Models of musical similarity

GERAINT A. WIGGINS
Centre for Cognition, Computation and Culture
Goldsmiths College, University of London

• **Abstract**

I attempt to align and compare the various papers in this Discussion Forum and to draw some general conclusions from them. Because the range of papers is so broad, it is not possible to compare them in detail, and so the comparison is made at the meta-level, comparing the nature of the models and techniques proposed, and the results produced, and discussing how these important papers combine to contribute to our understanding of music cognition. In conclusion, I propose a viewpoint based on memory and learning, to which, I claim, all the work ultimately points.

1. Introduction: How do I discuss these papers?

As Discussant to the ICMPC 2006 Symposia on Musical Similarity, I was faced with an impossible task, which is again asked of me here: to produce a coherent summary of the hugely broad range of work and ideas in this volume is quite unfeasible. There is, represented here, a rich diversity of work based around issues concerned with similarity and categorisation of music, with implications far beyond this relatively restricted application: on knowledge representation (Lemström and Pienimäki; Typke et al.), rule-based modelling — in the sense of Lerdahl and Jackendoff (1983) — (Lartillot and Toiviainen; Müllensiefen and Frieler; Ahlbäck), feature-based modelling — in the sense used in signal processing — (Eitan and Granot; Typke et al.; Eerola and Bregman; Müllensiefen and Frieler), higher-level "model frameworks" — at least by implication — (Deliège; Eitan and Granot; Müllensiefen and Frieler; Selfridge-Field) and musical memory and context — at least by implication — (Deliège; Eitan and Granot; Eerola and Bregman; Müllensiefen and Frieler; Ziv and Eitan; Selfridge-Field). In any case, the authors have presented their work in their own papers, and so there would be little point in rehearsing it again here. Instead, I propose to give a commentary on how we might bring all of these ideas together in a unified model of music perception. In doing so, I will include: my own interpretation of the implications of the work (as opposed to its specific results);

references to some of the efforts of my research group in this direction; some untestable assumptions about the (distant) past; and some speculation about the (not so distant) future.

In the rest of this paper, I will first set out my own stall of starting positions, which will inform my subsequent discussion of the relationships between the papers in this volume. I will then highlight what seem to me to be key issues in this area, with a view to drawing together the big picture behind the papers, relating these papers to them. Then, finally, I will propose a general position from which to view the papers, which underlines the importance of memory and learning in music cognition.

2. An Evolutionary Perspective

Before embarking on this discussion, it is useful to explain the underlying reasoning of all the thinking here presented, which is derived from a strongly evolutionary perspective. Following Plotkin (1998), it is my belief that it is not possible to give a general theory of cognition without considering how it arose through evolution, and this applies equally to more specific theories about aspects of cognition, such as musical behaviour. This stance will inform my proposal, at the end of what follows.

Pinker (1994) famously dismisses music as "auditory cheesecake", which, he argues, is enjoyable for reasons which are incidental and coincidental, but, in evolutionary terms, is not "good for us". Bown and Wiggins (2005) and Bown (2006) argue, however, that music is of real importance in *social* evolution, and the fact that many of the faculties which enable music audition evolved for other reasons is not an argument against music's importance in the development of humankind; they propose simulations which can supply supporting evidence for the validity of such arguments (though, of course, no amount of simulation can demonstrate what actually happened).

In social evolutionary terms, then, either musical behaviour was so strongly innate in the very first *homo sapiens* that it has not been bred out of any of their descendent societies over a time span of 50,000-100,000 years, or it is a strong selectional force in social evolution: witness the fact that musical behaviour (of some kind) is universal throughout human societies (possibly because musical behaviour evolved independently in all societies, or possibly because it evolved before those societies became distinct). In any of these cases, it seems extremely unlikely that such a complex and ubiquitous construct would evolve by chance, even on the back of such a clearly selectional force as language (even if we could demonstrate that music is a *post*-cursor of language, a claim against which there is plenty of argument — e.g., Trevarthen, 1999; Dissanayake, 2001).

Given that general musicality seems to be a universal human trait, we cannot avoid the questions of why different musics exist, and why, at least before electrical

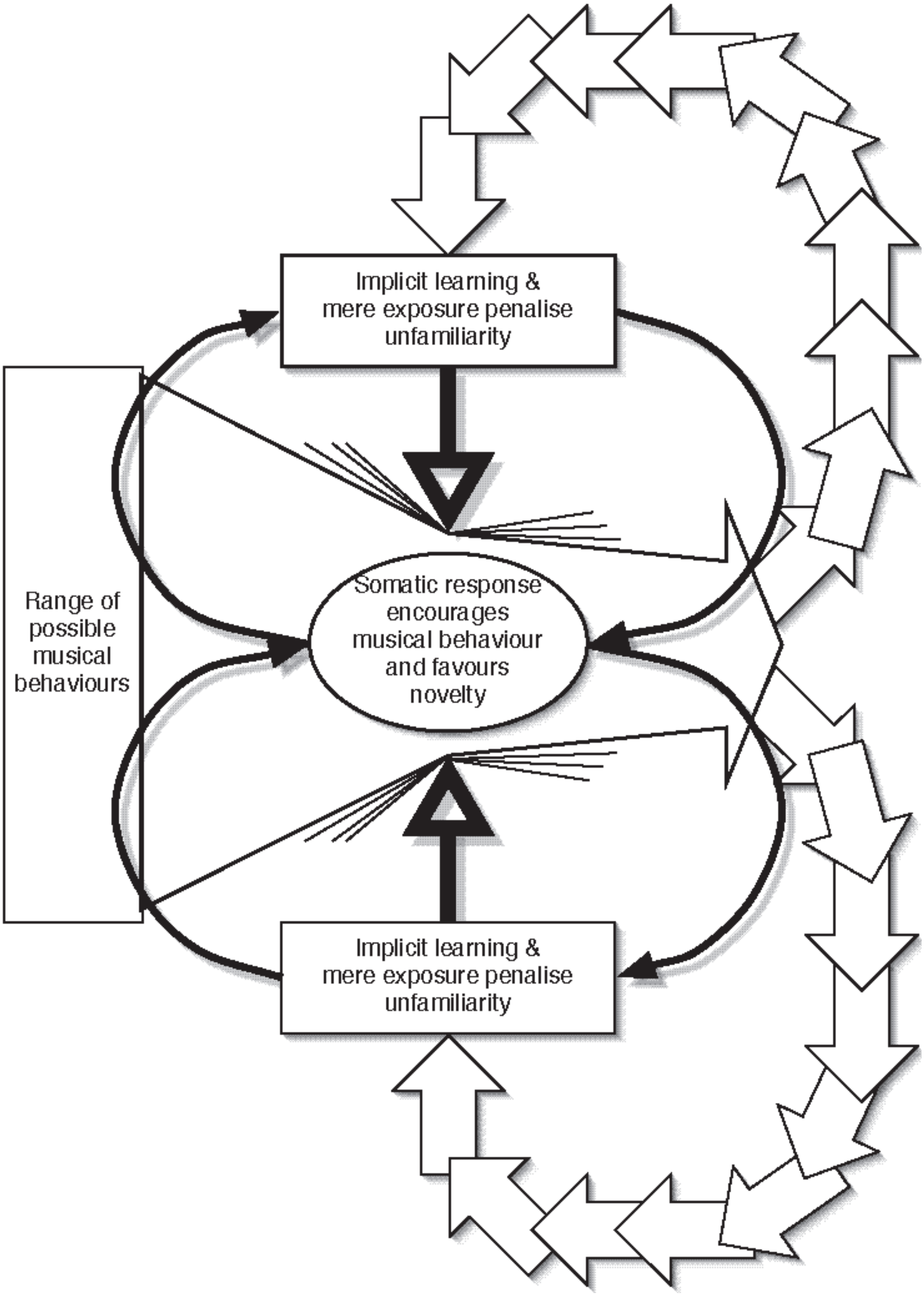

Figure 1.
Positive Feedback constrains musical language.

and electronic media, they were largely confined by societal boundaries. There is no evidence to suggest that this would arise from genetic differences between human societies. Therefore, the development must be in the musical behaviour itself, and in its expression by humans in society, driven, of course, by perception and cognition. One compelling account of this can be given in terms of *positive feedback*, based on the mere exposure effect (Zajonc, 1968), which implies that memory is crucial, and also implicit learning. The mere exposure effect acts like a resonant filter, allowing familiarity to govern acceptability, and, like the resonant filter in a radio receiver, to allow the energy of a signal (in this case the expressive language of any particular society's music) to be focused into a particular area of its space of potential values (frequency bands in the radio; specific musical languages in evolution). Biederman and Vessel (2006) propose an account of somatic response to human learning, which, coupled with the associative nature of human memory, may explain why there is a motivation towards music at all, and towards familiar music(al language); this is the amplification which causes resonance in the radio circuit. Thus, we can posit a model where a tendency to learn accompanied by positive somatic responses to stimuli similar to ones which are already remembered, but different enough to cause new learning, leads to a narrowing of the band of "possible musics" in any given society. This is illustrated in Figure 1. This model of musical development accounts for the development of different musics in different societies because starting points — even random ones — are likely to have large effects in the long term, because of the feedback. It is worth mentioning here, though it is not the topic of the current discussion, that a hypothetical account of individual and social creativity is implicit in the above reasoning, also.

