An Objective Basis for Music Theory: Information-Dynamic Analysis of Minimalist Music

Keith Potter⋆, Geraint A. Wiggins† and Marcus T. Pearce†
Departments of Music⋆ and Computing†
Goldsmiths College, University of London
New Cross, London SE14 5SG, United Kingdom
{k.potter,g.wiggins,m.pearce}@gold.ac.uk

Corresponding author: G. Wiggins

Abstract

We present evidence for a relationship between two objective measures of the information dynamics of music and points of structural importance in the music as analysed by human experts. Our approach is motivated by ecological validity: rather than taking musical stimuli and artificially simplifying them to make their study tractable, we have sought and found music which is appropriate to our study. We give a novel, detailed analysis of one piece, Glass' *Gradus*, and show how the analysis corresponds with the information dynamics of the piece as heard. To show that this correspondence generalises, at least to music in a similar style by the same composer, we go on to analyse Glass' *Two Pages*. We suggest that this research provides further evidence that information-dynamic modelling is a worthwhile approach to the study of music cognition and also has the potential to be a powerful tool to increase objectivity in data-based music analysis.

1 Introduction

Analysis of minimalist music poses unique challenges arising from the peculiar characteristics of its subject. One such characteristic reflects the use of compositional processes that are intended to be directly perceptible on the surface of minimalist music and (often) uniquely to determine its interpretation. Bernard (1996) suggests that, if there is nothing but surface, then there might seem to be little point in an analyst hunting for more subtle structural interpretations which do not exist. Such difficulties have proven significant barriers to development of general methods for analysis of minimalist music.

We propose a new method for the study of minimalist music, involving quantitative analysis by a computational model of human perception of music. Given the difficulties noted above, our method has several advantages: first, it analyses only the information objectively available on the musical surface; and, second, it is appropriate to use a model of human perception to examine surface processes intended to be directly perceived by listeners. Furthermore, in keeping with common characteristics of minimalist music, the model we propose operates over very simple representational primitives and is inherently dynamic in its mode of operation.

However, we argue that our method is not limited to minimalist music, and that more complete models of the information dynamics of music will, in future, provide an objective basis for the study of human cognitive—and hence musical—responses to music of all kinds. The current work, therefore, is the first step on a long journey.

In the following sections, we first outline our perceptual model and, in this context, explain the methodology of our approach. Then we show how it can be used to predict the detailed structure of a musicologist's analysis of a piece of music, Philip Glass' *Gradus* (1968), and to identify points in the music which, we argue, constitute structural boundaries, and which therefore constitute salient parts of the musical structure. We next demonstrate how the same system (without *ad hoc* re-configuration) predicts the structure of another Glass piece, *Two Pages* (Glass, 1969), by comparison with section boundaries given by the composer and two analysts. In conclusion, we argue that models such as that presented here have the potential to form the basis of a more objective, and therefore more rigorous, approach to music analysis.

It is important to note that, because we present this discussion from a music-analytic perspective, and we do not apply rigorous behavioural verification to the outputs of our model—to do so would be to miss the point of our work. Rather, the aim is to refute or establish evidence for the validity of our model in the *musicological* context. Further psychological and neurophysiological validation is available elsewhere (Pearce and Wiggins, 2006; Meyer, 1956; Huron, 2006) and will be the subject of our future research.

2 An Information-Theoretic Model of Music Perception

Pearce and Wiggins (2006) evaluate a statistical model of musical pitch perception which predicts the expectation generated by monodic tonal melodies; its theoretical roots are in the work of Narmour (1990) and it is based in practice on that of Conklin and Witten (1995). The expectancy model, which is illustrated in Figure 1, is built from two memory models, one short-term memory (STM) and one long-term (LTM).

Each model takes as its *musical surface* (Jackendoff, 1987; Wiggins, 2007a) sequences of musical events (in this case, notes as written in the score, and not as performed—see Wiggins, 2007b), defined in terms of properties such as onset time, pitch, duration and key. The representation scheme can

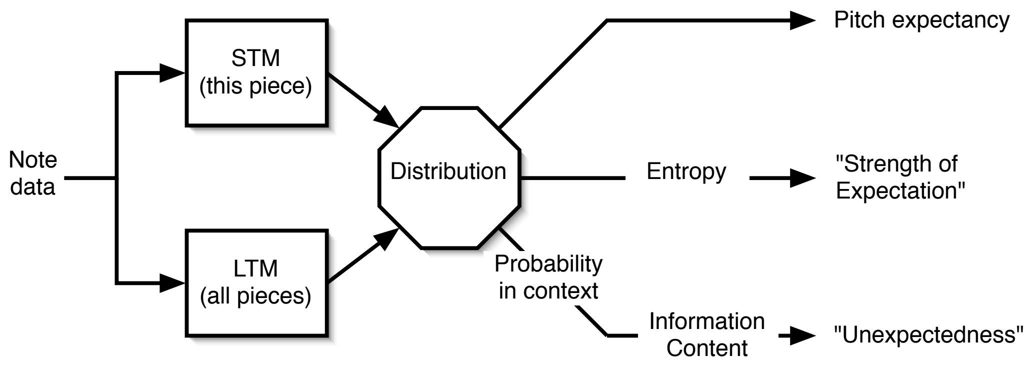

Figure 1: Our development of Pearce's (2005) cognitive model.

also express derived features (such as pitch interval) and interactions between features (denoted by $\otimes$). Here, we use three pitch based features: chromatic pitch (`Pitch`), chromatic pitch interval (`Interval`) and chromatic scale degree (`ScaleDegree`), and a feature representing the interaction of pitch interval and scale degree (`Interval` $\otimes$ `ScaleDegree`). In addition, the feature `Rest` represents a rest preceding an event and expresses interactions between the three pitch-based features described above and temporal structure in *Gradus*.

The two models are able to combine predictions derived from a number of different representations into a single distribution over some basic feature space: `Pitch` in the current research. They do not differ in their internal operation, but in their exposure to the data from which they learn. The LTM is trained, prior to listening simulation, on a database of about 900 tonal melodies (Pearce and Wiggins, 2006), and may or may not learn from the current piece of music, depending on parametric settings. The STM, conversely, has no prior knowledge, and learns dynamically, only from the current piece of music. In this way, we model "typical human Western musical experience" (LTM) and "on-going listening to unknown music" (STM). We expect, therefore, that the LTM will contain a broad model of tonal melody, derived by generalisation from its training melodies, while the STM will be very specific to whichever piece is currently being exposed to it; the current studies support these expectations. Pearce and Wiggins (2006) re-analyse behavioural data collected by Cuddy and Lunny (1995), Schellenberg (1996) and Manzara et al. (1992), demonstrating that the statistical model predicts the expectations of listeners in a variety of musical contexts significantly better than the two-factor model of Schellenberg (1997).

Armed with our statistical model of melodic expectation, we apply information-theoretic principles (Shannon, 1948) to its output. We appeal to two concepts from information theory: first, *information content*, which measures the amount of effort (literally, the number of binary digits) needed to transmit a piece of information from an originator to a receiver, *given a known, shared context*; and, second, *entropy*, which, by analogy with thermodynamic entropy, measures the amount of disorder in a signal (or, more positively, the inverse of the degree to which it is structured). Here, the shared context is a model of typical implicit musical knowledge (the LTM). The information content corresponds with unexpectedness: if something is expected and therefore not unfamiliar, it is already known in context, and therefore can be transmitted efficiently. Entropy corresponds intuitively with weakness of expectation of the outcome governed by a given distribution—that is, low entropy implies a strong expectation of what the outcome will be. Both measures are expressed in terms of the number of bits required to transmit information.

