Analysis: cues in Syrinx, and their information content

Preamble

The analysis will consist in identifying correspondences between the two information content signals, IC_{pitch} and $IC_{duration}$, and various structural aspects of the piece, as identified by Nattiez's analysis. To make these comparisons, which are estimates of agreement between two sets of binary judgements, I use Cohen's κ (Cohen, 1960), which is calculated thus:

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)}$$

where $P(a)$ is the actual agreement, calculated from the known data, and $P(e)$ is the likelihood that such agreement would be by chance. Landis and Koch (1977) give a qualitative scale of notional significance to κ: below 0.0 means no agreement; up to 0.2 is slight agreement; 0.2 to 0.4 is fair; 0.4 to 0.6 is moderate; 0.6 to 0.8 is substantial; and above 0.8 is almost complete agreement with 1.0 being the maximum; I use this scale as a rule of thumb, noting that it is not agreed by all statisticians. In some cases, I will use κ as an objective function for optimisation, and, in these circumstances, its absolute value will be of no interest nor consequence; but, in all cases, Cohen's κ tends to the conservative, and so, apart from a certain pathological case, is safe to use. That case arises in circumstances where relatively very few of the datapoints are in one category: agreements with relatively very many of the other category swamp the statistic, and it simply increases, topping out when all predictions are in

[4] "…constitute a paradigm by reason of the importance *accorded* to the triplet and the appoggiatura." To be clear on the terminological distinction between *accacciatura* and *appoggiatura*: I use the former to refer to the crossed-out form, and the latter otherwise. Both forms exist in the piece.

the larger category. In these circumstances, which are easily detectable by observation, precision and recall can sometimes be used, but not always.

This leads to an important methodological point. In this kind of modelling work, one can use statistics in ways which are complementary to those familiar in analysis of conventional empirical results. In particular, as mentioned above, various statistical analyses can apply leverage to particular dimensions of the model, and identify which of a range of possible parameterisations gives the best fit to the empirical data, thus identifying which is the best model. Of course, when one does this, one faces all the dangers of over-fitting that arise in supervised machine learning (Cleeremans and Dienes, 2008; Wiggins, 2010), and model selection (Honing, 2006) and so one must guard against them. I will take this approach in some of the analyses below, and, since this is a study of one piece of music, I will not make strong claims about conclusions. Rather, I emphasise that the ideas that are generated here have only the status of new hypotheses, which can be tested on different data, and by further empirical study, and not that of music-cognitive results in their own right.

$IC_{duration}$ and main beat

The first correspondence I consider is in $IC_{duration}$, the information content signal generated by predicting durations from sequences of durations and *accacciature*. The full detail of the signal is shown as the broken line in Fig. 2, where the x-axis is in bars (noting that bars 29 and 30 have only two beats), with grid lines at beats. *Syrinx* contains relatively many distinct note duration values, for such a short piece, and the results suggest that this is not perceptually insignificant.

In context of duration predictions, it is important to note the difference between duration and pitch as experienced parameters of notes: pitch is perceptible almost immediately the note starts; but duration is imperceptible until the note has ended. My working position is that retrospective listening (Narmour, 1990) addresses this point, and enables processing of pitch and duration in the majority of notes together, while still doing so in the perceptual present; this is in line with current thinking about the perceptual present in both music and language.

From the figure, one can immediately observe an apparent correlation between peaks in the $IC_{duration}$ signal and the beats of the score. To investigate this, using a method akin to, but simpler than, that of Pearce et al. (2008), I differentiate the signal, and search for rising values, comparing their positions with main beats where there is a note onset (otherwise $IC_{duration}$ is undefined). The two phenomena agree 84% of the time ($\kappa = .62$; substantial agreement). However, observation suggests that some of the rises in both the original signal and its derivative are extremely small, and might well be at the level of noise in the method. To address this, I introduce a gradient minimum of 1 bit per beat, and a minimum size of peak (measured between successive values) of 0.14, both these values being chosen by observation. The agreement between these larger rises and the main beats is 89% ($\kappa = .71$: substantial agreement).

Remembering that the model has been shown to predict human behavioural experience of expectedness, and related neural activity (Pearce et al., 2010), using it allows me to draw together musicology and cognition, and propose an explanation of why this compositional device works. Since evidence has previously been given that peaks in these information content signals predict segment boundaries (Pearce et al., 2008), I suggest that the piece is distinctly segmented into beats by the information in its rhythm; the one-beat length of the majority of Nattiez's Tableau II paradigms to some extent supports this position. This segmentation may help support the temporal structure of the piece, in the context of its extreme fluidity and very slow main pulse. One might hypothesise that the wide variation in note duration on these beats is deliberately chosen to create this effect, increasing uncertainty at each pulse and thus reinforcing the very slow tactus.

