Musicæ Scientiæ
Discussion Forum 4B, 2009, 7-24

How similar is similar?

EMILIOS CAMBOUROPOULOS
Department of Music Studies
Aristotle University of Thessaloniki

• **Abstract**

In the first part of the paper a theoretical discussion is presented regarding the fundamental concept of similarity and its relation to cue abstraction and categorisation. It is maintained that similarity is by definition context-dependent and strongly interrelated to cue abstraction and categorisation. Emphasis is given to determining the "musical surface" that can act as a musically pertinent lowest level of abstraction on which similarity between musical entities can be measured. Then, each of these concepts is examined in more detail with respect to a number of research studies presented in the recent special issue of *Musicæ Scientiæ* on musical similarity (Discussion Forum 4A, 2007). Views claiming that a geometric piano-roll-like representation is the most appropriate choice for polyphonic pattern matching, or that musical repetition is structurally significant if at least fifty percent of a pattern is equivalent (*i.e.* if it is more similar than dissimilar), or that "dramatic disparities" between musical similarities and corresponding categories can be found in empirical studies, are critically re-examined with a view to clarifying the fundamental concept of similarity.

Keywords: similarity, categorisation, cue abstraction, musical surface.

Introduction

Without similarity, music would not be possible, would not exist. Similarity, and its counterpart dissimilarity or difference, enables a listener to break down the acoustic continuum into smaller constituent parts (such as elementary events, segments, groups, streams), and to make associations between these parts (repetitions, variations, oppositions, transitions and so on). Local similarities and discontinuities/dissimilarities give rise to elementary discrete entities such as notes, and allow the formation of musical streams (*e.g.* voices) and of segments within streams; larger scale pattern associations enable the emergence of the very musical materials that make music memorable and enjoyable such as themes and variations, rhythmic patterns, characteristic harmonic progressions, opposing thematic materials, transitions between dissimilar

sections, motion away from and back to a point of reference (*e.g.* tonal region). Similarity is at the very heart of music, not only because it allows connections and associations to be made between different musical entities, but because it enables the emergence of the core musical entities themselves (for instance, a musical theme is not an object/event that exists out there in the world — it exists primarily by virtue of self-reference via repetition and variation).

Without similarity and difference music would be merely an acoustic blur; essential things such as pitch, timbre and time itself would fade away and disappear. It would not be possible to say what precedes and what follows, what "moves" to what. Music would be imperceptible: "… if events were absolutely smooth, without beginning or end, and even without modification or 'perceptible' internal roughness, time would find itself abolished. It seems that the notions of separation, of bypassing, of difference, of discontinuity, which are strongly interrelated, are prerequisite to the notion of anteriority. In order for anteriority to exist, it is necessary to distinguish entities, which would then make it possible to 'go' from one to the other." (Xenakis, 1989, p. 87).

If similarity is omnipresent in musical understanding, does it not lose ultimately its explanatory power? A notion that is too general, that attempts to explain everything and is used in too many different ways can become unworkable, if not useless (Goodman, 1972) does not hesitate to discard similarity altogether). Still, it is almost impossible to study musical structure without implicit or explicit recourse to some aspect of musical similarity/dissimilarity. It is the aim of the paper to highlight various problematic aspects of similarity and to propose a theoretical framework that enables a workable and consistent use of this invaluable "tool" in the study of music.

The main assertions made in this paper is that similarity always depends on context (*i.e.* it is contextually defined), that it is inextricably bound to categorisation processes, and that, even, the features/properties of entities become more or less prominent in a given context for a specific categorisation task or for a similarity judgement. It is maintained that context affects perception of similarity, categorisation and, even, the perception of entities per se.

In the first part of the paper a theoretical discussion appears regarding the fundamental concepts of similarity, categorisation and feature/cue abstraction. Then, each of these concepts is examined in more detail, and a number of research hypotheses and results presented in the recent special issue of Musicæ Scientiæ on musical similarity (Discussion Forum 4A, 2007) are critically re-examined with a view to clarifying the fundamental concepts.

Similarity, Categorisation, Cue Abstraction

How similar is similar? When are two objects/events similar *enough* to be considered similar or different *enough* to be considered dissimilar? In what respects are two things similar? What event properties play a more significant role in judging similarity? What role does context play in measuring similarity and grouping together similar entities in categories?

In this section we will examine the fundamental notions of similarity, categorisation and cue abstraction, and relationships between them.