We emphasise that our model uses *no* domain-specific, programmed rules: we have explicitly modelled pitch, time and key and their mathematical derivatives as our musical surface (Wiggins, 2007a,b), but no knowledge of musical *structure* (i.e., time-sequential relationships between events on the surface) is programmed into the model. Therefore, all the model's knowledge of musical structure, and hence its expectations, are derived through statistical induction of regularites in the data with which it is trained. For this reason, we claim, this model may be an explanatory theory (in the sense of Wiggins, 2007b) underlying such descriptive theories as the Gestalt Principles (Pearce and Wiggins, 2006).

3 Methodology

The methodology we have adopted in this study is different from the usual psychological route. Rather than resorting to behavioural studies of significant numbers of human subjects, to supply evidence or otherwise of *cognitive* validity, we compare the predictions of our model with the independently produced analyses of the first author, who is an expert analyst of minimalist music (Potter, 2000), to evaluate *musical* validity. We suggest that this is an appropriate approach because the basis of our model has already been validated in the cognitive context (Pearce and Wiggins, 2006; Huron, 2006). Of course, it might be argued anyway that music analysis actually summarises the perceptions of many listeners; however, we make no such assumptions here: we are interested in the music-analytic results themselves, and in their relationships with our model's predictions. If the relationships are strong, then our method may be useful for music analysis; further, a successful explanatory theory of that analysis will help us understand the relationship between music analysis and cognition.

4 Method

4.1 Feature Selection

As discussed above, the model is capable of drawing together inferences made over multiple representations derived from the musical surface. This raises the question of choosing features that yield the richest structural representations of the works analysed here. In order to address this issue, we have selected sets of features that maximise the overall predictability (equivalent in practice to minimising the average information content) of each work as seen by the model. This was achieved by running a forward stepwise selection algorithm (Aha and Bankert, 1996; Kohavi and John, 1996; Blum and Langley, 1997) which, given an empty set of features, considers on each iteration all single feature additions and deletions from the current feature set, selecting the addition or (preferably) deletion that yields the most improvement in the performance metric, terminating when no such addition or deletion yields an improvement.

4.2 Information Dynamics

We are interested in the *Information Dynamics*—the changes in information content and entropy—of our model as it simulates musical listening. We hypothesise that (changes in) these measures will correlate with musically significant events, and we attempt to refute this hypothesis by comparing our measures with composers' perceptually relevant score annotations and our independent analyses.

As well as considering the changes in information content event-wise, *throughout* a piece of music, we can consider the average information content *across* the piece as a measure of the degree to which it is structured (perhaps what an analyst would call “rigour”). So we hypothesise that a strictly formal, systematic piece would have a lower average information content (i.e., it would be *more* structured) than an unsystematic piece. This is another means by which we attempt to refute a connection between our model’s output and the music-analytic data.

4.3 Segmentation

Musical segmentation is a common goal in music-cognitive theory and simulation (e.g., Lerdahl and Jackendoff, 1983; Cambouropoulos, 1996; Wiggins, 2007b). Our model can predict both large-scale and small-scale boundaries in music; here we discuss just the larger scale structural boundaries in detail—essentially, those which are not musically moot points—and attempt to refute the predictions of our model by comparing them with those of two music analysts (York, 1981; Potter, 2000) and the composer’s scores. However, it will become clear that many of the points identified in the detailed analysis of *Gradus*, below, correspond broadly with change-points of different kinds in the information-theoretic measures.

Our prediction of large-scale structure works by looking for relatively large, simultaneous, positive change-points in the information content and entropy of the music. This is predicated on the notion of *closure* as described by Narmour (1990) and others: as a section, phrase, or other structure draws to an end, the music becomes less unexpected (i.e., information content decreases) and predictions that can be made about it become more certain (i.e., entropy decreases). In contrast, when a new section starts, there is little expectation (about its content), so information content rises; correspondingly it becomes more difficult to predict the up-coming structure, so entropy increases as well.

When estimating the relative size of the predicted boundaries, there are three factors to consider, which we explain here in terms of the graphs (Figures 8 and 12): first, the *depth* of the “valley” in the graphs, whose minimum is at the predicted boundary, which indicates the magnitude of the change; second, the *gradient* of the graphs immediately following the change, which must be large; and, third, the *length* of the rising edge, since a long rising edge denotes a sustained (i.e., not temporary) change.

4.4 Minimalist Music

There are three reasons for choosing minimalist music for this research: one, simply, is that there is relatively little analysis of minimalist music in the literature, and so a useful contribution can be made directly by supplying some: indeed, this is the first published detailed analysis of *Gradus*; another arises from Bernard’s (1996) point about the nature of minimalist analysis, to which we refer above: if there is no deep structure in minimalist music, the question is “what makes it music?”—perhaps a cognitive modelling approach can answer this, since public response suggests that it does indeed constitute “music” for many listeners. Finally, minimalist music, due to its very nature, reduces the possible sources of musical variation, allowing for a tightly controlled study without sacrificing ecological validity by using artificial stimuli, as often happens in studies of music perception.

5 Philip Glass’ *Gradus*

5.1 Introduction

Our first case study is Philip Glass’ *Gradus* (1968). *Gradus* was written for the saxphonist Jon Gibson (and is sometimes referred to under the title *for Jon Gibson* in consequence). It is monodic and isochronous, though it does include rests, presumably partly because a saxophonist needs to breathe.

Figure 2: Incipit of *Gradus* (Glass, 1968)

Figure 3: Bars 41–42 of *Gradus*, showing the boundary change from strong scalic structure to a new configuration.

The piece, while clearly minimalist, and clearly in the style of its composer, is not, like most of his music of the late 1960s and early 1970s, systematic (Potter, 2000, page 280); an incipit excerpt is given in Figure 2. The piece is readily amenable to many extant computational techniques for analysis, which we defer for future work; instead, we focus first on a detailed tonal analysis of the piece (which, broadly, may be heard as an extended perfect cadence), by our human expert musicologist. In this analysis, we mark references to bar numbers identified as points of significance by our model with square brackets, like this: [45]. These points are then marked in the same notation in Figures 8 and 9, and we explain the figures in the subsequent subsection. We emphasise that these markings are added *post hoc* to an analysis produced *independently from the model's output*.

5.2 Analysis by Human Musician

5.2.1 Part I

The initial pitch-class gamut of *Gradus* consists of just four pitch classes—A, B, D and E—with A also repeated an octave higher. This can be interpreted as close to a pentatonic scale (only F♯ is missing). But the important thing is that this gamut of two whole tones separated by a minor third, while clearly articulating a kind of "soft" modality, is largely dependent on its articulation—via melodic outline and, in particular, via rhythm and phrasing—to achieve any recognisable pitch centre.

A is immediately prominent, being also the lowest and highest notes heard until a low G is introduced in Bar [42] (Figure 3). Yet the dominance of A is, familiarly, often interpretable by the listener as having a dominant quality relative to D. For example, the constantly repeated sub-unit, ABD (first occurring as a separate sub-unit as early as Bar 1), is readily heard as v-vi-i in, or on, D.