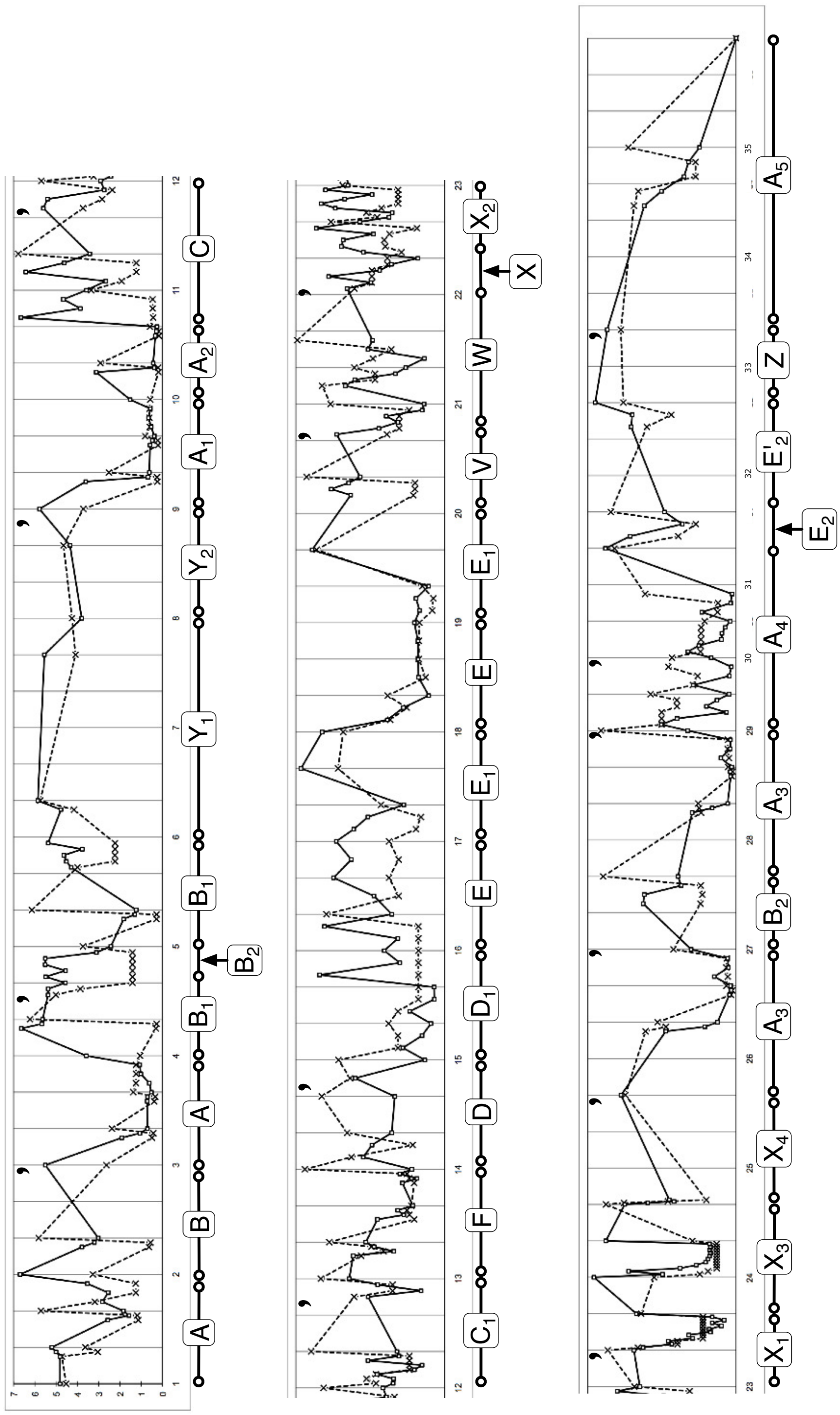


Figure 2: Information content of duration predicted from duration (broken line) and Information content of chromatic pitch predicted from the complex system described in the text (solid line). Nattiez's analysis is shown below the graph, and the commas in the score are placed above it.

My suggestion that sharp rises in this signal help to make the beat clear would evidently be undermined by the opposite (sharp drops in the information content signal, on beats), and this something that can be checked easily. Only one decrease happens in the entire piece, on the 3rd beat of bar 7, in the middle of paradigm Y, with a very small gradient of -0.053, and is ruled out by the magnitude threshold applied above. I therefore argue that this one aberration does not undermine my segmentation proposal.