This means that we can posit a reasoned hypothesis on the development of human music over time. Perceptual mechanisms require little argument beyond biological evolution; the higher-level cognitive functions, in the above model, can develop as channelled by the positive feedback; the very existence of the somatic response posited by Biederman and Vessel (2006) gives a possible reason why it would happen at all. This hypothesis is satisfying to Ockham's Razor: it requires no assumption of universality of music, but only of universality of learning, which is well-established.

There is a chicken-and-egg question here, though. Even if the somatic propensity of "infovorous" humans (Biederman and Vessel, 2006) accounts for a tendency to enjoy music, there needs to be a level on which its development is grounded. It is reassuring, then, that some (proto-)musical faculties are hard-wired (or at least learned so early/generally that we can safely treat them as such) — for example: pulse grouping into twos and threes; spectral fusion; and auditory streaming (Bregman, 1990). Those faculties which seem to arise from our propensity for implicit learning — for example, harmony (Ponsford et al., 1999), compound rhythm, segmentation (Pearce and Wiggins, 2006b) — can then be viewed as constructs. Thus, physics provides the basis on which general audition is founded; survival provides the selection force that perpetuates general audition; human (associative) memory and learning (also a strong selectional faculty) provides the rest.

3. The Papers

In Figure 2, I present one way of locating the papers in this volume within the field of music cognition and in relation to each other, which highlights my particular preoccupations in relation to the topic (as discussed above and in detail below). In this section, I explain the structure and notation of the diagram. In later sections I will analyse and synthesise the material in the papers.

First, the colours in this diagram are significant to varying degrees. Each paper is represented by a box and an arrow, intended to indicate the nature of the contribution. Each box outline and associated arrow is coloured differently, to allow easy reading of the diagram. Each box filling is chosen from five colours, as specified in the key of the diagram, indicating the primary and, in some cases, secondary way(s) in which the paper content is of interest; boxes filled with gradient colours indicate multiple contribution; some papers have tertiary contributions which are not represented. Boxes with black outlines and lilac filling indicate primary topics of interest; and, finally, dotted boxes with black outlines indicate subsidiary areas of interest in which contributions are made. The meanings associated with the colours will be explained below.

The boxes are arranged as far as possible to indicate the level of abstraction at which they position themselves. For example, Lemström and Pienimäki (this volume) are interested in a very fundamental and detailed issue in the application of computers to cognitive modelling; in contrast, Selfridge-Field is interested in modelling social behaviour, which is, in at least some senses, a much more abstract enterprise.

Finally, the arrows connect their parent boxes to the topics and sub-topics to (at least) which they contribute. The outline colour-coding should make these relationships, which are convoluted, easier to see. The only type of contribution which I will not discuss in detail below is “Ground Truth” — the nature of and need for this is uncontroversial.

4. Cognitive Modelling Methodology

4.1. Descriptive and Explanatory Cognitive Models

The first issue I consider underpins the whole of the methodology of Cognitive Science. It is the distinction between models which are (in my terms) *descriptive* and those which are *explanatory*. I draw this distinction within a positivist framework: my

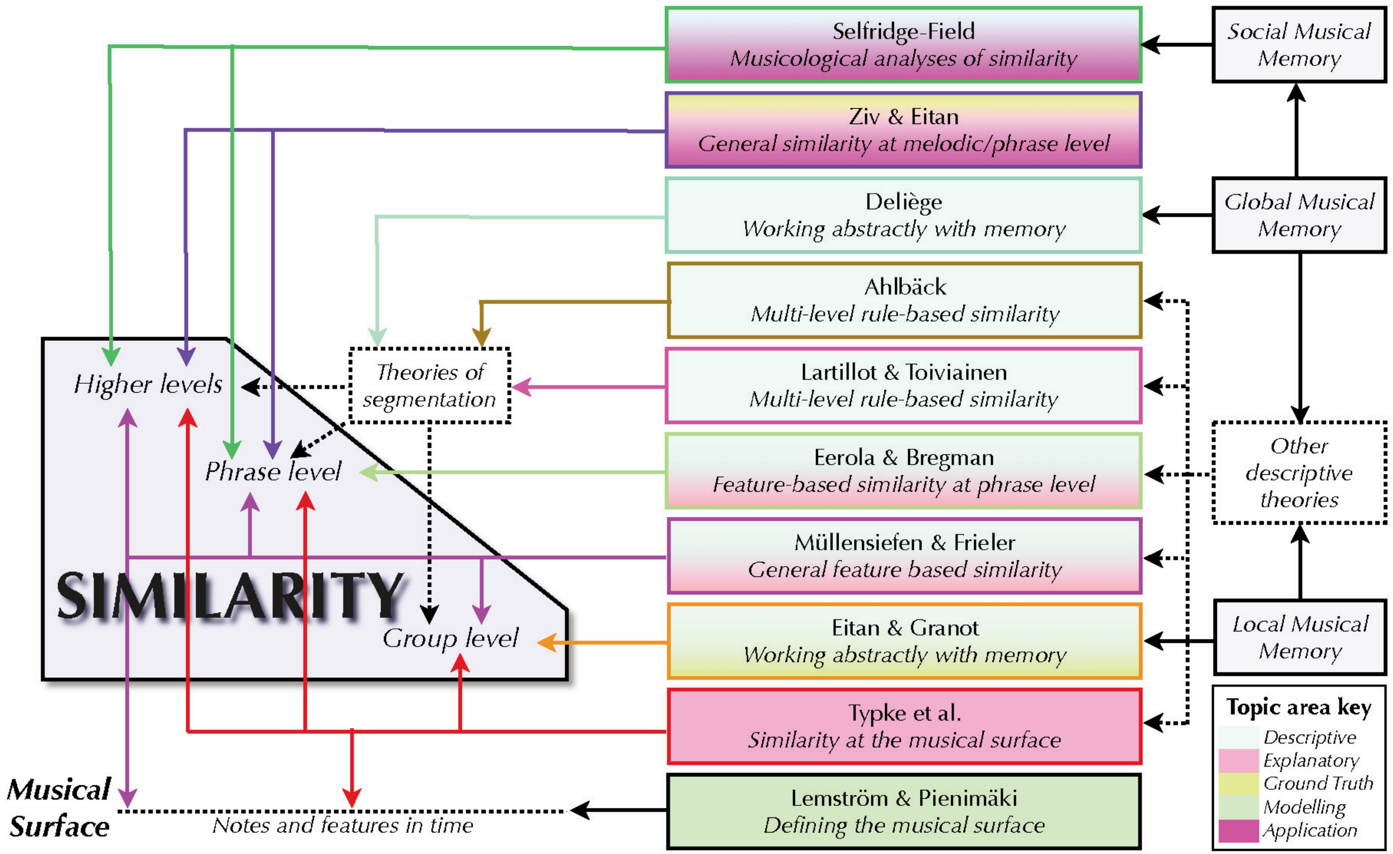
Selfridge-Field
Musicological analyses of similarity
Ziv & Eitan
General similarity at melodic/phrase level
Deliège
Working abstractly with memory
Ahlbäck
Multi-level rule-based similarity
Lartillot & Toiviainen
Multi-level rule-based similarity
Eerola & Bregman
Feature-based similarity at phrase level
Müllensiefen & Frieler
General feature based similarity
Eitan & Granot
Working abstractly with memory
Typke et al.
Similarity at the musical surface
Lemström & Pienimäki
Defining the musical surface
Social Musical Memory
Global Musical Memory
Other descriptive theories
Local Musical Memory
Topic area key
Descriptive
Explanatory
Ground Truth
Modelling
Application
Theories of segmentation
Higher levels
Phrase level
Group level
SIMILARITY
Musical Surface
Notes and features in time

Figure 2.