D, supported and emphasised by its dominant, A, is thus already perceivable as the putative focus of the work's early stages, and arguably suggestive of the music's ultimate goal. Yet the lack of any

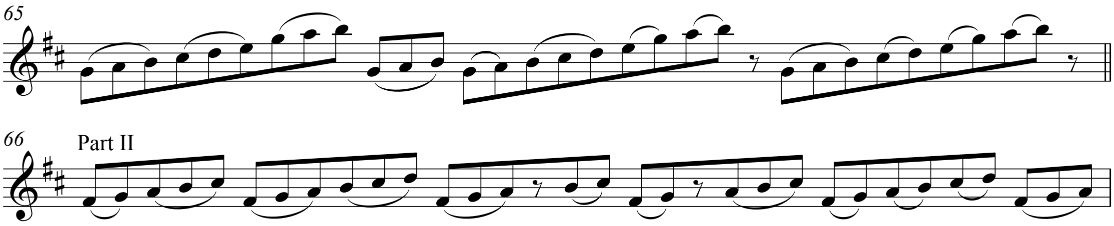

Figure 4: Bars 65–66 of *Gradus*, showing the boundary between Parts I and II.

semitones at all in the work's first twenty bars renders any such modal focus ambiguous and somewhat provisional.

The introduction of C♯, providing a (semitonal) leading note to the tonic of D, only occurs, as suggested above, in Bar [21]. Before that, the original four-pitch gamut is expanded to five pitches by the addition of G in Bar [4]. This new pitch is readily incorporated into the "early stages" of attempting to identify *Gradus*'s material, pitch operation and process: modally, G is, of course, iv of D, thus enhancing both the dominant quality of A and the tonicisation of E.

It is worth drawing attention here to the general matter of how the unfolding articulation of pitches and rhythms in *Gradus* contributes to the listener's sense of modal or tonal focus. To give just a single example: once the pitch G is added, it is used to articulate a potential "dominant 7th" quality to the modal mix around D. A good example would be in Bars [6 and 7], in which can be noted the gradual detachment of G from its initial function as merely part of the rising sequence from A to A, towards a first exploration of the role of G in arpeggiated sequences emphasising A, E and G. If it is D, and not yet C♯, that here helps fill out these further forays into the area of arpeggiation as opposed to scale patterns (the suggestion of arpeggiation had arguably begun as early as Bar 2), then that only draws attention to the subtlety with which the move towards greater modal/tonal clarity is achieved in *Gradus*. The contractions to just a few pitches (for example, the abandonment of G, and high A, in Bars [9 and 10], to focus on the lower notes in the gamut, occasionally right down to AB alone) should also, however, be noted.

C♯ is eventually reached, as indicated above, in Bar [21]. Its potential leading-note status is immediately emphasised by the incorporation of C♯ into rising patterns that emerge as a natural consequence not only of the general trend but also of the preceding Bar [20]. As with AB earlier on, C♯Ds quickly detached from time to time from the broader upward movement surrounding it. And the feeling of a first-inversion dominant-seventh chord on A is already quite strong in Bars [21–23].

Since the basic building blocks of *Gradus* are now firmly in place, and no further new pitch class is introduced until the F♯ which announces the arrival of Part II of the work (Bar [66]), let us make a few observations on two matters which seem significant here.

Firstly, the gradual expansion of pitch register. With six pitch classes (ABC♯DEG) now in play, attention focuses on the way in which registral expansion builds on the gamut, C♯–A, the focus of Bars [21 and 22]: first re-establishing the full octave, A–A, then expanding outwards from this in both directions. There is in fact quite soon an expansion downwards to B (in Bar 23), then to the original low A (Bar [26]); though one short passage (the second half of Bar [33] and the whole of Bar 34) reverts to B. The original A–A register is emphasised by the way in which, for the first time in the piece, more than a single quaver rest is inserted within an individual pattern.

Further ways in which rhythm helps articulate and mould the listener's approach to pitch focus include the tendency to introduce more than one rest in a pattern. This begins to manifest itself in Bar

Figure 5: Bars 34–35 of *Gradus*, showing a change from scalic to more fragmented structure.

[28]; one shouldn't, though, make too much of it here, on account of the fact that it has been difficult from the piece's outset to distinguish clearly between patterns as indicated in the notation.

Yet the fragmentation at which it hints is then taken much further from Bars [35-37], where rests intervene further to break up the line (Figure 5). Though longer sequences soon disrupt this tendency, the effect is clear. The dominance of simple rising patterns, without descending detours, from near the start of Bar [31], is further emphasised by such fragmentation. So is their assertion of A7, itself also increased by the way in these rising patterns flip back from high G to low A. Such deployment of rests can also, perhaps paradoxically, help to make the patterns sound at all regular. When they do this, as for instance in Bars 35–37, they actually make the flow of the music easier to digest and recall.

The first move outside the A–A frame occurs with the descent to low G at the start of Bar [42]. The registral frame now shrinks for a while to G–E, and the consequence is to increase the tension but, if anything, to assert A7 still more firmly. Yet the anticipated descent from iv to iii in D is delayed all the way to the beginning of Part II. Before this, upward registral expansion occur (progressively via E, G and A to the new note B at Bar 60), as well as several examples of contraction supplying at least a local level of tension. In the course of this, A7 achieves a rare moment of unadultered focus via fourteen consecutive notes in Bar [47].

With the return of high A, in particular (most prominently from Bar [50]), high Gs and As become audibly detachable from the melodic line, as they are subject to a modest degree of repetition. More obvious, partly due to being more prolonged, is the contraction of register from Bar [53] onwards, continuing the suggestion of iv–i in D.

The registral expansion upwards is swift, however: firstly to E (end of Bar [53]), then all the way to A (towards the end of Bar [56]). In Bar [58], expansion downwards to G once more sees a low G to high A range with the now characteristic GA emphasis on top, with a brief shrinkage down to G–D at the end of Bar [59], before the next expansion upwards, to B. With GAB sequences now prominent at the top of the registral span of a major tenth, the descent to F♯ in Part II is prepared with something of a flourish.

5.2.2 Part II

Part II is close to just one third the length of the whole piece. It continues the basic strategies already described for Part I, extending the pitch range further to achieve a total span of a perfect twelfth (low F♯–high C♯); but also introducing a new strategy, that of consecutive repeated notes on the same pitch, which will ultimately unravel the previous scalic-based strategies and conclude the piece. Further advantage is also taken of the ambiguities inherent in the pitch gamut deployed in *Gradus* to tease the listener with the possibility of a tonal resolution that is only, ultimately, half-delivered: via its dominant, the favoured modal strategy of *Gradus* as a whole.

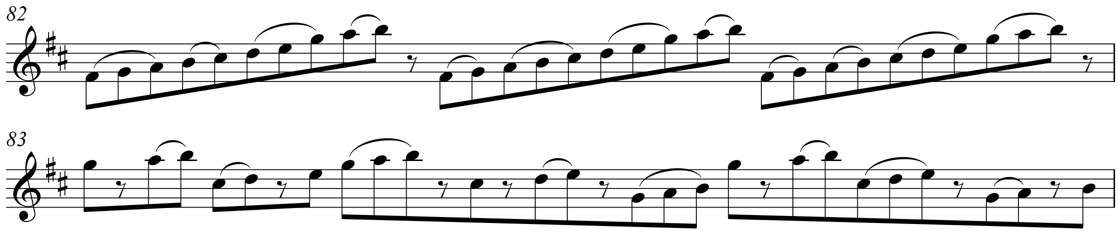

Figure 6: Bars 82–83 of *Gradus*, showing the third large boundary change from scalic to more fragmented structure.