Nattiez's Tableau II identifies small-scale paradigms, most of which have a length of one beat. The model above, with different parameters (gradient threshold = 2.4 bits per beat; absolute threshold = 0.13 bits) predicts the Tableau II boundaries with $\kappa = .57$ (substantial agreement). Note that the larger gradient threshold is requiring larger rises in $IC_{duration}$ to motivate boundaries, suggesting that Nattiez's Tableau II boundary selection process is more selective than mere beat identification, as one would expect. This result again supports the suggestion that perceptual segment boundaries correspond with sharp rises in information content (Pearce et al., 2008), both by reference to Nattiez's analysis and to the simpler main beat. They also provide a mechanism which might support my claim, above, that the piece is carefully constructed to maintain a sense of rhythm even though its tactus is extremely slow. However, the maximal agreement with the tactus is much stronger than the maximal agreement with with Tableau II, and so I conclude that $IC_{duration}$ is detecting the former and not the latter, if it is detecting anything.

One remaining obvious question to ask is whether there is a correspondence between sharp rises in $IC_{duration}$ (perhaps thresholded, as above) and the larger-scale boundaries in Tableau I. In this duration sub-model, this seems not to be the case: the model does not locate a maximum in κ, and one observes the swamping effect mentioned above: a higher agreement score is obtained by assuming no boundaries than by using the model. Precision and recall measures are equally unsuccessful. Therefore I conclude that $IC_{duration}$ is coinciding with the tactus, and not with the musically pertinent segments of Nattiez's analysis, except where these two coincide. Part of the listening experience, therefore, probably resides in the interaction between this small-scale effect, and the larger scale, pitch-based segmentation, which I discuss next.

IC_{pitch} and segmentation

The model I have used for pitch prediction is substantially more complicated than the duration prediction model, above. The attempt here is to model more than just the temporal content of the music, and to bring in not only *basic* viewpoints describing basic quantities such as pitch, but also *derived* viewpoints, describing quantities derived from basic ones, such as pitch interval, and *linked* viewpoints, which allow predictions from combinations of parallel viewpoints. The linked viewpoints are important, because they can capture regularities which arise in parallel in different features of the musical stream (for example, particular duration sequences which always occur with particular intervals between their notes—and this is very relevant here, for example in the opening 1-beat figure of *Syrinx*). The basic viewpoints are chromatic pitch, metrical duration and *accacciatura*; the single derived viewpoint is pitch-interval. All possible combinations of linked viewpoints are available, so long as they contain pitch information (otherwise, they cannot predict pitch). The pitch representation in all cases is purely semitonal, and does not include enharmonics, though it is important to understand that this does not presuppose particular tuning: it presupposes 12-way pitch categorisation, but that is all.

Here I focus on predicting pitch: only basic viewpoints can be predicted, for reasons of statistical validity (Pearce and Wiggins, 2006). I do not make a strong claim about the vocabulary of viewpoints from which I construct this particular model, and, indeed, there are several others, such as pitch contour, which are clear candidates. However, in the current work, I have limited the selection to improve tractability: once I have understood how these viewpoints work in this context, I will move on to others. Because there are multiple viewpoints, there is a choice to be made about which ones to include: it does not follow that adding more viewpoints makes a better model. As with the duration model, above, I make this choice objectively, by minimising the entropy of the resulting model.

The model that is selected as the most compact is one which predicts (chromatic) pitch from a combination of pitch-interval, pitch $\otimes$ pitch-interval, pitch-interval $\otimes$ duration, and pitch-interval $\otimes$ duration $\otimes$ *accacciatura*. Note that pure chromatic pitch is discarded by the selection process; it is easy to see why this would be: octave equivalence is important here, and the pure pitch viewpoint occludes it. The system's selection of the link between pitch and duration suggests that that relationship may be worthy of further, separate musicological study; here, however, I am interested primarily in the information content predicted by the model. The IC_{pitch} signal is shown as the solid line in Fig. 2. The system's selection of pitch $\otimes$ pitch-interval is interesting, since this partly encodes the tonality of the piece[5], in an implicit tonal profile, and confirms that tonality is important; likewise, the selection of a separate link combination including *accacciatura* supports Nattiez's argument that the embellished figures are indeed significant. In short, the compression heuristic suggests that relative pitch gives a more compact representation than absolute, and that using quasi-tonal information (the relationship between pitches and the intervals from them, which encodes information about scale) is more compact than not doing so. Both of these issues point to a hypothesis regarding general pitch and tonality perception: perhaps these phenomena arise because, in the context of Western musical culture, they permit more efficient representations of musical memory.