A map of the papers: see Section 3 for commentary.

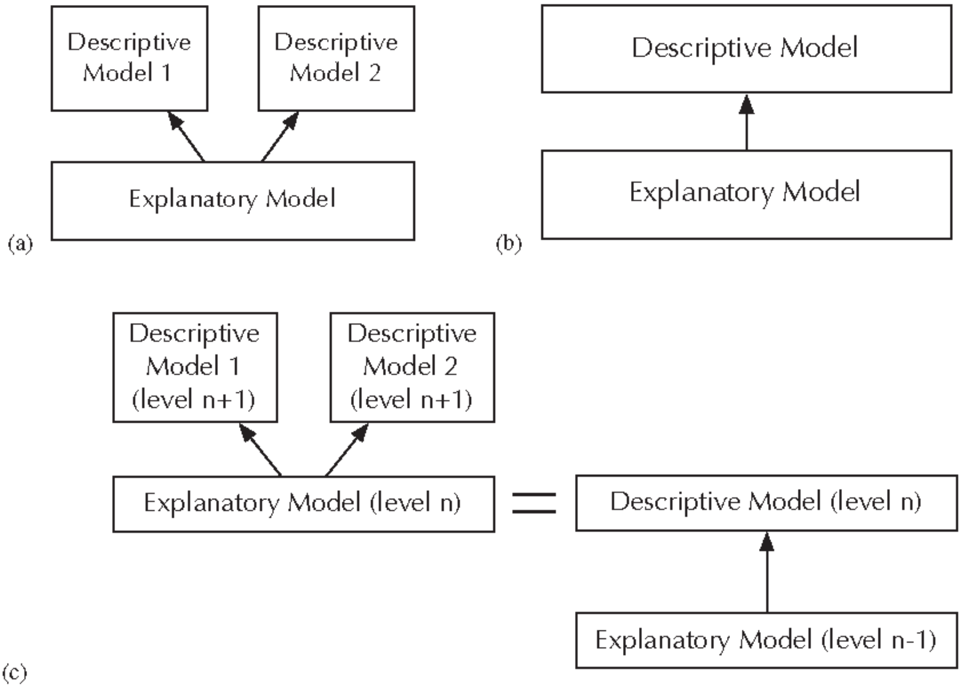
Descriptive Model 1
Descriptive Model 2
Explanatory Model
(a)
Descriptive Model
Explanatory Model
(b)
Descriptive Model 1 (level n+1)
Descriptive Model 2 (level n+1)
Explanatory Model (level n)
=
Descriptive Model (level n)
Explanatory Model (level n-1)
(c)

Figure 3.
An Explanatory Model can underlie one (or, preferably, more) Descriptive Models; Models can be explanatory at one level of abstraction and descriptive at another — this allows us to dig further into our topic.

aim is to achieve models which are as close as possible to the behaviour observed in the world, not to search for some more mysterious "truth" underlying them.

By "a descriptive model", I mean a model which describes a behaviour; it says what will happen in response to an event (in our case a musical stimulus) and it may say when it will happen. A descriptive model is usually derived by careful observation of empirical results and then formulated in terms of declarative or imperative rules, each (combination) of which deals with a specific case. Examples of such theories are the Gestalt Principles and the Generative Theory of Tonal Music (Lerdahl and Jackendoff, 1983).

By "an explanatory model", I mean a model which not only describes behaviour, but one which does so in terms of a (possibly more general) *underlying theory*; in other words, it says not only what will happen in response to an event, or when, but also (in some terms and at some level of abstraction) why and/or how. From this perspective, the rules of a descriptive model are codifications of the visible effects of a deeper, and perhaps invisible, process captured by the explanatory model.

These two definitions can be looked at in various ways. First of all, in a general cognitive-scientific kind of Grand Plan, we can look for new explanatory models which can underpin existing descriptive ones. An example of this would be the current interest in statistical learning-based theories of cognition (e.g., Pearce and Wiggins, 2006a; Lotto and Purves, 2002), which can account for behaviours in more than one modality of perception. This gives us an arrangement like that shown in Figure 3a. However, it is not a requirement that an explanatory model unite multiple descriptive ones. By definition, an explanatory model has more *explanatory power* (because it explains how and why, and not just what) and therefore, if correct, is an advance beyond a corresponding descriptive theory (Figure 3b). However — and therefore — it does not follow (as might naïvely be supposed) that descriptive models are less useful or worthwhile than explanatory ones. Indeed, the level of understanding required to formulate a descriptive model unambiguously is often the level of understanding needed to begin to build an explanatory one. Further, formal descriptive models can be very important in testing explanatory models (at least in a preliminary way) without recourse to expensive and time-consuming empirical work. In particular, and importantly for cognitive science, a model which is explanatory at one level of abstraction may be descriptive at another, and thus may call for a new layer to be added to explain *further* (Figure 3c). For example, the statistical methods to which I refer below, which I claim may be explanatory of other descriptive models, are an abstraction of apparent *mind* operation: they say nothing about how a *brain* works, but just describe what it does (at a certain abstract level). A new model which accounts for the predictions of these models in terms, for example, of actual brain function, would be explanatory in the terms I am aiming to define here.

An alternative way of looking at the descriptive/explanatory distinction is in terms of computer programming (of models). If one can write a program whose behaviour is explicable in terms of nothing more than case-by-case if-then rules, then, as a model, it has less explanatory power than a program which embodies a uniform mechanism which happens to produce a desired outcome (especially, again, if that mechanism can account for other models also). This is tied up with the idea of *emergence in complex systems* (e.g., Solé et al., 2002), and therefore with notions of complexity itself, and it is important to understand that it is possible to write rule-based systems which are capable of producing emergent behaviour which could not have been predicted merely by looking at the rules alone — this is one of the many consequences of the Halting Problem (Turing, 1936) — and therefore this is not an argument against case-by-case rule-based systems in general. In other words, not all rule-based systems are merely descriptive.

This idea is coupled with the fact (often unpalatable and/or incomprehensible to the public at large) that it is possible for computers to do things for which they have not been explicitly programmed by a person. Machine learning systems (Widmer, 2006, gives a snapshot of the state of the art in musical machine learning) whose behaviour matches either empirical data or empirically derived descriptive models hold out a promise of many more explanations in future. So there is a further related distinction here: models whose behaviour is *programmed*, and models whose behaviour is *emergent* — either because it is learned or because the model is sufficiently complex (in the mathematical sense) to give rise to emergent behaviour anyway. Generally speaking, then, these two terms correspond with descriptive and explanatory, though we must be aware of this confusing grey area between them.

Ultimately, therefore, cognitive scientific methodology could be summarised as a process of descriptive modelling, followed by explanatory modelling (Figure 3a and b), which is then followed by a re-description of the domain of discourse as the phenomena under study are better understood (Figure 3c). This last step might also be described as a progressive redefinition of the terms of the debate as understanding improves. In any case, we repeat this process until, ultimately, we reach the level of brain chemistry.

Perhaps all this is obvious to some readers; I would argue that it is not so to all, and discussions at the Symposium of which this volume is an outcome support my argument. I suggest that it is important to understand that descriptive modelling, which necessary and important, is not enough on its own to achieve a full understanding of music (or any other psychological phenomenon). It is therefore also important to understand which models contribute in which ways, so that a clear path through the landscape of possibilities may be found. To this end, I have coloured those papers in this volume, in Figure 2, which are directly concerned with modelling either blue (for descriptive) or pink (for explanatory), or both where a paper makes a contribution in both areas. Those papers not featuring these colours are not directly concerned with actual modelling (though they all support that activity in one way or another).