Having accomplished the descent to low F♯, Part II initially contracts to spans of no more than a major sixth for any individual pattern, with the highest note for Bars [66]–67 being E, and for Bars [68]–[70], D. Perfect fifths and perfect fourths are, indeed, much more common than sixths, and soon, at Bar [68], the span shrinks to a perfect fourth (F♯–B). The effect of these shorter scalic patterns, contrary to the effect achieved by the gapped pentatonic scale of the work's opening, is to focus more clearly on particular pitch centres, especially the D major that had been the more implicit modality of Part I.

But Part II indulges in no simple and explicit insistence on, or final celebration of, D major as the tonic key. Or at least not without teasing the listener along the way. When the focus shifts to G, for example, the largely triadic Bars [69-75] prolong this territory for a length of time that is surprising and disruptive in this context. The high A that is inserted into the G♯7 iterations (G major plus its leading note) from the end of Bar 74 [74-75] only lead to a return of scalic patterns at Bar [76], where it is the reoccurrence of C♯ that causes the move back to an D major emphasis. The return of high B in Bar [78] not only continues the gradual extension upwards of the scalic patterns brought back with the previous rearrival of A, but a progression of the highest note in Part II from C♯ and D (Bar [66]), E (Bar 67), and, in particular, G (Bar [71], via the G major triads, not the scalic patterns [71-74]). And while D major is rearticulated with abundant clarity at the start of Bar [83], following the increasing (and actually quite systematic) unravelling of low F♯ to either high A or high B patterns over the course of Bars [78-82] (Figure 6), a G-major flavour insinuates itself almost immediately (the GAB sequences in Bars [83-85] are offered in both the lower and high octaves, playing iv–v–vi in D major but appearing sufficiently often, and in sufficiently fragmented form, to draw attention to themselves in their own right, as it were).

With the final extension to high C♯ from Bar [87], the range now reaches its maximum size, a F♯-C♯ perfect twelfth. Its full articulation proves sufficiently long to require the spilling out of the 32-quaver bar lengths over into successive bars, as it expands. (Such "spillage" had already occurred as early as Bar [69], in Part I.) The more continuous quaver movement combines with the fact that almost all patterns not commencing on low F♯ start on C♯ to reinforce F♯ and A as possible alternative focal pitches to D, even as F♯ and A could be argued to reinforce its centrality triadically.

Before the arrival of repeated notes on the same pitch, D major makes a clear return to unambiguous focus (in Bar [92]), via a clutch of leading-note/dominant [seventh] emphases homing in progressively on the tonic. High G and A are the only two pitches to "silt up" through immediate repetition (Bar [94]; Figure 7); the context in which this is initially done instantly suggests iv–v in D major. The gradual process whereby such repetition spreads out, like a virus, eventually to take over the final five bars of *Gradus* entirely, is an interesting one which should be explored on rhythmic and systematic levels as well as the tonal/modal one. Suffice it to say here that the earlier tendency

Figure 7: Bars 93–95 of *Gradus*, showing the full range of the piece, and the boundary (in the middle of Bar [94]) to the "coda".

towards low C♯ as the main alternative to low F♯ as the starting point for the rising sequences is now embraced enthusiastically and comprehensively from Bar 90 onwards.

5.2.3 Conclusions of the Tonal Analysis

The overall progression of *Gradus* can be described as a move from the initial four pitch classes, ABDE, quickly to five pitch classes (+ G, from Bar [4]), and then, at a more leisurely pace, to six (+ C♯, initially from Bar [21]), and finally seven (+ F♯, from the beginning of Part II at Bar [66]). This moves from an ambiguous, basically pentatonic, scale, ABDEG, to a seven-note diatonic sequence, F♯GABC♯DE, that most readily, in actual articulation, suggests a D major scale.

This progression is, however, complicated by a number of factors in addition to the changing extent of registral repetition of pitch classes. The most important is the way in which contraction as well as expansion of both the number of pitch classes in play and the overall number of different notes, permit local explorations of different parts of the total gamut. For example, C♯ first appears in Bar [21] as part of a five-pitch gamut rather different from that of the piece's opening (C♯DEGA), and only from Bar [23] as part of the six-pitch gamut, BC♯DEGA.

It is partly via this means that Glass plays with the listener's expectations as far as pitch centres are concerned. Such toying with the listener's attempts to make sense of *Gradus*pitch progressions also includes the tantalising deployment of the repeated notes by which its final pages are increasingly dominated.

Resolving firmly onto D itself would arguably have been crass in a modal context such as the one *Gradus* creates. Yet the expansion of the pitch range upwards to C♯ and downwards to F♯ implies a final resolution on D that is never in fact delivered. The only two pitches selected for repeated treatment, G and A, offer iv and v in D major, but not the tonic of D itself. Coming as it does at the work's end, this gambit is perhaps the most telling as well as most extended example of Glass' rather subtle engagement of the interface between a modality sufficiently ambiguous to allow constantly shifting pitch perspectives, and one rooted in a familiar diatonic scale sufficiently inflected by repetition on a variety of levels to insinuate the more focused perspective suggested by invoking the term tonality.

5.3 Information Dynamics of *Gradus*

Now we discuss the information dynamics of our model, as it simulates listening to *Gradus* in the context of the above analysis. First, we consider the outputs of our Short-Term Memory (STM) model, shown in Figure 8. Using the STM alone, the best-performing feature set (see "Feature Selection", above) contained `Interval` ⊗ `Rest`, `Pitch`, `Interval`, `Pitch` ⊗ `Rest` and `Interval` ⊗ `ScaleDegree`, yielding a mean information content of 1.56 bits per event. Recall that this model is untrained, and therefore contains only knowledge of local structure in the piece, learned dynamically as the music proceeds. This knowledge is expressed in terms of the selected features. Selection of these viewpoints suggests that tonal function (which intervals, and where in the scale) is important as well as the articulation implied by the presence of the rests in the piece.

The figure shows both our measures: the solid line is entropy; the dotted line is information content. The vertical rules drawn at regular intervals along the graph denote "bar" lines in the score; the x-axis is time expressed in quavers.

To read the graph, it is important to understand that both measures are smoothed with a sliding window, 64 events (2 bars) wide—otherwise the local jaggedness of the graph would make the effects we are studying very hard to see. Therefore, a sharp change in gradient of the graph denotes a *step* change in the local mean value of the data, and we focus on the temporal location of these change points: we wish to explore correlation between them and our independent analysis.

Globally, the STM information content is generally lower than the entropy, which means that the model is doing relatively well at predicting what comes next at each stage. This gives us a measure of the internal coherence of the piece: there is some coherence, but nothing like the systematic coherence of *Two Pages* (see below).

The LTM model, shown in Figure 9, on the other hand, does relatively poorly at predicting: 3.34 bits per event, and information content is consistently higher than entropy. Automatic feature selection for the LTM yields poorer prediction performance than the STM using a best-performing feature set consisting of `ScaleDegree` ⊗ `Rest`, `Interval` ⊗ `ScaleDegree`, `ScaleDegree` and `Pitch` ⊗ `Rest`.

Apart from the features in common between the two optimal feature sets, the set selected for the STM exhibits a greater emphasis on local melodic structure (`Interval`, `Interval` ⊗ `Rest`), while that selected for the LTM shows a tendency towards tonal-melodic structure (`ScaleDegree`, `ScaleDegree` ⊗ `Rest`).