Now, I consider the segmentation of the piece predicted by this model. Again, according to the information-theoretic interpretation of Narmour (1990), a segment boundary coincides with a reduction in information content towards the end of a segment, followed by a sharp rise on the first event of the next, as found by Pearce et al. (2008), and in the duration study, above. Such structures can be clearly seen in Fig. 2: the question is whether they coincide with musical boundaries. Ways to answer this include considering the boundaries of Nattiez's paradigms and the phrase and comma markings in the score. I begin with the last of these, since it is most straightforward.

Commas

I now consider the commas in the score (see Appendix). The first thing to note is that there is some doubt as to the provenance of this edition, and therefore one cannot be sure if these markings are Debussy's own. However, they are performed in the three different recordings to which I have access, so at least some performers take them seriously. There are 15 commas marked in the score, seeming to denote performance breaks, and therefore, presumably, strong segment boundaries. 13 of them correspond with the characteristic IC_{pitch} rises described above. The rises and gradients are quantified in Table 1, and the commas can be seen against the IC_{pitch} signal in Fig. 2 and in the score in the Appendix.

The first of the two unpredicted commas is that in bar 4, after the fourth note; the model predicts a boundary after the first note in bar 4, and indeed, but for the overriding phrase markings this would not be an unreasonable interpretation. The second is in bar 33, and here a *decrescendo* carries through from the previous note, again giving a strong reason to override the model's prediction, on the basis of data unavailable to it. Of course, one might argue that one reason for placing these commas in the score is to instruct the performer to emphasise segments precisely where they would otherwise *not* be

[5]Note that the word "tonality" is used broadly here: this is not text-book harmony!

Table 1: The positions of the commas in the score, and the change in IC_{pitch} across them; the position given is the beat immediately following the comma. 13 of the 15 correspond with rises.

Position				Position				Position			
Bar	**Beat**	**Rise**	**Grad't**	**Bar**	**Beat**	**Rise**	**Grad't**	**Bar**	**Beat**	**Rise**	**Grad't**
3	1	2.50	3.73	14	3	1.86	11.18	27	1	1.72	20.61
4	$2\frac{2}{3}$	-0.25	-1.14	20	$3\frac{1}{6}$	1.11	2.84	29	1	2.01	24.13
9	1	1.44	4.32	22	$1\frac{1}{6}$	1.20	2.53	30	1	1.10	19.82
11	$3\frac{1}{4}$	2.18	5.24	23	2	0.28	0.83	31	2	5.97	14.33
12	$3\frac{1}{2}$	1.36	2.73	25	3	2.27	2.37	33	2	-0.57	-0.85

perceived, so it may be the case that a perceptual model cannot be expected correctly to predict all such phenomena. However, it is interesting to note that the Narmour-inspired correspondence between score annotations often taken to mean segmentation (and clearly meaning that here in some cases) and the model's output is as strong as it is.

Predicting Nattiez's paradigm boundaries and imprint formation

I begin to explore whether there is any relationship between the model's output and Deliège's imprint formation by taking Nattiez's paradigm boundaries as given, and examining differences between those which introduce new material, such as the first occurrence of paradigm A, those which repeat material, such as the second occurrence of paradigm A and those which present variations on previously exposed material, such as A_1 and (in extremis) A_5. The expectation would be that the first case would contain relatively large amounts of information, the second would contain relatively small amounts, and information content in the third case would vary, depending on the degree of similarity between the material and the original of which it is a variation.

I test this proposal by comparing the average information content across each paradigm with a threshold, chosen to maximise agreement between the two categories so produced and, first, Nattiez's analysis, on the (strong) assumption that unmodified letters denote new material while modified ones denote variations. A threshold average of 2.5 bits categorises the notes to match Nattiez's analysis with 73% agreement ($\kappa = 0.51$; moderate agreement). However, the most cursory examination of the score shows that some paradigms that share the same letter are really rather different (though it is still clear why Nattiez grouped them together in terms of musical function: mostly, the first half of the paradigm is the same and the second varies considerably, or the two paradigms are of very different lengths, and so the match is with an initial subsequence of one of them). The paradigms with modified letter names, which appear clearly different from their "parent" unmodified segments are: B_1, B_1', E_1, X_1, X_3, E_2, A_5. B_2 differs too, and this is an interesting case for the current analysis: because the paradigms are of different lengths, the model is unable to identify the match that Nattiez has annotated[6]. An alternative comparison, marking these paradigms as new material, achieves a maximum agreement at a threshold average of 2.2 bits, of 97% ($\kappa = .93$; almost complete agreement). This is a very promising match between a reasonable variant of Nattiez's analysis and the output of the model, but with the analysis

[6]In fact, the IDyOM system is capable of identifying such matches, but I have not included the necessary advanced category of *threaded* viewpoints in the current study, in the hope of simplifying what is a rather complex narrative. See Pearce and Wiggins (2006) for details.