4.2. An Example

As a simple example of how the interaction between description (i.e., rule-based) and *potentially* explanatory (in this case, data-driven) models can be enlightening, I have applied SIATEC[1] (Meredith et al., 2002), a purely geometrical (and non-cognitive) algorithm for discovering structure in sets of data points to the data used by Lartillot and Toiviainen (this volume). I enhanced SIATEC with one (admittedly *ad hoc*) rule: that if a discovered structure had a gap of more than 1 beat, it would be broken into two structures; however, I applied this rule uniformly throughout the computation, and not as a special case. To summarise: SIATEC used in this way on monophonic music finds sequences of notes which are repeated, and their repetitions, where the sequences contain gaps or unmatched notes of no more than 1 beat in duration.

This is an interesting comparison because Lartillot and Toiviainen's model is descriptive, rule-based and — crucially — very knowledge-rich. SIATEC's "analysis" is shown in Figure 4. The letter annotations shown where SIATEC corresponds with

[1] Many thanks to David Meredith for re-implementing SIATEC for me, so that I could do this.

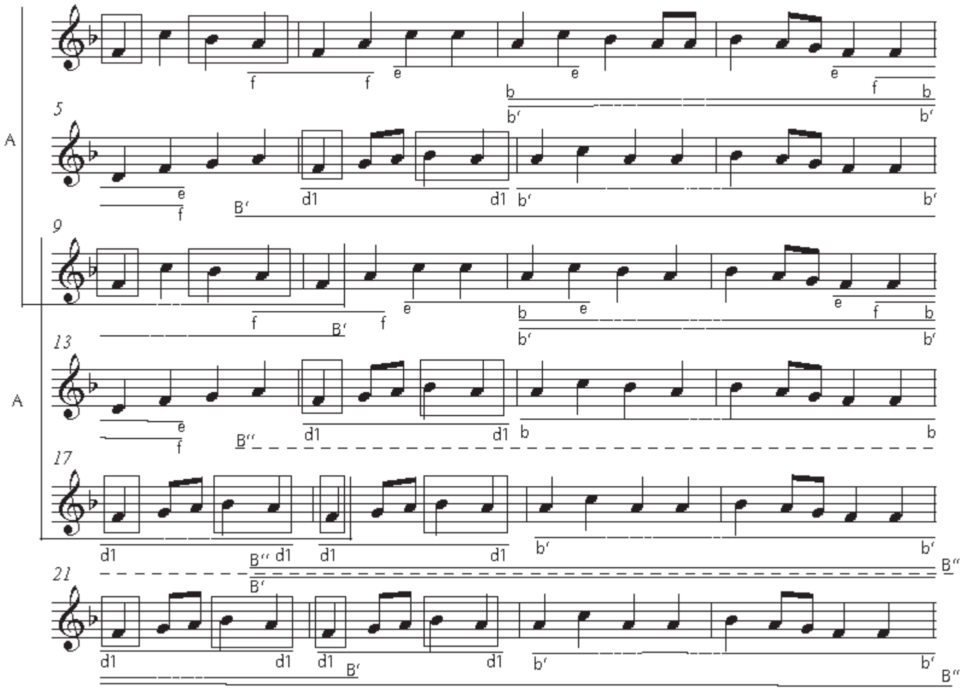

Figure 4.

SIATEC on Maria muoter reinû maît. The letters correspond with Lartillot and Toiviainen's; gaps in the patterns discovered are denoted by dotted segments of lines; the first occurrence of B″ is marked with a dashed line to make it easier to see. Note that the omitted notes in the original analysis are omitted here automatically as part of the discovery process, and not as an ad hoc rule. The boxes denote patterns not discovered by Lartillot and Toiviainen's system; I suggest they are musically significant figures, notwithstanding their non-contiguity.

Lartillot and Toiviainen's — these are not always exact matches, and the ways in which the two approaches differ are interesting. The fact that a very simple uniform process based on purely geometrical principles and no psychological content, coupled with a very simple but appropriate representation of music, can generate near-identical answers to the knowledge-rich model is perhaps surprising and begs the question of whether it is an explanatory model underlying the descriptive one.

However, I emphasise that I am making no such claim at this juncture. Rather, I suggest it is interesting to look at the ways in which the predictions of these two approaches differ. SIATEC works on two very simple principles: first, that repetition is significant in music; and, second, that musical time can be modelled geometrically to useful ends. There is no claim that these principles cover all possibilities: this is clearly not the case. But what we see here, in the mismatch between the outputs of these two models, is data which could, in principle help us to refine SIATEC into an explanatory theory, which would in principle be a step forward, because SIATEC can deal with polyphonic music, where Lartillot and Toiviainen's model cannot.

Thus, we might propose SIATEC+ (SIATEC plus some future refinements) as an explanatory model of Lartillot and Toiviainen's model, and conversely their model would be a test generator for SIATEC+. But we could also compare SIATEC+ with Typke et al. on a playing field refereed by Lartillot and Toiviainen, the two more uniform (and therefore potentially explanatory) models vying to be the better explanation of the descriptive one.

5. The Musical Surface and Beyond

5.1. Perception vs. Cognition

There is no precise definition of which I am aware of the boundary between what we call "perception" and what we call "cognition"; however, we sometimes use these words as though there were such a dividing line. From my perspective, as one interested in computational modelling of cognitive processes, it is particularly important, when attempting to model human behaviour, to clarify such questions: unless one has a precise specification, one simply *cannot* write a correct computer program.

For the purposes of the current discussion, there is a very natural point at which to draw a line between perception and cognition, which also happens to be the *musical surface* (Jackendoff, 1987) on which all the work presented here is based: the level of musical notes as heard. The argument for this is that the musical note is a perceptual construct, constructed of phase-related harmonics, almost irreversibly *fused* together. While listeners can demonstrably and effortlessly group notes together (into melodies, chords, etc.), it requires some learned skill to "de-fuse" the harmonics of a given note, and even when we are able to hear individual harmonics within a note, we are unable to deconstruct the entire percept: those harmonics are heard in relation to the timbre of the note and not as separate notes themselves. Our musical surface, then, is the surface of notes as heard in a musical listening experience; anything above that level may be deemed cognitive, and everything below it may be deemed perceptual. Grouping and metre are multi-layer phenomena which straddle this boundary: low-level grouping (in time and in pitch — auditory streaming) is a perceptual process over which we have little control (like fusion); higher-level grouping seems to be a higher-level kind of process, involving relatively detailed memory of past experience. This is by no means the only reasonable musical surface to consider, but it is that which is considered by default in this volume.

In summary, then, the choice of musical surface here ultimately arises from an implicit, but well-established, perceptual/cognitive boundary. The authors are looking at those structures which are, to some extent, introspectively sub-divisible, and not at those which are not. This seems to be a strongly defensible position, since it is much easier to make the case in the evolutionary context that these higher level cognitive features are *musical*, as compared with the more fundamental auditory processes shared by mammals, which must have evolved much earlier since they are related to basic survival mechanisms.

5.2. Dimensions of the musical surface

The identification of the musical surface is further complicated by the implicit assumption, made in some of the work here presented, that a musical score directly represents such a musical surface, when a piano roll is actually a more accurate approximation; of course, all the authors well know that a score is only an abstraction of the perceived surface — but we sometimes allow ourselves to gloss over the point, and sometimes precision can be lost as a result. Ziv and Eitan (this volume) address this issue head-on, and provide some evidence that musicological analysis (based on the score — or perhaps on the musicologist's prediction of the psychological effect on an average listener of a given performance of the music?!) does not, even for well-known, tonal repertoire, match precisely with the reported experience of listeners, even when they are skilled musicians.

Because the different papers are working at different levels of abstraction, it is necessary, when we come to consider computational models, to think about the *affordances* offered by each level to the work using it. Lemström and Pienimäki (this volume) and Typke et al. work explicitly in a generalised mathematical abstraction and discuss the affordances of their representations to their models. This is important because, obviously, one cannot compute with data which is not available in one's representation; and, less obviously, choosing an inappropriate representation can compromise or enhance the operationalisation of models as programs — I return to this issue below.

In nearly all music modelling, time seems to be a distinguished dimension: we may model perceptual time (perhaps as score time) or real time — and we may be interested in the difference between the two. And, *within time*, many dimensions are changing. It is far from clear how these dimensions and, crucially, the *dynamic* relationships between them, should be represented.

Lartillot and Toiviainen (this volume), in more detailed abstract thinking, distinguish between absolute note name (without reference to key or temperament) and scale degree with respect to key note (i.e., local harmonic function). These two ideas are familiar to musicologists — but they have different mathematical properties, and so will produce different analyses (for example, the latter, in most models, will lead to transposition-invariance, in the obvious way). Similarly, Typke et al. propose enhancements to their model based on explicitly encoded tonal function, and work on the basis that, if one chooses the right representation(s), one can sometimes apply mathematical methods from elsewhere in science directly as cognitive models. Figure 4 (above) is an example of this.