Notwithstanding the relatively poor prediction, there are still interesting change-points in the LTM model, and these are marked, in the same way as before, in Figure 9. A particular area of interest here is the relatively low information content in Bars 33–42 and Bars 57–68. The first of these is a combination of two effects: in [35-38] the frequency of rests increases dramatically, so the rather non-melodic, arpeggiated quality of the piece breaks down; in the bars before and after, on the other hand, there is strongly scalic writing. The second, 57–68 is an area of purely scalic writing, which gives rise to much higher predictability, though still being not conventionally melodic. Finally, note how the very extended scalic sections, [42-46] and [59-66], dramatically decrease the information content of the notes: we argue below that contributes to the feeling of musical closure at these points.

Between the two graphs, it is possible to isolate change points which correspond with all of the significant points in our expert analysis above, and it is appropriate at this point to recall that this analysis was produced independently from the graphs. The fact that not all of the points of interest are visible in the output of both models underlines the need for a *dynamic* system here: not only do the feature selectors of each model need to be dynamic (as they are) but also the combination of the two models, if the system is to help in producing analyses such as this.

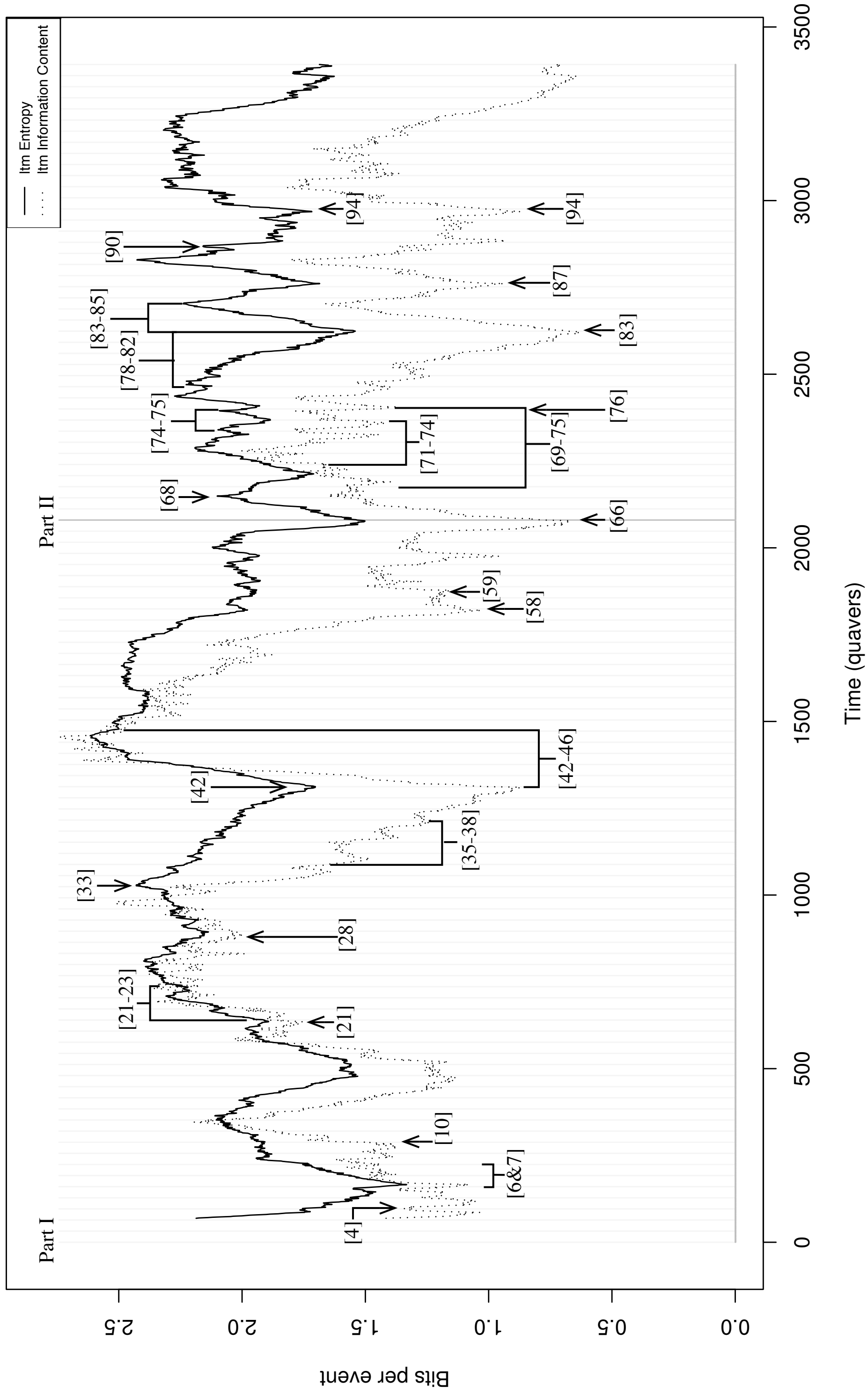

Figure 8: The STM Event-wise Information Content and Prediction Entropy of *Gradus*. Vertical lines mark boundaries between bars.