Table 2: Agreement between prediction of new vs. old material, by comparing average IC_{pitch} across the time period shown, with a new vs. old material judgement based strictly on Nattiez's letter notation as explained in the text.

	Nattiez's letters		Wiggins' modifications	
Time period	**Threshold**	κ	**Threshold**	κ
Semiquaver	2.02	0.30	1.78	0.59
Quaver	2.00	0.32	1.75	0.62
Crotchet	2.4	0.40	1.9	0.59
Bar	2.5	0.35	3.3	0.54

as a given: so while the conclusion can be drawn that information about the paradigms may be in the signal, the model is not in any strong sense predicting. A more interesting kind of match, now that there is reason to believe that the required information is present, would be the converse, where the model predicts both the segmentation and the newness of the material. I examine this possibility next.

To do the comparison, I consider the information content of the piece averaged not over the variant of Nattiez's paradigms as givens, but over four different fixed time periods (semiquavers, quavers, crotchets, and bars), and asking, first, whether there is a threshold which will allow grouping of the values produced into two categories (again, new/not new), to match Nattiez's analysis. The thresholds and agreements achieved by doing so are shown in Table 2. The agreement is quite robust against the averaging time period, which is promising; it is maximal with quaver time periods and a threshold of 2.00 ($\kappa = .59$; on the boundary between moderate and substantial agreement). This relatively strong agreement suggests again that the information required to distinguish new from old material and possibly therefore to identify paradigm boundaries is indeed in the IC_{pitch} signal. The acid test of whether there is meaningful information here will be in a detailed analysis of the difference between the model's prediction and my modification of Nattiez's paradigms: if the model can predict which *parts* of the modified paradigms are repeated/modified and which are genuinely new, then it is predicting something real.

The first place in the piece where a literal (as opposed to purely gestural, like the similarity between B and B_1) but non-identical similarity happens is at the start of B_1' at bar 5: the first 3 notes are identical with those of B_1, the fourth is the same pitch and a different length, and the rest are very different. The model correctly suggests that the first two beats are familiar material, while the third is new. The data can be seen in Figs. 1 (for the introduction) and 2.

The next point at which such a comparison can be made is between the third beat of bar 9, in paradigm A_1. This is an octave transposition of the opening bars of the piece. The model suggests unfamiliarity during the first beat, but then identifies the transposition, and the information content decreases throughout A_1, finishing with a fairly low average across the segment.

Next, the two beats of A_2 are identical with the first two of A_1, while the first beat of C is rhythmically identical to the third of A_1, but the pitch pattern is different, its contour being reversed. The model suggests that A_2 is familiar material, but that C is not, again corresponding with what is musicologically evident from the score[7]. This point perhaps best emphasises the relationship between the listening experience, Nattiez's analysis and cue abstraction. Paradigm A is evidently a very salient

[7]It is, perhaps, appropriate at this point to recall that the model has been given no musicological knowledge at all—all it has is its simulated perception of the piece, which is based on implicit learning.

trope in the piece, with its initial dotted structure, followed by a repeat or something else; this structure fits exactly with Deliège's notion of cue: "a salient element that is prominent at the musical surface", and, indeed, it forms "a basic point of reference for the comparisons between musical structures that occur throughout the listening process" (Deliège, 2001, p.238), since it forms an initial boundary for many of those structures. What seems to happen is that the first two beats of A_1 and the whole of A_2 are set up to frame the contrast between the third beat of A_1 and the first of C; and this is certainly my own subjective experience: the first beat of C attracts one's attention strongly, because it is the first time a rising structure has been experienced in this context. Once it has had time to notice the approximate similarity between A and A_1, the model predicts familiar material through A_2, and then a sharp rise in IC_{pitch} corresponds with the boundary between A_2 and C, as the model identifies the unfamiliar material, whose information content is accordingly high. It is interesting also to note that the rhythmic model is predicting familiarity (as one would expect), but that both signals rise from this point, with the very new material in C, and when this is repeated in C_1, though it is less familiar than A and its relatives, IC_{pitch} is still relatively high; what is more, the boundary between Nattiez's two paradigms is not predicted by the model, whereas the comma in the score (see above) is. Paradigm W, later, suggests that a boundary before the low E♭ is not unreasonable, but then the score confuses the whole issue by supplementing a comma with a phrase mark *and* a definite slur. Nattiez groups C, C_1, V and W together in his commentary, though the connection between V and C is a good deal less clear than that between V and W. So it is very difficult to draw conclusions here; but the point is that this is a different kind of construction from the initial cue in A. This is the interstitial material that cues coordinate, in Deliège's theory, and the model concurs: V and W are still predicted as new material, though not as radically new as C.