Many of the papers here included use feature-based musical surfaces, which are viewed differently from the note-as-percept-based approach. They do not presuppose any particular underlying representation (though, here, they all actually use the musical surface specified above), and they are defined in terms of the dynamics of properties (or *features*, in engineering terms) of note percepts and/or groups of note percepts in time. Eitan and Granot (this volume) place their work precisely on the boundary between note-based and feature-based representations: they identify categories of features or notes whose dynamics relate together as perceptually similar, supplying important ground truth data for this kind of model. Müllensiefen and Frieler take a different approach, pulling together a large collection of different features, and attempting mathematically to find a universal model of melodic similarity by manipulating the multidimensional space defined by the different features; Eerola and Bregman's view is similar, but they aim to identify the contextual significance of the different features.

Nearly all the authors are interested in structures which lie above the musical surface and which may be interpretable independently of it. This begs interesting questions about what cognitive (and computational — see below) representations may encode these structures. For one example, in the terms used by Deliège (this volume): What distinguishes a cue from the structure cued, and what relates a structure to its imprint? Deliège gives no specific definition of a cue: it is specified as something abstract from the individual notes, with high information content, in context, so it attracts attention (this interpretation will be significant below). For another, in the terms of Eitan and Granot: What are "intensity" dynamics and how are they encoded? These aspects of our research have barely begun to scratch the surface. Now that we have abstract models which seem to be good, we can begin to mine deeper by attempting to *explain* them in the sense of my term defined above.

5.3. Knowledge and Data Representation

Having decided what information to represent, we must design the *formal* representation of these models. Ultimately, for a model to be useful in any sense other than academic, therapeutic or educational, we need to operationalise it using a computer. This means (at least with current technology) that we must not only choose the right musical surface, but also that we must represent it, within the detail of our theory, in an appropriate way. This is far from trivial, as the extent of work on music representation in the 1980s and '90s testifies (Wiggins et al., 1993, provide a survey), and, indeed, the debate still rages on, fuelled, not least, by the necessity of providing engineering solutions to commercial problems (Wiggins, 2007). To help in this direction, the literature gives us some simple thinking tools.

For example, Wiggins et al. (1993) give two dimensions for the evaluation of music representations (computer-based or otherwise): *expressive completeness*, which is the extent to which the original form of the music may be reconstructed from the representation; and *structural generality*, which is the extent to which structure implicit in the original form may be denoted and annotated. In general, it is good to maximise both the expressive completeness and the structural generality of one's representation; however, there are trade-offs to be made, for example, in that the musical surface of any given activity (such as any of those discussed in this volume) is a cut-off point, below which detail becomes irrelevant.

A very common representation for (monodic) music has been the so-called "symbol string", or less disingenuously, the symbol sequence. There is a large literature of work on applications of "string algorithms" (algorithms for processing sequences of symbols) to music (Clifford et al., 2005, survey the state of the art). However, Lemström and Pienimäki (this volume) argue that this representation is not appropriate to music in general, and that, instead, a geometrical representation should be used, at least for polyphonic music; indeed, unless action is taken to force the string representation into a quasi-geometrical one (Mongeau and Sankoff, 1990), this representation is not expressively complete enough, in the sense of Wiggins et al. (1993), above, for the many of the tasks discussed in this volume. This is a careful contribution to the most fundamental aspects of computer modelling: without the right representation, it is unlikely that a correct model will be achieved.

6. Grouping, Structure, Grouping Structure, and Segmentation

To return to the discussion of the present volume: for the purposes of these papers, the musical score, perhaps enhanced with detailed expression data, and certainly with repeats, etc., written out in full constitutes an ecologically valid musical surface. The note percept is the atomic symbol, which can be broken up no further, and categorical pitch perception lets us abstract to scale degree. We can be content with a score-like representation of time because our metrical mechanisms are below the level of this particular musical surface, and therefore we are concerned not with simulating them but merely using their output (in other words, the metrical presentation given in a score). However, the perceptual processes which I am placing below our musical surface also have effects beyond it: fusion contributes to the chord percept and to timbre; and rhythmic perception which gives us tactus also gives us higher level rhythmic structures, and we need to represent these structures which are built on top of the musical surface. Smaill et al. (1993) discuss these issues in detail, and present the beginnings of a formal system to facilitate such representation. These groupings and their abstract descriptions will be crucial to computer implementations of models such as that of Deliège (this volume) and Eitan and Granot.

Many of the papers discuss segmentation, which is the flip-side of grouping, because it is fundamental to similarity: to continue the radio analogy, structural similarity seems to be a kind of carrier wave for more subtle kinds of melodic and harmonic similarity. Deliège, Eitan and Granot, Lartillot and Toiviainen, Müllensiefen and Frieler and Ahlbäck all give models or partial models of segmentation/grouping as part of their contribution, and Selfridge-Field implicitly requires segmentation *a priori.*

Segmentation is relevant here for several reasons. Firstly, on a practical and general level, perceptual segmentation is a means of reducing the amount of information to be processed by the brain. That segmentation is related to memory storage is demonstrated by the elegant experiments of Tan et al. (1981) and Peretz (1989), in which "phrases" straddling musical boundaries predicted by Gestalt rules were less well remembered than those which were between boundaries. And given this, structural similarity is likely to be *determined* by segmentation, because phrases straddling boundaries are not perceived as phrases and therefore similarity between them cannot be perceived at all. Those authors in this volume who do not address or presuppose segmentation do not need to either because their musical focus is within segments (e.g., Eitan and Granot), or because their topic does not require such musical precepts (e.g., Lemström and Pienimäki).

But here we see another chicken-and-egg problem, this time for segmentation models (and thence for similarity models) based on emergent properties of other underlying models; this is a general problem for what I am calling explanatory models of anything. In the purely descriptive case, where we assume rules *a priori* (Figure 5a), we simply turn the handle and our required behaviour (in this case, segmentation) drops out of the model. In an explanatory model, there is nothing, other than the properties of the model itself to give us that behaviour. We have to hope, therefore, that these properties will give rise to the same effect as, and hence account for the rules of, the descriptive model (Figure 5b).

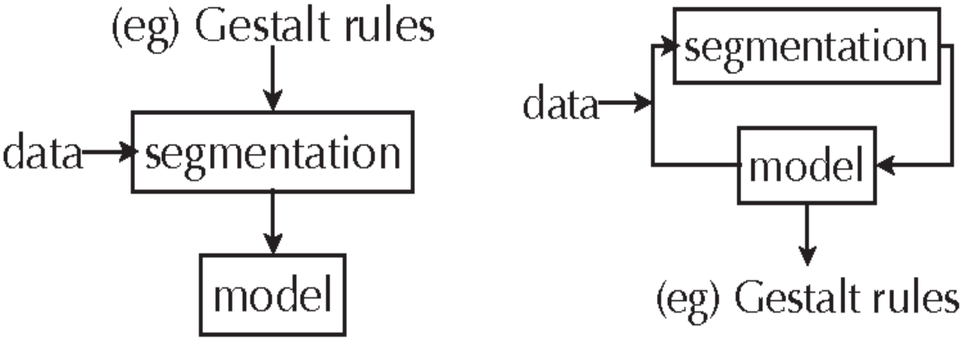

Figure 5.

Accounting for Segmentation in Music: (a) is a descriptive account and (b) is explanatory.