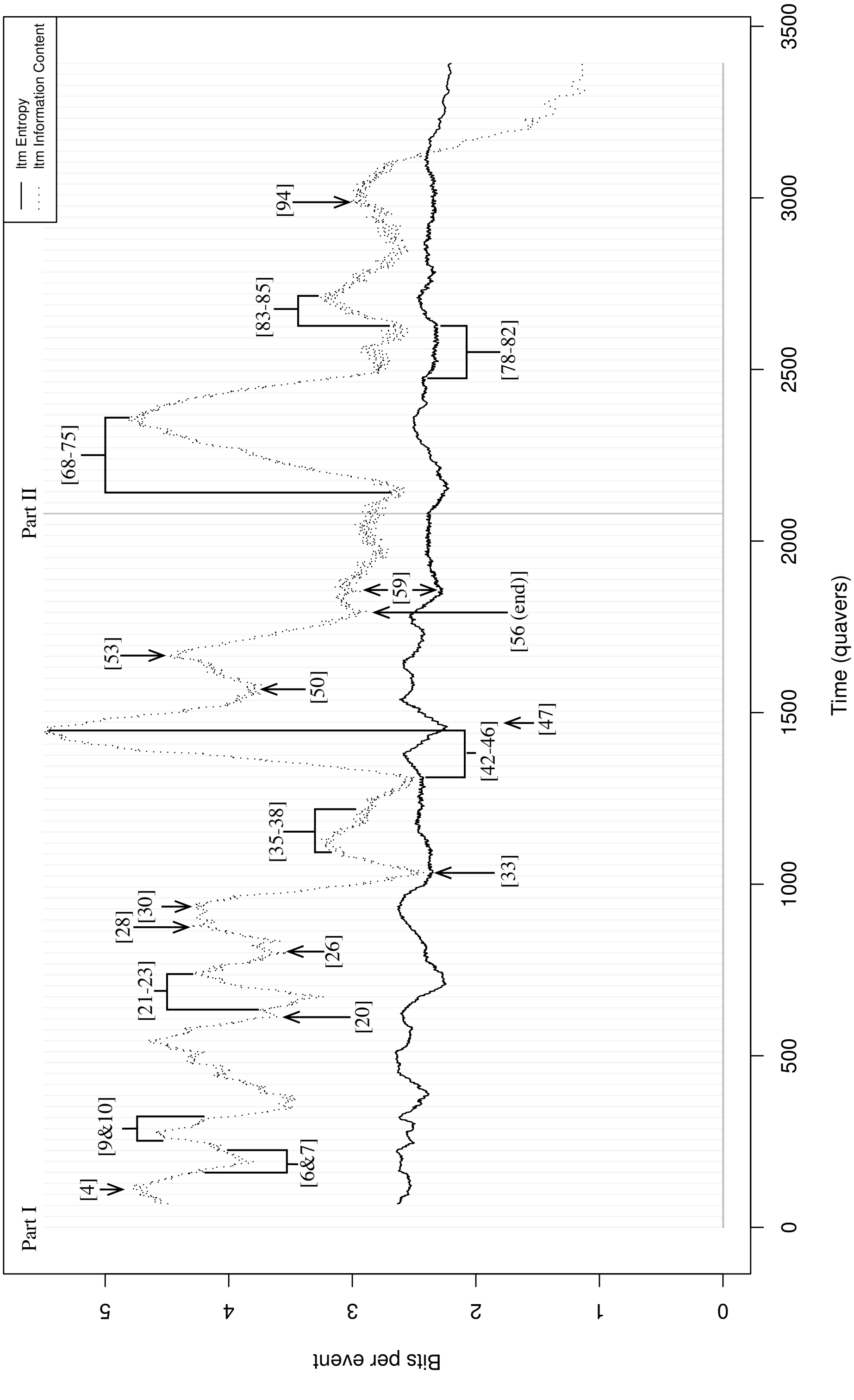

Figure 9: The LTM Event-wise Information Content and Prediction Entropy of *Gradus*. Vertical lines mark boundaries between bars.

Nevertheless, we suggest that the outputs presented here demonstrate that our model has clear potential not only to be used as a musicological tool but also to serve as a concrete, quantifiable link between musicological analysis and music cognition. We present further evidence for this claim below.

5.4 Segmentation

In the score of *Gradus*, Glass marks only one boundary: that between Part I and Part II. However, there are, in fact, three clear, large-scale boundaries in the piece, arising in the music from the division of the piece into four sections, each of which is a transformation of more arpeggiated passages like that at the start of the piece (Figure 2) into extended scalic passages (Figure 3), the last having a sort of coda of repeated notes.

The largest such boundary is predicted at Bar [42], which is the first time this kind of change has been encountered (Figures 3 and 8). Note that both entropy and particularly information content have decreased significantly over the preceding bars, and that the increase in both happens precisely on the first event of Bar 42.

The second large boundary is predicted at Bar [66], also marked by the composer as the start of Part II. It is marked in Figure 8 and 4. This boundary is structurally very similar to that at Bar [42]; for this reason, i.e., because many of the structures are shared, the relative changes associated with closure and the start of a new section are less pronounced, though they are still uncontrovertible in the graph.

The third large boundary is predicted at Bar [83], which point is again between some of the scalic passages to which *Gradus*' sections tend and new, more arpeggiated material (Figure 6).

We must now explain, therefore, why are we not predicting large boundaries at, for example, Bars [58], [87] or [94]. At Bar [58], the "valley" in information content is not matched by one in entropy. Therefore, closure as we model it above has not been achieved, so this is not a candidate boundary. At Bars [87] and [94], the "valleys" are quantifiably smaller than at our proposed boundary points, but also the gradients of the graphs to the right of the minima are substantially (about a factor of 2) smaller, and, by a similar factor, less sustained.

As a final piece of evidence for our "closure" interpretation, note how both measures significantly decrease towards the end of the piece.

We suggest that this discussion demonstrates that our model is capable of finding large-scale boundary points in *Gradus*. However, this might still be an *ad hoc* result, arising from relationships between our model and the structure of this particular piece. Next, we generalise the result to another work, whose structure is musically quite different, to refute this possibility.

6 Philip Glass' *Two Pages*

6.1 Introduction

Two Pages, written only a year after *Gradus*, is a very different piece of music. An incipit excerpt is given in Figure 10. The two pieces share significant surface properties—they are both isochronous, monodic and clearly minimalist—and deeper musical properties—in particular, subtle tonal/modal manipulation (Potter, 2000); but *Two Pages* is strictly systematic. Its system is explicit in the score, either as repetitions of given figures (see Figure 10), or as specifications of figures which change in time (see Figure 11).

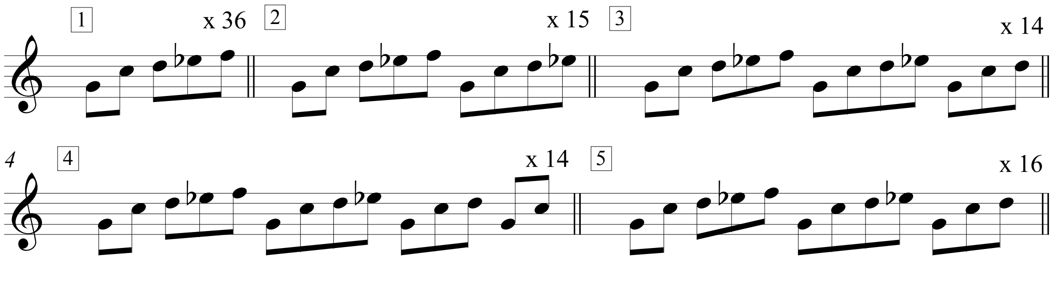

Figure 10: Incipit of *Two Pages* (Glass, 1969)

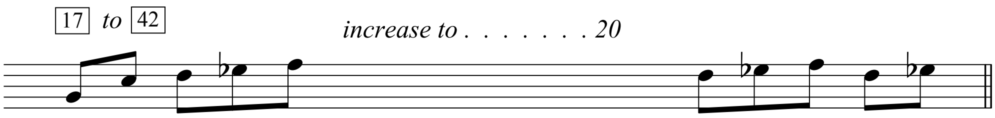

Figure 11: The specification of a changing figure from *Two Pages* (Glass, 1969)

Although the scalic sequences of *Two Pages* do convey tonal implications, the behaviour of our LTM model in the piece is much less interesting that that of the STM model, because *Two Pages* is even less traditionally tonally-melodic than *Gradus*. We focus here on the STM only. For evaluation, we need a structural standard with which to compare, and we have three: one by the composer, in the form of (musical) figure and section boundaries and others by York (1981) and Potter (2000). We now consider the information dynamics of the combined learning model as it simulates listening to *Two Pages* in terms of these structural boundaries and of the (musical) figures in the piece.

6.2 Information Dynamics

The best-performing feature set for the STM in *Two Pages* contained only the feature `Interval` $\otimes$ `ScaleDegree`, indicating that tonal function is very important in the piece. The selection of just one feature, compared with five in the *Gradus* study, is probably a result of Glass's insistence on exploring the additive processes in this composition using an unrelentingly minimal structure set. The mean information content obtained using this feature, 0.36 bits per event, is much lower than that obtained for *Gradus*, reflecting both the more constrained materials (5 distinct pitches and 10 distinct intervals as opposed to 11 and 32 respectively in *Gradus*) and the more strictly repetitive structure of the piece. So, as we predict above, in these pieces, mean information content does decrease with increasingly systematic composition.

Figure 12 shows event-wise information content and entropy of the STM while simulating listening to *Two Pages*; Figure 13 shows the corresponding LTM output. As before, the graph is smoothed with a sliding window to assist visualisation; the same artefacts arise and are significant here: a sharp change in graph gradient denotes a step change in local data, and we are interested in the times of these change points.