The later occurrences of paradigm A (A_3, A_4, A_5) are a greater challenge for the model, because they are varied in more extreme ways than A_1. A_3 surprises the listener, the first time, because its first note is suspended across the bar line, and therefore it is not recognised until the end of its second beat, and this can be seen, reflected in the IC_{pitch} signal, in Fig. 2; once that recognition is achieved (after the second note of the paradigm), information content drops and the model again predicts familiar material. The second statement of A_3 is (by now) familiar, however, according the model. A_4 is, yet again, identified as familiar, now beginning to demonstrate that the model is doing something more than mere identity matching, and, finally, A_5, while being radically different from the rest of paradigm A, shows a decreasing profile as it proceeds, indicating progressively stronger familiarity as more pitches are encountered: this process provides the characteristic decrease that Narmour argues should accompany the ending of a piece, and matches the strong sensation of relaxtion that this final phrase conveys to the listener (or, at least, to this listener).

There is not space here to analyse the interactions between paradigms D and E in detail, but there is enough variation in the signal to suggest that something important happens in the two (non-identical) occurrences of E_1. As Nattiez says, the triplet with its *accacciatura* forms a salient unit, and, again, one sees such units framing and emphasising other salient structures: in this case the two large upward leaps from crotchet to crotchet, which cause a strong rise in information content and attract the listeners attention equally strongly. Most segmentation theories would predict a boundary before the higher of each pair, and the score is marked clearly to avoid this: the *crescendo* and the phrase mark instruct the performer to avoid that interpretation. This is a further reminder of a limitation of the current study: the learning data includes neither dynamics (in the musical sense) nor phrase marks, and so cannot avoid the boundary prediction. This aside (for future work), the salient nature of the crotchets is again predicted.

Simulating the listening experience

While resisting making overly strong claims, one can draw an analogy between both of these signals and an important aspect of the listening experience: the feeling of recognition. It is not surprising that a model of this nature is able to determine what has been heard literally before and what has not. However, since Pearce et al. (2010) have shown that the signals it produces correlate (inversely) with the *psychological experience* of expectation, the way it does so may be important. What is more, the capacity of the model to generalize (for example, in its apparent recognition of the relationship between A and A_5) is an important part of its capacity as a cognitive model.

The gross effect to which I refer can be seen in Fig. 2. Consider bar 3, which is an identical repetition of bar 1/paradigm A. The average level of the signals is substantially lower in this bar than in the previous one, because the sequence has been observed identically before, and therefore can be predicted with some certainty. Bar 4, paradigm B_1 has new pitch information, however, and so it has higher information content. This effect is repeated throughout the piece, for example at A_1 and A_2. These sections, at which the information content is low, therefore, correspond, in Deliège's terms, with the process of imprint formation: they identify which pieces of musical material contribute to the imprint. In a more complex model, with background knowledge of other music, they would encode more information: for example, musical clichés and quotes would be detected, and this too could inform the imprint formation process. I emphasise that, since the imprint is more abstract than the individual notes, it does not do to suggest that the model is actually *forming* imprints. However, one can imagine a similar model, informed by the workings of this one, that might begin to simulate that process. This idea can be compared with Nattiez's analysis, too, since the analysis makes explicit what is viewed as (varied) repetition and what is not.

However, neither cue abstraction nor Nattiez's analysis operates on one level only: multiple cues can operate concurrently on different time-scales to give the feeling of structure in a piece; multiple levels of paradigm can express this. I now consider again the introduction of Syrinx, combining the two signals, $IC_{duration}$ and IC_{pitch}, into one analysis, referring again to Fig. 1. I use the notation *Bar:Beat* to refer to particular beats in the piece, with fractional *Beat* component where necessary to identify particular notes. Recall that IC_{pitch} includes information from the duration viewpoint, and therefore to some extent follows $IC_{duration}$ in places where duration contributes strongly to the whole.

At the start of beat 1:1, both signals are quite high, and almost equal: there is no information available to either. This remains true of IC_{pitch}, which is being presented with new pitches, but the rhythmic repetition in beat 1:2 reduces $IC_{duration}$: the beginnings of imprint formation of that important rhythmic figure. Immediately, then, here is a way in which the duration information operates independently of the pitch information, to define the piece's structure. Beat 1:3, however, is clearly new material, with new note durations and new pitches. The pitch information content remains medium-high, but the duration information quickly drops as the model becomes accustomed to repeating semiquavers; duration information yields to the more important pitch sequence. At the start of beat 2:1, is the first peak in IC_{pitch}, corresponding with the start of paradigm B.