Fortunately, for the case I am making here, there *is* evidence that segmentation can indeed be predicted from memory-based models alone, both in speech and "tone words" (Saffran et al., 1999) and music (Ferrand et al., 2002; Ferrand, 2005; Pearce and Wiggins, 2006b), without recourse to *ad hoc* musical rules. As an example, Figure 6, shows how the information-theoretic properties (Shannon, 1948) of an autonomously learned statistical model of music (based on the same musical surface as the papers here) can predict segmentation. In this preliminary work, the music studied was *Two Pages*, a minimalist piece by Philip Glass, which has the useful properties of being monodic, monotimbral and isochronous. Therefore, it is possible to study it with relatively simple models, which increases the probability of being able to understand what is going on, but it has more ecological validity (as a published piece of music) than more artificial stimuli. The model (Pearce, 2005) consists of a long-term and a short-term memory, each of which is a complex variable-length n-gram model: the long-term memory is trained on the Essen Folksong Database (Schaffrath, 1992). The heavy line in the figure shows, event by event, the (information-theoretic) entropy of the distribution predicting each note immediately before it is "perceived". The numbers are smoothed by taking a moving average over a narrow band of values to the left of the value shown (which explains why the changes in entropy are sloping and not vertical). The vertical lines in the diagram indicate properties of the score: the darker ones are the sections of the piece as annotated by the composer; the lighter ones are the boundaries between the repeated figures (under additive/subtractive process) of which the piece is constructed. This graph is suggestive evidence (but by no means conclusive proof — the work is still in progress) that there is a link between entropy and structural boundaries in music. More detail is given by Pearce and Wiggins (2006b); the relevance of the work here is that the model contains no hard-wired musical rules at all, and the only musical knowledge encoded here is that required to achieve the note percept (which is below the musical surface in which we are interested in this volume).

7. An Alternative Emphasis: Memory and Context

Several of the authors in this volume appeal to memory, short- or long-term, as part of their theory. This seems to me to be important, especially in view of my proposal, above, that these behaviours may be explicable (in my strong sense) without appeal to *a priori* musical rules.

The Cue Abstraction theory (Deliège, this volume) is stated in terms of memory, and seems very ripe for explanation in terms of memory based models akin to that of Pearce (2005); there is a strong flavour of information theory about the notion of a cue being something *salient in context*. Similarly, but on a different level, Selfridge-Field's models relate strongly to questions about the development of musical languages over time, and how this might be one factor reinforcing the formation of societies (Bown and Wiggins, 2005; Bown, 2006) and between individuals within societies (Chan and Wiggins, 2005, 2006). Eerola and Bregman and Müllensiefen and Frieler include memory and learning models explicitly in their experiments.

A very important difference between the work here presented is the (somewhat fuzzy) line between models which are *models of similarity* — that is, the perception of musical similarity modelled as a general and quasi-universal behaviour — (e.g.,

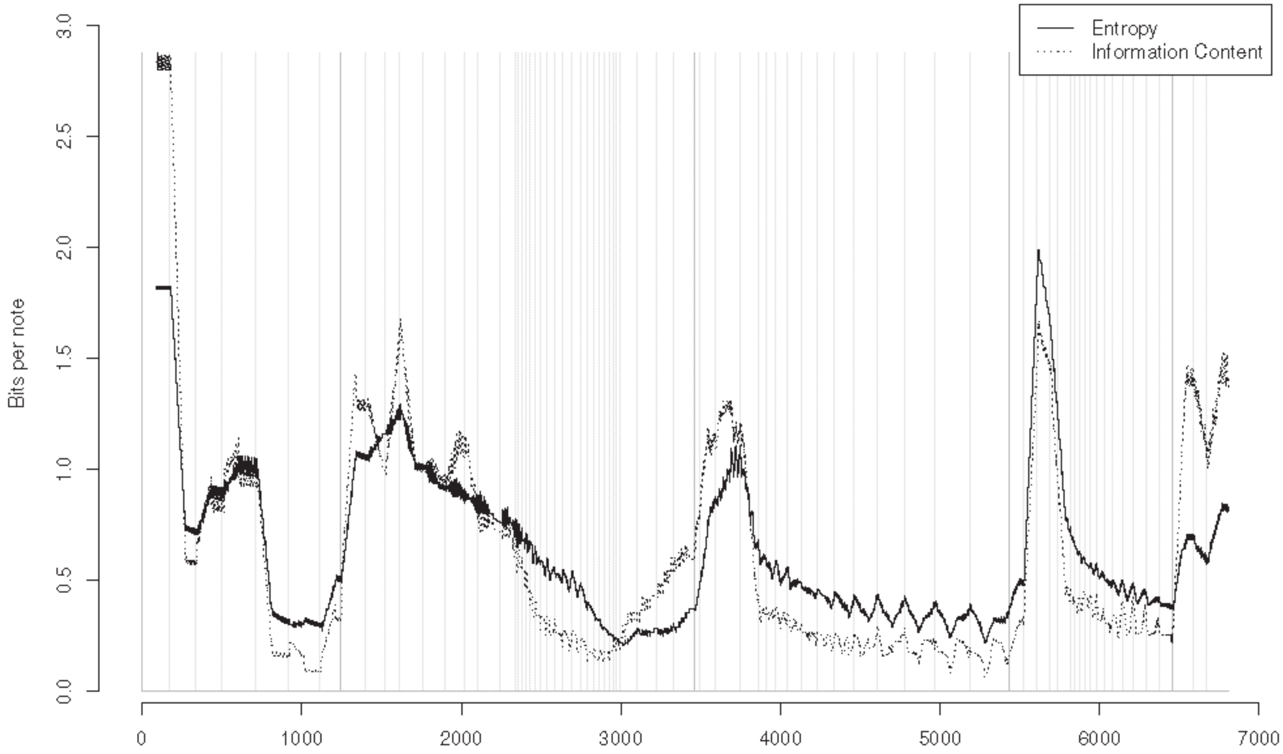

Figure 6.

Segmentation of Philip Glass' Two Pages. The heavy curve shows the entropy of the note distribution predicted by the model as the piece proceeds (the x-axis is note number, which, since the piece is isochronous, is the same as real time), shown here with moving average to make the graph clearer. The heavy vertical lines mark the sections of the piece as defined by the composer; the lighter vertical lines mark the boundaries between the repetitive figures from which the piece is constructed. Thanks to Marcus Pearce for preparing this graph.

Müllensiefen and Frieler) and *frameworks within which similarity may be considered* (e.g., Deliège). First, I clarify this distinction a little, using these two examples.

Müllensiefen and Frieler (this volume) aim for a *general* measure of musical similarity, expressed via dimensions in a Euclidian space, each of which is the output of one feature applied to the data. The aim is either to model similarity as an absolute measure, independent of the individual making the judgement, or, perhaps, to arrive at a sort of "average" similarity which is supposed to be at the core of all similarity judgements for the kind of music considered.

Deliège (this volume), on the other hand, gives a framework, unrelated to specific data, within which more individualistic views of similarity may be expressed — because the idea of the cue and the imprint are less precisely specified, and so do not tie us to a particular, fixed model.

This difference is important because of the lack of agreement between individual respondents about musical similarity. There is no room in a general model based on the idea of a fixed, external ground truth, for individual variation, unless it be by building a separate "general" model for each individual. However, even such a model will not account for asymmetry of similarity judgements, nor of the effect of third stimuli on individuals' judgements (Tversky, 1977). The framework-type theories leave wiggle-room for this, but do not actually account for it unless they (implicitly or explicitly) model memory. This has consequences for experimental technique in this context: we take care to control for priming effects by randomising the order of stimuli in our similarity experiments — but by doing so, I claim, we arrive at what can only be a weak model of similarity, because the priming effect is *part* of similarity perception, and not a confounder at all.

The idea that musical similarity can be studied as an absolute, external to individuals, is akin to the Romantic notion that there are absolutes in music itself (Schenker, 1925, 1926, 1930). It relies on the existence of something like a Platonic ideal (Plato, 1997) of music, against which all else can be measured, and, indeed, we must assume this supposition to be insidiously implicit in any activity of studying music unless it is explicitly denied. My own position is strongly opposed to this: that "music" is a social and psychological construct, and therefore that it can be studied only *approximately* if it is dissociated from these essential causes — it is for this reason that I have tended to write "musical behaviour", rather than "music" in the current paper and elsewhere.

The specific point I make here is that musical behaviour, while (obviously) reliant on auditory perception, is fundamentally a matter of memory: (associative) memory encourages us to engage in musical listening and production, it allows us to parse musical sound, and the nature of memory can account for the development of musical behaviour in society, as well as for the "speciation" of music between societies.