Similarity

Similarity is very often defined as partial identity, that is, two entities are similar if they share *some* properties, but not necessarily all (see extended discussion on the notions of identity and similarity in Cambouropoulos 2001). Pairs of entities may be compared, and one pair may be judged as being more similar than another, if its members share more common properties than the members of the other pair. Similarity between two entities may be calculated by simply counting the number of matches between their properties. Alternatively, similarity may be defined as a (weighted) function of the differences between all the pairs of properties these objects possess (see extended discussion on cognitively-based similarity metrics in Müllensiefen & Frieler, 2007).

The notion of similarity is meaningful only when the ways or respects in which two entities are similar are specified (see Medin, Goldstone and Gentner, 1993). This derives from the very definition of similarity given above, *i.e.* entities are similar in respect to certain properties they share. These shared properties may be specified implicitly or explicitly; in any case, they have to be there, otherwise similarity becomes meaningless.

Goodman (1972) argued that similarity is like motion in that it requires a frame of reference. Just like it is necessary to say that one object moves *in relation* to something else, so it is necessary to say that two things are similar *in relation* to some properties that are pertinent in some context. Without specifying the respects in which two things are similar similarity becomes unwieldy: everything is similar to everything else in some respect and everything is different from everything else in some other respect (*e.g.* a mouse and an airplane are dissimilar in that one is a living thing and the other not, but, they are similar in that they both have the capacity of independent motion).

Similarity often creates the illusion that it is "objective" or "fixed" due primarily to perceptual constraints. For instance, a violin and a cello would be considered similar by most humans, at least, in regards to appearance (they look similar); this similarity seems to be "objective" because it is constrained by our perceptual mechanisms. However, this similarity judgement is not without an underlying "in regards to" relationship. The fact that perceptual constraints are shared by most

humans does not render similarity "objective" — it simply provides an implicit shared framework that allows similarity to operate in everyday judgements. In analogy, humans commonly talk about motion of things assuming implicitly earth being the frame of reference — this does not mean that motion is absolute — motion is by definition a relative concept — it just happens that earth is a convenient frame of reference for most practical purposes and is implicit in most judgements of motion.

Deliège (2007, p. 10) quotes Henri Piéron's (editor of *Vocabulaire de la psychologie* — Piéron, 1951) definition of similarity as being the "psychological assertion of resemblance that can be assessed, either objectively, and is thus a property appertaining to a physical dimension, or subjectively and is an attribute of pairs of stimuli". This definition discriminates between objective similarity that relates to physical properties of entities and subjective similarity that arises depending on circumstantial relations between entities. Following the above discussion, I would claim that this definition is misleading. Similarity is by definition relational and, therefore, cannot be objective. Medin, Goldstone and Gentner (1993) refute the claim that similarity is "restricted to hard-wired perceptual processes" and give psychological evidence that supports "the idea that respects [for similarity] are determined by processes internal to comparisons." (p. 254). They maintain that "…the problem with viewing similarity as fixed is that it leads researchers to ignore the processing side of similarity. Our studies, as well as others, show similarity to be dynamic and context dependent." (p. 271)

Categorisation

A category is commonly taken "to refer to a class or set of entities (they could, for example, be objects, actions, states, qualities) which are grouped together on the basis of some criterion or rule" (Hampton and Dubois, 1993, p. 13). The "modern" view of categorization gives emphasis to gradedness, typicality and fuzziness of boundaries rather than strict necessary and sufficient conditions of membership that characterise the "classical" view.

A commonly encountered hypothesis on which many categorization models are grounded is that categorization is strongly associated with the notion of similarity. The idea that similar entities tend to be grouped together into categories is intuitive, and is supported by data collected by Rosch and colleagues in the 70s (*e.g.* Rosch 1975, 1978). Such categories reflect "natural" partitions of the world and are also informative and predictive (*i.e.* if we know in which category an object belongs to we can infer what attributes it likely has).

Criticism on the similarity-based view grew in the 80's (apart from Goodman's renown critique). It was maintained that similarity was not powerful enough to account for human categorisations. For instance, the fact that listeners may consider a long silence (such as Cage's "4.33") or a recording of natural sounds in a concert as a musical work, has to do with a complex framework of ideas and theories of what humans think music is and the kinds of things that may happen in musical concerts

rather than features of the sound material per se. Murphy and Medin (1985) have been strong proponents of the so called theory-based view on categorisation.

More recently, however, there has been a second-wave of similarity-based approaches that address shortcomings of the earlier similarity-based and theory based approaches (see collection of papers in Hahn & Ramscar, 2001). This has been accomplished primarily by allowing more sophisticated views of similarity that consider structural representations (rather than spatial or featural representations) or take in account the processing history of perceived similarity between objects. For instance, Heit's exemplar based model (2001) is a similarity driven model that accounts for the effects of prior knowledge and theories by allowing appropriate exemplars — this model shows that similarity-based accounts can be very powerful and useful, and can account for many things that were previously thought to be out of reach for similarity-based approaches. A richer account for similarity can form the basis for sophisticated categorisation tasks. For instance, "4.33" by John Cage is similar to other musical works in that it is "performed" by a musician in a musical concert where listeners expect to listen to musical works — music is a culturally-defined category that is meaningful within a certain context (there exists no absolute definition of what counts as a music).