At the start of beat 3:1, the model is again uncertain (information content is high) in IC_{pitch}, because of the long high note in bar 2, which leaves no matching context (not only is this the first minim in the piece, but there is no example of a repeated note, either), but $IC_{duration}$ is low in recognition of the repeated rhythm. IC_{pitch} provides the boundary between paradigm B, and the repeat of paradigm A takes both signals low, signalling imprint formation (and contributing the hypothesis that this is the phase we are most likely to remember from the piece).

The beginning of paradigm B_1, at the start of beat 4:1, is identified by a rise in IC_{pitch}, though

$IC_{duration}$ stays low for one beat; this can be seen as further emphasis of the dotted rhythm imprint, and diversion of attention to the new pitch information.

Bar 5 begins (5:1–5:2) with a literal repetition of bar 4, which has the same effect as before, with both signals low, again corresponding with imprint formation and an experience of familiarity. Both signals now remain high, corresponding with the feeling of floating uncertainty produced by these long, high notes, until the end of the introduction.

There is not space here to analyse the whole piece in this way. However, Fig. 2 shows the full analysis, and may be compared with Nattiez's analysis and with the score; further correspondence of the kind described above will be found. While these results are not quantitative, they are suggestive of support for my proposal that a model of this kind can be usefully engaged in music-analytical tasks.

Conclusion

I have presented an interpretation of the relationship between an information-dynamic model of music cognition—one which, specifically, predicts *expectation*—and aspects of Deliège's cue abstraction on one hand and of Ruwet's paradigmatic analysis (as applied by Nattiez) on the other. I have identified quantitative and qualitative correspondences between the model's objective output and the rational but subjective analysis, which, though by no means perfect, are nonetheless promising. I have suggested how this information might relate to a process of cue abstraction, and how this model might contribute to a wider model which took account these factors to model the process of imprint formation. I argue that the correspondences seen here lend support to the IDyOM model as a model of some aspects of the *experience of listening* to music. This in turn raises the possibility that the model can be used itself as a music analysts tool.

There is plenty more useful work to be done on the theory of cue abstraction and I am pleased and honoured, on the occasion of its inventor's retirement, to be able to contribute to mapping out just a small part of that future potential.

Acknowledgements

The pitch expectation model underlying IDyOM was invented by Marcus Pearce (2005), and the information-theoretic aspects were developed under the EPSRC grant GR/S822220, "Information Dynamics of Music" and some of the writing was funded by its successor, EP/H01294X, "Information and Neural Dynamics in the Perception of Musical Structure ". Alan Smaill introduced the author to Nattiez's analysis of *Syrinx* in 1991. Irène Deliège supplied the inspiration—as for so many researchers in music cognition.

References

Cleeremans, A. & Dienes, Z. (2008). Computational models of implicit learning. In Sun, R. (ed.), *Cambridge Handbook of Computational Psychology*, pages 396–421. Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20(1)*, 37–46.

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, 51–73.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's grouping preference rules. *Music Perception*, *4*, 325–360.

Deliège, I. (1996). Cue Abstraction as a Component of Categorisation Processes in Music Listening. *Psychology of Music*, *24(2)*, 131–156.

Deliège, I. (2001). Introduction: Similarity perception ↔ categorization ↔ cue abstraction. *Music Perception*, *18(3)*, 233–243.

Honing, H. (2006). Computational modeling of music cognition: a case study on model selection. *Music Perception*, *23(5)*, 365–376.

Landis, J. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realistation Model*. Chicago: The University of Chicago Press.

Nattiez, J.-J. (1975). *Fondements d'une sémiologie de la musique*. Paris: Union Générale d'Editions.

Pearce, M. T. (2005). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. PhD thesis, Department of Computing, City University, London, UK.

Pearce, M. T., Herrojo Ruiz, M., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioural and neural manifestations of musical expectation. *NeuroImage*, *50(1)*, 303–314.

Pearce, M. T., Müllensiefen, D., & Wiggins, G. A. (2008). Melodic segmentation: A new method and a framework for model comparison. In *Proceedings of ISMIR 2008*.

Pearce, M. T. & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception*, *23(5)*, 377–406.

Potter, K., Wiggins, G. A., & Pearce, M. T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Scientiae*, *11(2)*, 295–324.

Ruwet, N. (1972). *Language, musique, poésie*. Paris: Editions du Seuil.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423, 623–56.

Smaill, A., Wiggins, G. A., & Harris, M. (1993). Hierarchical music representation for analysis and composition. *Computers and the Humanities*, *27*, 7–17.

Wiggins, G. A. (2007). Models of musical similarity. *Musicae Scientiae, Discussion Forum 4a*, pages 315–337.