Since musical similarity tells us, at least some of the time, something fundamental about the "meaning" of music — I use scare-quotes here because I do not know what this phrase really means (Wiggins, 1998) — it follows from the hypotheses above

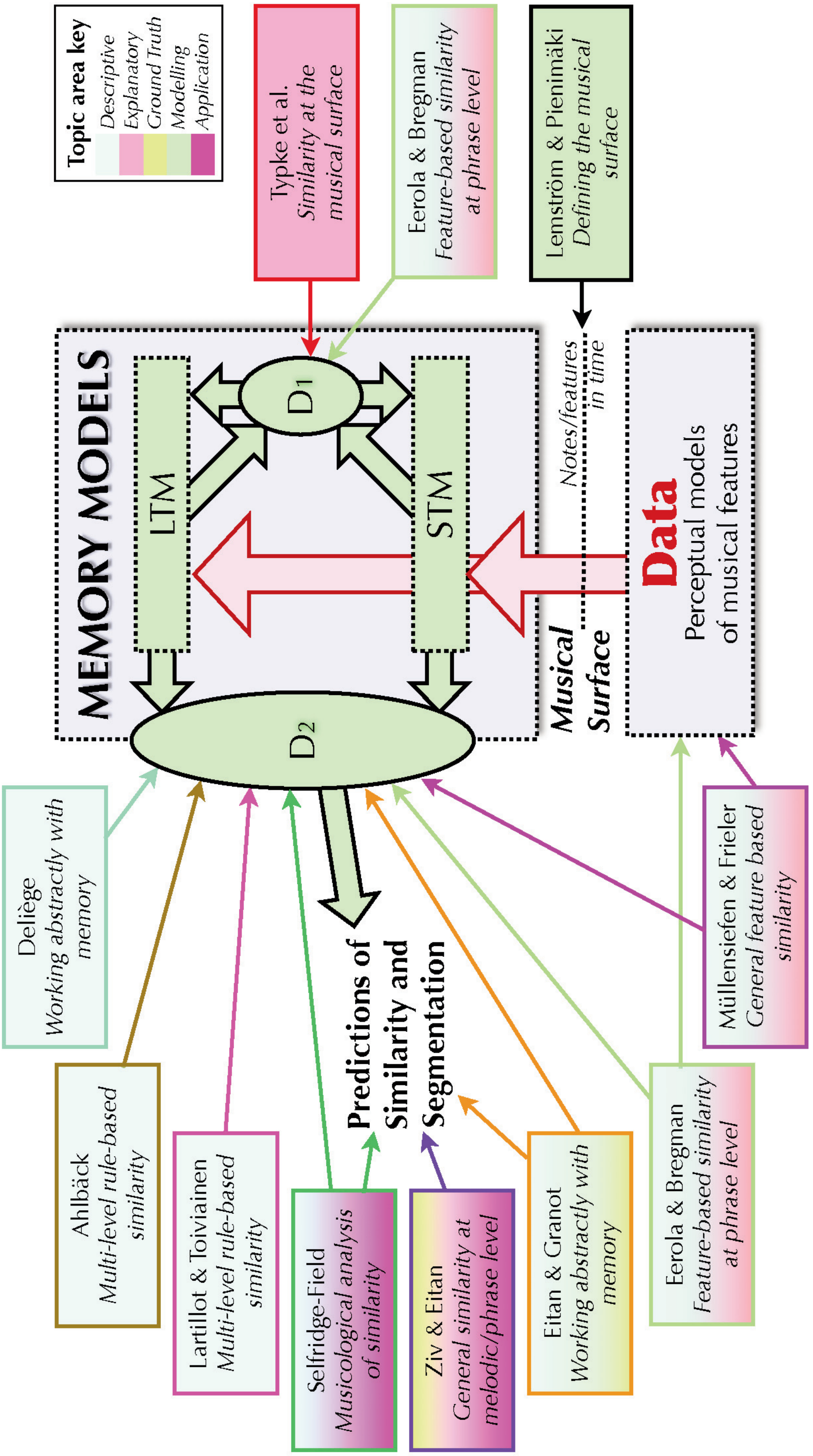
Topic area key
Descriptive
Explanatory
Ground Truth
Modelling
Application
Typke et al.
Similarity at the musical surface
Eerola & Bregman
Feature-based similarity at phrase level
Lemström & Pienimäki
Defining the musical surface
MEMORY MODELS
LTM
STM
D1
D2
Notes/features in time
Musical Surface
Data
Perceptual models of musical features
Predictions of Similarity and Segmentation
Deliège
Working abstractly with memory
Ahlbäck
Multi-level rule-based similarity
Lartillot & Toiviainen
Multi-level rule-based similarity
Selfridge-Field
Musicological analysis of similarity
Ziv & Eitan
General similarity at melodic/phrase level
Eitan & Granot
Working abstractly with memory
Eerola & Bregman
Feature-based similarity at phrase level
Müllensiefen & Frieler
General feature based similarity

Figure 7.
Map of the proposed view of the research.

that it cannot be modelled without explicit reference to memory, and Tversky's (1977) work supports this.

My proposal, then, is that we should take the line followed by many of the researchers here, implicitly and explicitly, and study musical behaviour from a mimetic, memetic and information-theoretic standpoint. Only thuswise can we begin to give an account of human musical behaviour as whole which is not ethnocentric and which is nonetheless able to predict the behaviour of individuals. It must be noted, however, that taking this view does not in any sense devalue the contributions of the authors in this volume: all of the papers here contribute in one way or another to a memory-centred view of similarity, and I have attempted to illustrate this in Figure 7. This figure is, in a graphical sense, almost the results of turning Figure 2 inside out, moving memory to the centre, instead of similarity.

Thus, the modelling of musical similarity ceases to an aim itself, but becomes a lens through which we can view musical behaviour as a whole, in terms of cognitive constructs (Ahlbäck, Deliège, Eitan and Granot, Lartillot and Toiviainen, this volume), of musical memetics and sociology (Selfridge-Field), To achieve this end, we need data to work with (Eitan and Granot, Eerola and Bregman, Selfridge-Field, Ziv and Eitan) and appropriate methods to represent that data on computers so that models can be built (Lemström and Pienimäki). We also need to elucidate those behaviours inadequately captured by general models (Ahlbäck, Eerola and Bregman, Lartillot and Toiviainen, Müllensiefen and Frieler, Typke et al.) so that we can move in the direction of more detailed and precise models in future.

Acknowledgements

I gratefully acknowledge the input of my colleagues in the Intelligent Sound and Music Systems (ISMS) group at Goldsmiths, especially of Hamish Allan, Oliver Bown, Tak-Shing Chan, David Meredith and Marcus Pearce. Several participants at the ICMPC symposia in Bologna of which this volume is an outcome made helpful contributions. Some of the work presented here was funded by EPSRC project grants GR/N08049 and EP/D038855 and Andrew W. Mellon Foundation grant "MeTAMuSE" to the author, and an EPSRC DTA studentship award to Marcus Pearce, via City University, London.

Address for correspondence:
Prof. Geraint A. Wiggins
Department of Computing
Goldsmiths College, University of London
New Cross, London SE14 6NW
United Kingdom
e-mail: g.wiggins@gold.ac.uk

• References

Ahlbäck, S. (2007). Melodic similarity as determinant of melody structure. *Musicæ Scientiæ, Discussion Forum 4a.*

Biederman, I. & Vessel, E. A. (2006). Perceptual pleasure and the brain. *American Scientist*, *94*, 247-53.

Bown, O. (2006). The extended importance of the social creation of value in evolutionary processes: A proposed model. In Colton, S. & Pease, A. (eds), *Proceedings of the ECAI'06 Workshop on Computational Creativity*. URL: www.doc.gold.ac.uk/map01ob/bown_ijcai06.pdf.

Bown, O. & Wiggins, G. A. (2005). Modelling musical behaviour in a cultural-evolutionary system. In Gervàs, P., Veale, T., & Pease, A. (eds), *Proceedings of the IJCAI'05 Workshop on Computational Creativity*. URL: www.doc.gold.ac.uk/ mas02gw/papers/Bown Wiggins05.pdf.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA: MIT Press.

Chan, T.-S. T. & Wiggins, G. (2005). A computational memetics approach to music information and aesthetic fitness. In Colton, S., Gervás, P., & Veale, T. (eds), *Proceedings of the 2nd International Joint Workshop on Computational Creativity*. URL: www.doc.gold.ac.uk/ mas02gw/papers/ChanWiggins05.pdf.

Chan, T.-S. T. & Wiggins, G. (2006). More evidence for a computational memetics approach to music information and new interpretations of an aesthetic fitness measure. In Colton, S. & Pease, A. (eds), *Proceedings of the 3rd International Joint Workshop on Computational Creativity*. URL: www.doc.gold.ac.uk/ mas02gw/papers/ChanWiggins05.pdf.