Cambouropoulos (1998, 2001) has proposed a computational model where the notions of categorisation, similarity, and the representation of entities/properties are strongly interrelated. It is not simply the case that one starts with an accurate description of entities and properties, then finds pairwise similarities between them and, finally, groups the most similar ones together into categories. It seems more plausible that as humans organize their knowledge of the world, they alter their representations of entities concurrently with emerging categorizations and similarity judgments. Different contexts may render different properties of objects/events more diagnostic concurrently with giving rise to certain similarity relationships and categorisations. If context changes, it affects similarity, categorisation and the way the objects/events themselves are perceived.

Cues and Properties

Establishing similarity and category membership relations requires accepting an appropriate level of abstraction at which such relations can be judged. Such a level of abstraction is not absolute, but depends on the nature of the data and the context. If one is interested in cognitively pertinent similarities and categories, then perceptual constraints can be taken into account in establishing cognitively plausible abstractions and measures of similarity. However, even such perceptually pertinent abstractions are not absolute, as they depend on context, *i.e.*, from among different perceptually plausible representations one might be preferred over others in a given context (*e.g.* a saxophone appears to be more perceptually similar to other brass instruments, however, in the context of musical organology it is considered more similar to woodwind instruments).

Deliège has introduced the notion of a *cue*: "Cues… are brief but meaningful and significant structures, which stand out from the sound background." (Deliège, 2007, p. 13) A cue not only "attracts the listener's attention" to a certain musical sequence but also "'summarises' the sequences from which it arose into a succinct representation, a sort of label, that reduces the memory load required to internalise the whole structure." (Deliège, 2001, p. 238) No strict definition is given to *cue* by Deliège, but it is suggested that a cue is an abstraction/reduction of a musical sequence that contains significant musical information and is therefore highlighted in perception.

Cues emerge within a specific musical context — "Cues are specific to each work…" (Deliège, 2007, p. 13). Additionally, cues are strongly related to similarity and categorisation. "A cue always acts in concert with the… principles of similarity and difference — in the context of which a cue becomes a cornerstone in the process of musical categorisation: its primary dynamic function in music cognition." (Deliège, 2001, p. 238)

Deliège states that "a *cue* is a salient element at the musical surface." (2001, p. 237). This begs the question of what is an appropriate level of abstraction for the musical surface (see next section). If we can define the musical surface, then, we can attempt to describe in more detail what a cue is.

In this section, it has been maintained that similarity, categorisation and cue abstraction are strongly interrelated. Each of these cannot be defined in isolation but only in relation to the others. There are still many open questions as to how cues are abstracted, what exactly role context plays, how similarity and categorisation affect each other and so on. Perhaps similarity cannot explain all aspects of categorisation, but gaining a better understanding of similarity processes and object/event representations can assist understanding categorisation.

It is suggested in this paper that, the least a researcher can do, is explicate the meaning of the term similarity, its use in relation to categorisation, and the representation on which it is applied. This way, it will be clear what one means by the powerful yet often controversial notions of similarity and categorisation in the context of certain research.

In the next sections, we will examine the hypotheses and results of a number of studies reported in the Special Issue on Similarity Perception in Listening to Music (Musicæ Scientiæ, Discussion Forum 4A, 2007).

Finding an Appropriate Level of Abstraction: The Musical Surface

The acoustic continuum is broken down into elementary events by a listener. "The identification of each event is an endproduct of the ongoing perceiving process. Without rules to segregate elements, events could not be perceived." (Handel, 1989, p. 217). The elementary events perceived as constituent units of an acoustic

continuum are further grouped together into elementary categories. Research in categorical perception has investigated especially various facets of musical pitch and time perception — see overviews and discussion in (Dowling and Harwood, 1986; Handel, 1989). It is generally admitted that categorical perception depends not only on the physical acoustic source or on the perceptual sensitivities of the human auditory system but on contextual effects and background knowledge as well (Handel, 1989).

Jackendoff (1987) describes the *musical surface* as being the "lowest level of representation that has musical significance" (p. 219). In relation to tonal music he states: "… the *musical surface*, encodes the music as discrete pitch-events (notes and chords), each with a specific duration and pitch (or combination of pitches, if a chord). Standard musical notation represents the pitch-events of the musical surface by means of symbols for discrete pitch and duration;…" (p. 218).