The vertical lines, as before, denote the boundaries between the musical figures of which the piece is constructed. In *Gradus*, these were all of the same length (32 quavers); in *Two Pages*, they are of variable lengths, so in the graph the x-distance between the lines varies. The bolder vertical lines denote the Part boundaries as annotated in the score by the composer; their numbers are shown on the

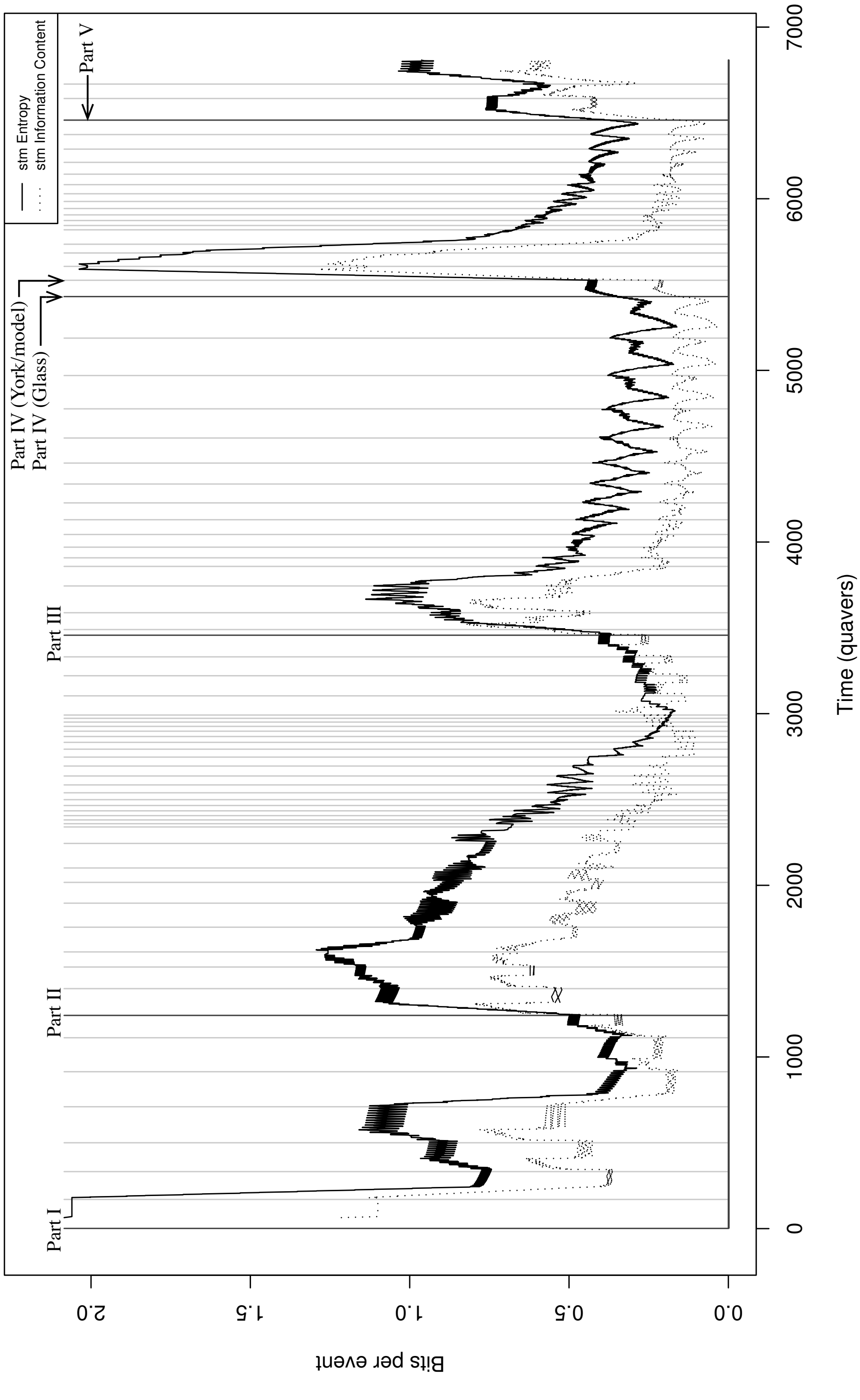

Figure 12: The Event-wise STM Information Content and Prediction Entropy of *Two Pages*. Vertical lines mark boundaries between musical figures; bold vertical lines mark Part boundaries as annotated by composer.

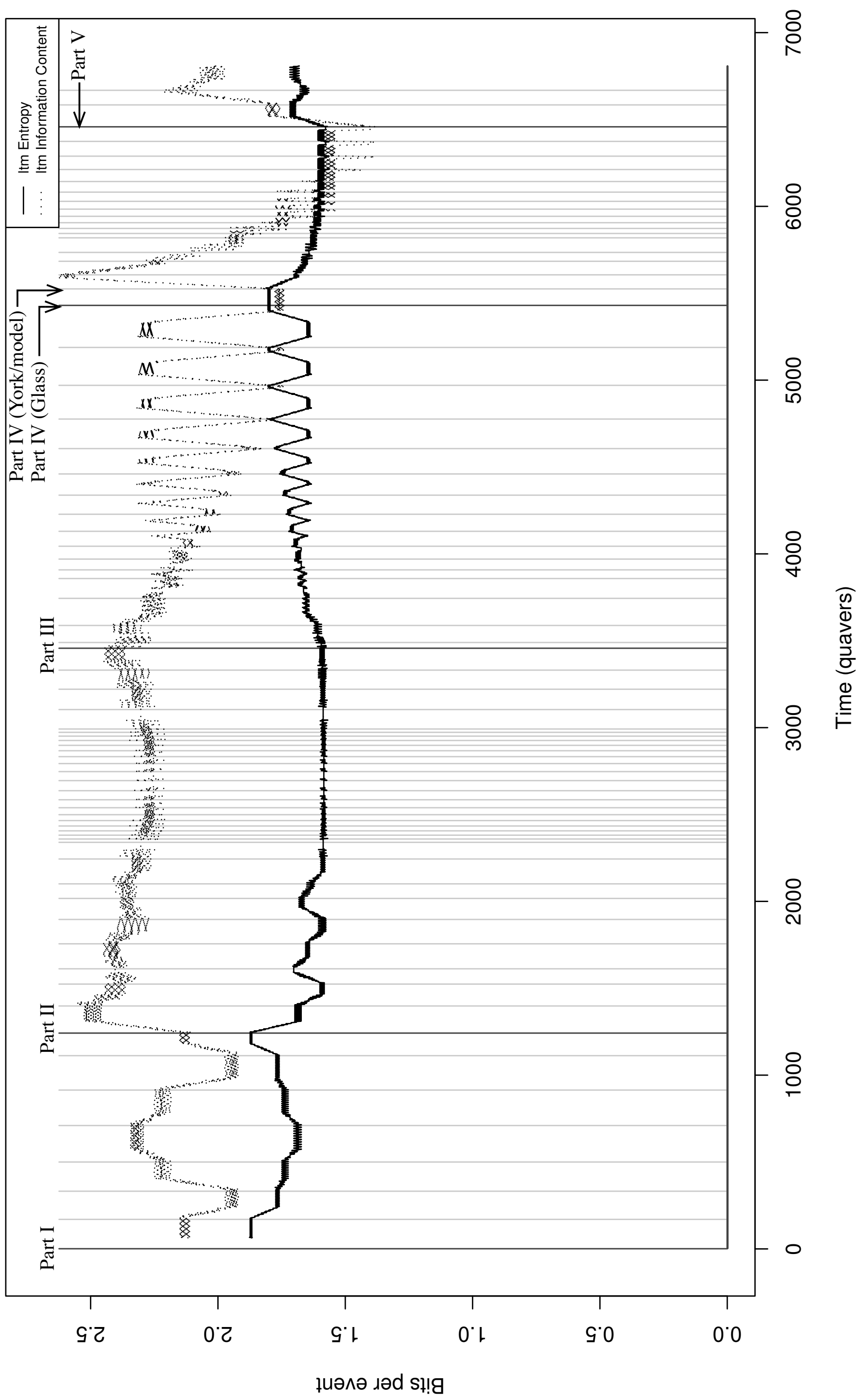

Figure 13: The Event-wise LTM Information Content and Prediction Entropy of *Two Pages*. Vertical lines mark boundaries between musical figures; bold vertical lines mark Part boundaries as annotated by composer.