Wiggins, G. A. (2010). Computer models of (music) cognition. In Rebuschat, P., Rohrmeier, M., Cross, I., & Hawkins, J. (eds), *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

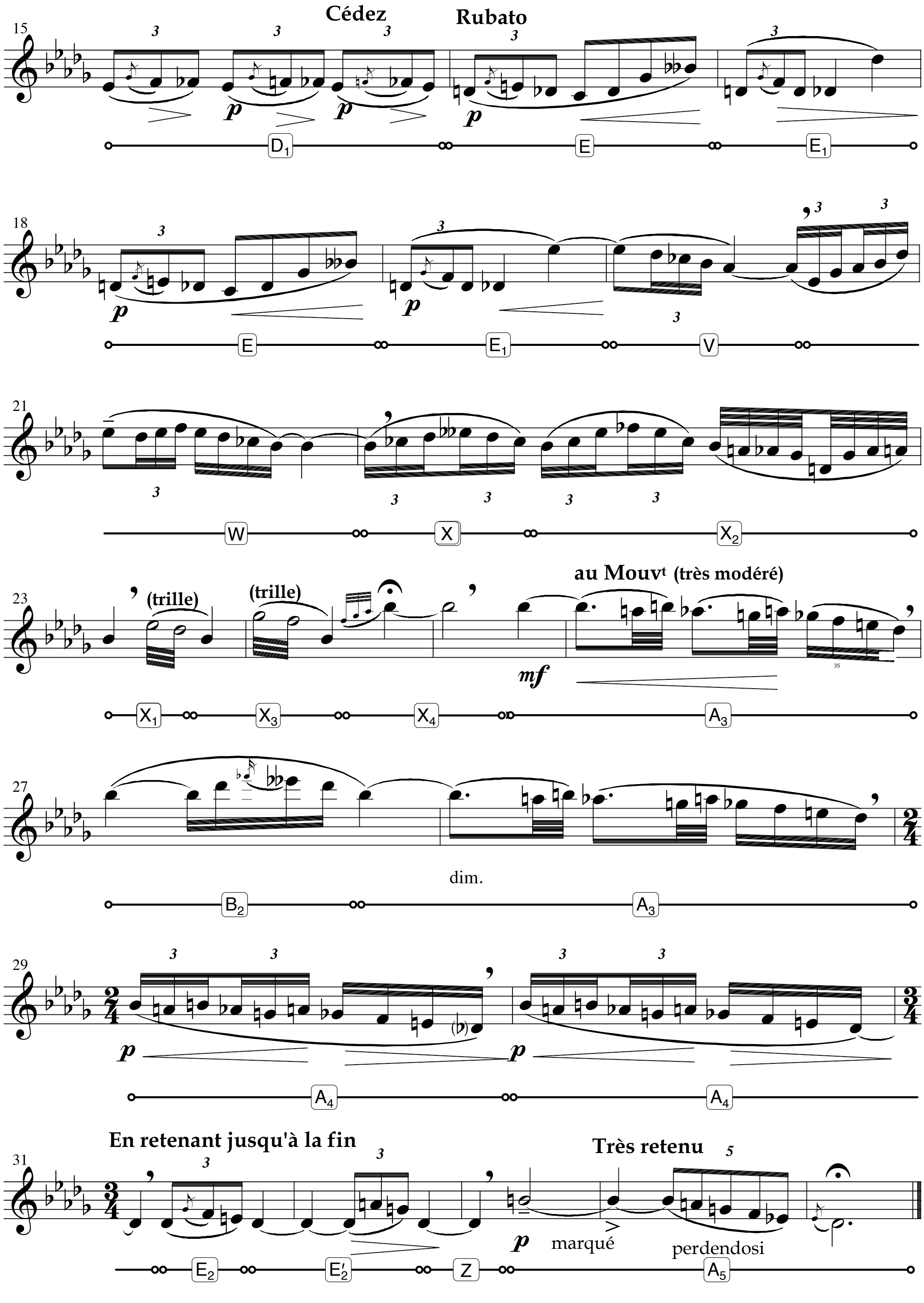
Wiggins, G. A., Pearce, M. T., & Müllensiefen, D. (2009). Computational modelling of music cognition and musical creativity. In Dean, R. (ed.), *Oxford Handbook of Computer Music and Digital Sound Culture*. Oxford University Press.

Appendix: Debussy's *Syrinx*

Syrinx

à Louis Fleury

Claude Debussy (1913)


Cédez
Rubato
D_1 E E_1
E E_1 V
W X X_2
(trille)
(trille)
au Mouvt (très modéré)
X_1 X_3 X_4 A_3
dim.
B_2 A_3
A_4 A_4
En retenant jusqu'à la fin
Très retenu
marqué
perdendosi
E_2 E'_2 Z A_5