Clifford, R., Crawford, T., Iliopoulos, C., & Meredith, D. (2005). Problems in computational musicology. In Iliopoulos, C. & Lecroq, T. (eds), *String Algorithmics*, NATO Science Series. London: KCL Press.

Deliège, I. (2007). Similarity relations in listening to music: How do they come into play? *Musicæ Scientiæ, Discussion Forum 4a.*

Dissanayake, E. (2001). An ethological view of music and its relevance to music therapy. *Nordic Journal of Music Therapy*, *10 (2)*, 159-75.

Eerola, T. & Bregman, M. (2007). Melodic and contextual similarity of folk song phrases. *Musicæ Scientiæ, Discussion Forum 4a.*

Eitan, Z. & Granot, R. (2007). Intensity changes and perceived similarity: Inter-parametric analogies. *Musicæ Scientiæ, Discussion Forum 4a.*

Ferrand, M. (2005). *A new cognitive model of musical melody segmentation (working title).* PhD thesis, University of Edinburgh.

Ferrand, M., Nelson, P., & Wiggins, G. A. (2002). A probabilistic model for melody segmentation. In *Proceedings of ICMAI'02*. Springer-Verlag.

Jackendoff, R. (1987). *Consciousness and computational mind*. Cambridge, MA: Mit Press.

Lartillot, O. & Toiviainen, P. (2007). Motivic matching strategies for automated pattern extraction. *Musicæ Scientiæ, Discussion Forum 4a.*

Lemström, K. & Pienimäki, A. (2007). Approaches for content-based retrieval of symbolically encoded polyphonic music. *Musicæ Scientiæ, Discussion Forum 4a.*

Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: The MIT Press.

Lotto, B. R. & Purves, D. (2002). The empirical basis of color perception. *Consciousness and Cognition*, *11*, 609-29.

Meredith, D., Lemström, K., & Wiggins, G. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, *31 (4)*, 321-45.

Mongeau, M. & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities*, *24*, 161-75.

Müllensiefen, D. & Frieler, K. (2007). Modelling experts' notions of melodic similarity. *Musicæ Scientiæ, Discussion Forum 4a.*

Pearce, M. & Wiggins, G. A. (2006a). Expectancy in melody: The influence of context and learning. *Music Perception*, *25 (5)*, 377-406.

Pearce, M. & Wiggins, G. A. (2006b). The information dynamics of melodic boundary detection. In Baroni, M., Addessi, A. R., Caterina, R., & Costa, M. (eds), *Proceedings of the 9th International Conference on Music Perception and Cognition.* CD-ROM, University of Bologna, Italy.

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.

Peretz, I. (1989). Clustering in music: An appraisal of task factors. *International Journal of Psychology*, *24 (2)*, 157-78.

Pinker, S. (1994). *The language instinct.* New York: William Morrow & Co.

Plato (1997). *The Republic* [Original 385BC, translated by J. L. Davies & D. J. Vaughan]. Ware, Hertfordshire, UK: Wordsworth Editions Ltd.

Plotkin, H. (1998). *Evolution in mind.* Cambridge, MA: Harvard University Press.

Ponsford, D., Wiggins, G. A., & Mellish, C. S. (1999). Statistical learning of harmonic movement. *Journal of New Music Research*, *28 (2)*, 150-77.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by hman infants and adults. *Cognition*, *70*, 27-52.

Schaffrath, H. (1992). The ESAC databases and MAPPET software. *Computing and Musicology*, *8*, 66.

Schenker, H. (1925). *Das Meisterwerk in der Musik, Volume I.* Munich: Drei Maksen Verlag.

Schenker, H. (1926). *Das Meisterwerk in der Musik, Volume II.* Munich: Drei Maksen Verlag.

Schenker, H. (1930). *Das Meisterwerk in der Musik, Volume III.* Munich: Drei Maksen Verlag.

Selfridge-Field, E. (2007). Social dimensions of melodic identity, cognition, and association. *Musicæ Scientiæ, Discussion Forum 4a.*

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. Also from Edinburgh as DAI Research Paper No. 511.

Solé, R. V., Ferrer, R., Montoya, J. M., & Valverde, S. (2002). Selection, tinkering and emergence in complex networks. *Complexity*, *8*, 20-33.

Tan, N., Aiello, R., & Bever, T. G. (1981). Harmonic structure as a determinant of melodic organization. *Memory and Cognition*, *9 (5)*, 533-9.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae*, *Special Issue 1999-2000*, 155-215.

Turing, A. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *2 (42)*, 230-65.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-52.

Typke, R., Wiering, F., & Veltkamp, R. C. (2007). Transportation distances and human perception of melodic similarity. *Musicæ Scientiæ, Discussion Forum 4a*.

Widmer, G. (2006). Special issue on music and machine learning. *Journal of Machine Learning*, (In press).

Wiggins, G. A. (1998). Music, syntax, and the meaning of "meaning". In *Proceedings of the First Symposium on Music and Computers*, Corfu, Greece.

Wiggins, G. A. (2007). 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., Miranda, E., Smaill, A., & Harris, M. (1993). A framework for the evaluation of music representation systems. *Computer Music Journal*, *17 (3)*, 31-42. Machine Tongues series, number XVII; Also from Edinburgh as DAI Research Paper No. 658.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, Monograph Supplement*, *9*, 1-27.

Ziv, N. & Eitan, Z. (2007). Themes as prototypes: Similarity judgments and categorization tasks in musical contexts. *Musicæ Scientiæ, Discussion Forum 4a*.

• Modelos de similitud musical

Intentamos alinear y comparar varios artículos en este Foro de Debate y extraer algunas conclusiones de ellos. La amplia extensión de los artículos impide llevar a cabo una comparación en detalle, y dicha comparación se lleva a cabo en el nivel formal, comparando la naturaleza de los modelos y técnicas propuestas, y los resultados obtenidos, y discutiendo cómo estos destacados trabajos contribuyen a nuestra comprensión del proceso cognitivo de la música. En conclusión, proponemos un punto de vista basado en la memoria y el aprendizaje para el que reclamamos todo el esfuerzo últimamente desarrollado.

• Modelli di similarità musicale

Ho cercato di allineare e mettere a confronto i diversi articoli di questo Forum di Discussione, e di trarne alcune conclusioni generali. Dato l'ambito così vasto degli articoli, non è possibile confrontarli in dettaglio, quindi la comparazione si è effettuata al meta-livello, paragonando la natura dei modelli e delle tecniche proposte e i risultati prodotti, e discutendo il modo in cui l'insieme di questi importanti articoli contribuisce alla nostra comprensione della cognizione musicale. In conclusione, propongo un punto di vista basato su memoria e apprendimento, sul quale a mio parere si orienta in definitiva tutto il presente lavoro.

• Modèles de similarité musicale

J'essaye ici de comparer les différents articles de ce Forum de Discussion et d'en tirer quelques conclusions générales. Étant donné la vaste gamme d'articles, on n'a pu tenir compte des détails ; la comparaison a donc été faite au méta-niveau en prenant la nature des modèles et des techniques proposés ainsi que les résultats présentés ; nous avons exposé la contribution que ces articles importants apportent à notre compréhension de la cognition musicale. En conclusion, nous proposons un point de vue fondé sur la mémoire et l'apprentissage qui, à mon avis, est le but ultime de tout ce travail.

• Modelle zur musikalischen Ähnlichkeit

Ich versuche die verschiedenen Aufsätze dieses Diskussionsforums aufzulisten und miteinander zu vergleichen und dabei einige allgemeine Schlussfolgerungen zu ziehen. Der Bereich, aus dem diese Aufsätze stammen, ist so weit gefasst, dass detaillierte Vergleiche nicht möglich sind und daher Vergleiche auf der Meta-Ebene hinsichtlich der Art der Modelle, der vorgeschlagenen Techniken sowie der Ergebnisse gezogen werden. Es wird diskutiert, wie diese wichtigen Aufsätze zusammen unser Verständnis musikalischer Kognition bereichern. Als Schlussfolgerung schlage ich einen Standpunkt vor, der auf Gedächtnis und Lernen basiert, wohin meines Erachtens letztlich alle Forschungsarbeiten weisen.