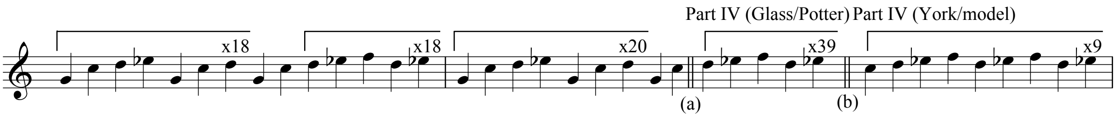

Figure 14: The transition to Part IV of *Two Pages*.

graph. The smoothing window is 64 quavers wide, as in the *Gradus* graphs.

Globally, we immediately see that the vast majority of figure boundaries coincide with a change in direction of both of our measures. Because of the nature of the music, we would expect the measures trivially to coincide with change, but not necessarily a large change each time; so there seems to be a suggestion that something information-theoretically (and therefore musically) stronger than just arbitrary addition of symbols is happening here. But what is more striking from the graph, and rather harder to explain in musical terms, is the very large increase in both measures at the start of each of the composer's marked Parts (with the exception of Part 4, where there is a delay of one figure; we return to this below). These changes are hard to explain because the changes in the music that produce them are apparently very small, and not apparently very different in kind from the changes that take place at the other weaker boundaries in the piece; in both information-theoretic and music-analytic terms, however, they are very significant. This suggests that our model is accurately predicting (using exactly the same rules as for *Gradus*) the segmentation proposed by both the composer and two analysts for four out of five part boundaries (I–III, V), and suggests interesting future detailed work on the nature of these changes and what musical effects they have.

The Part 4 boundary requires further consideration, because the large detected change, which, we claim, marks this Part boundary is apparently one figure late, according to the composer's marking. To understand why this is, we need to resort to the score fragment shown in Figure 14.

The figure shows the alternatives for the start of Part IV, marked (a) and (b). In Potter's (2000) analysis, which was made from the manuscript score of the composer, the Part IV boundary at (a) is assumed to be *a priori* given; indeed, in that score, Glass goes into uncharacteristic detail about the compositional method, marking some of the notes as a "tag", which seems to mean the marker at the end of a repeated sequence. However, comparison with the preceding (musical) figure, shown in Figure 14, demonstrates that, notwithstanding the means of construction, point (a) is in fact in the middle of another figure which is structurally congruent—the only difference is the number of repetitions. Furthermore, point (b) marks the beginning of a (musical) figure which is *substantively* different from the two preceding ones. This, we argue, is the *perceptual* Part IV, rather than the annotated compositional one at point (a).

Support for our claim comes from York's (1981) analysis of *Two Pages* which was made from the analyst's own transcription of an audio recording, and without sight of Glass' manuscript score. York analyses the piece in the same way as our system, placing the Part boundary at point (b) in Figure 14, and viewing the two (musical) figures immediately before point (b) as different parameterisations of the same structure. In this way, we suggest, our system (along with York) has more accurately captured the music *as heard* than the alternative analysis based on (or perhaps obscured by) the inaudible compositional method; we might even suggest, on the basis of Bernard's (1996) description of the minimalist aesthetic (see "Introduction"), that its section marking is more "correct" than that of the composer.

7 Summary and Conclusions

We have presented evidence for a relationship between two objective measures of the information dynamics of musical works, as heard by a simulated listener, and points of varying structural importance in the music as analysed by expert human analysts. We have identified various ways in which such a relationship might be refuted, and failed to do so.

We have given a novel, detailed analysis of one piece, and shown how it corresponds with the information dynamics of the music as heard, and we have shown that this connection generalises, at least to music in a similar style by the same composer. We will attempt to demonstrate further generalisation elsewhere.

In taking this approach, we have demonstrated a promising methodology for the computational and cognitive study of music: rather than taking musical stimuli and artificially simplifying them to make their study tractable, we have sought and found music which is amenable to the kind of treatment of which our system is capable, thus enabling a more ecologically valid approach, and, therefore, in principle, more ecologically valid results. However, more work is needed before this latter claim can be justified in practice.

At a more general level, we suggest that this research provides further evidence that dynamic information-theoretic modelling of musical listening is a worthwhile approach to the study of music cognition and also has the potential to be a powerful tool to increase objectivity in data-based music analysis. Our immediate future work will seek to refine the current models with further data, to extend our coverage to more minimalist and other music, and to study the relationship between the predictions of our model and human brain function.

Acknowledgement

The third author's involvement in this work was funded by EPSRC research grant GR/S82220 "Information Dynamics of Music".

References

Aha, D. W. & Bankert, R. L. (1996). A comparative evaluation of sequential feature selection algorithms. In Fisher, D. & Lenz, H. J. (eds), *Learning from Data: AI and Statistics V*, pages 199–206. New York: Springer.

Bernard, J. W. (1996). Theory, analysis, and the "problem" of minimal music. In Marvin, E. & Hermann, R. (eds), *Concert Music, Rock and Jazz Since 1945*, pages 259–284. Rochester, NY: University of Rochester Press.

Blum, A. & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97(1–2)*, 245–271.

Cambouropoulos, E. (1996). A formal theory for the discovery of local boundaries in a melodic surface. In *Proceedings of the III Journées d'Informatique Musicale*, Caen, France.

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

Cuddy, L. L. & Lunny, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgements of continuity. *Perception and Psychophysics*, *57*, 451–62.

Glass, P. (1968). *Gradus*. New York: Dunvagen Music.

Glass, P. (1969). *Two Pages*. New York: Dunvagen Music.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Bradford Books. Cambridge, MA: MIT Press.

Jackendoff, R. (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.

Kohavi, R. & John, G. H. (1996). Wrappers for feature subset selection. *Artificial Intelligence*, *97(1-2)*, 273–324.

Kostelantz, R. (ed.) (1997). *Writings on Glass*. Berkeley, CA: University of California Press.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA.: The MIT Press.

Manzara, L. C., Witten, I. H., & James, M. (1992). On the entropy of music: An experiment with bach chorale melodies. *Leonardo*, *2*, 81–8.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

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

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

Potter, K. (2000). *Four Musical Minimalists: La Monte Young, Terry Riley. Steve Reich, Philip Glass*. Cambridge, UK: Cambridge University Press.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realisation model. *Cognition*, *58(1)*, 75125.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy. *Music Perception*, *14(3)*, 295–318.

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

Wiggins, G. A. (2007a). Computer-representation of music in the research environment. In Crawford, T. T. & Gibson, L. (eds), *AHRC ICT Network Music Expert Seminar*. Oxford: Ashworth.

Wiggins, G. A. (2007b). Models of musical similarity. *Musicae Scientiae, Discussion Forum 4a*.

York, W. (1981). Form and process in *Two Pages* of Philip Glass. *Sonus*, *1(2)*, 28–50. Reprinted in Kostelantz (1997, pp. 60–79).