Wiggins (2007) suggests that "there is a very natural point at which to draw a line between perception and cognition, which also happens to be the *musical surface*…: the level of musical notes as heard" (p. 325), and proposes that "a piano roll is actually a more accurate approximation" (p. 326) of the musical surface than the musical score.

Most computational, and, often, even perceptual, models assume a musical surface that appears at the level of individual notes in the musical score (or in a piano roll). Categorically perceived notes (discrete pitches and durations) are considered to be the most primitive musical entities that form the musical surface.

But, is the note level the lowest level of representation that has musical significance? There is evidence that things such as melodic and harmonic pitch intervals, chords or larger configurations such as tone clusters, tremolos, trills, glissandi are commonly perceived by listeners as wholes rather than combinations of atomic lower-level components. For example, especially for pitch, it has been suggested that the majority of listeners, for whom musical pitch is relative, perceive pitch intervals categorically prior to individual pitches (Dowling and Harwood, 1986; Handel, 1989). Tenney suggests that larger sound complexes such as tone-clusters or other dense chords "cannot usually be analysed by the ear into constituent tones, and [he suggests] are not intended to be analysed." (Tenney, 1961:6) — see also (Cook, 1990); even simpler triadic chords may be perceived as elementary chord types — or even tonal chord function types — before being analysed into their constituent tones and intervals. A glissando is also perceived and can be represented as a single entity (see Bregman 1990, p. 644) with start-pitch and end-pitch, slope of transition, duration and intensity (a linear transition between the two pitches may be implied as a default).

Listeners break down the acoustic continuum into *musical streams* prior to identifying individual notes. They perceive streams of musical events such as streams of notes (*e.g.*, melodic streams) or streams of chords (*e.g.*, accompaniment). The principles of auditory stream segregation (McAdams & Bregman, 1979; Bregman

1990) enable a listener to integrate or fuse co-modulating components (*e.g.*, partials or notes moving in parallel) into coherent sequences due to a number of factors. Such streams can be represented in different ways; for instance, a melodic stream may be represented as a sequence of diatonic intervals or chromas, or a chord stream as a progression of harmonies. In most cases, individual notes are sub-surface elements — listeners do not perceive musical sequences as sequences of (concurrent) individual pitches.

It is, herein, suggested that the musical surface comprises of (complex) musical events perceived as wholes within coherent musical streams — the musical surface is not merely a sequence of atomic note events. In the case of monophonic sequences, the importance of a richer higher-level musical surface is not so obvious, as it is relatively straight-forward to derive higher-level representations from a mere string on atomic notes; for instance, a string of pitches can be readily converted into a sequence of chromatic intervals or diatonic intervals or steps & leaps or gross contour (ups and downs) and so on. Of course, which of these will be chosen as an appropriate level of representation is critical for a given task (for instance, pattern matching on a string of absolute pitches cannot reveal transposed patterns — see more on representation issues in pattern matching in Cambouropoulos, Crawford & Iliopoulos, 2001).

In the case of non-monophonic music, however, defining a musical surface in terms of more complex musical events organised in musical streams, can play a significant role in extracting further musically pertinent information. In this case, it is not trivial to derive higher-level representations, such as musical streams, from a sequence of mere individual notes — see more on voice/stream separation modelling in (Karydis *et al.*, 2007; Cambouropoulos 2008).

Lemström and Pienimaki (2007) compare edit distance string based approaches to two-dimensional geometric frameworks in content-based retrieval of symbolically encoded polyphonic music. They argue that string based algorithms are efficient, and adequate for linearly ordered sequences of events such as monophonic melodies, but are severely handicapped when faced with polyphonic music. For polyphonic music, they argue that a two-dimensional geometric framework, employing computational geometric techniques on a piano-roll-like representation, "is intuitive, allows natural handling of polyphony and is easily visualised." (p. 142) — attempts to introduce the geometric approach to musical retrieval includes work by Clause *et al.* (2000) and Meredith *et al.* (2002). Lemström and Pienimaki (2007) conclude their study stating that "although the edit distance framework provide somewhat more efficient and robust tools, the clearly more effective representation for polyphonic music of the geometric framework compensates that. Moreover, as the edit distance framework faces severe problems in combining polyphony and transposition invariance, which both are very important and intrinsic features of the musical task at hand, we conclude that the geometric framework is the choice for successful polyphonic content based music retrieval." (p. 148)

It is herein argued that the above argument is problematic and misleading. The main objection is that the geometric piano-roll-like representation suggested is not cognitively relevant, *i.e.*, it is not the musical surface. A listener does not extract a geometric representation from the acoustic input; a listener organises the acoustic continuum into musical streams (Bregman, 1990) and encodes the music into a representation that is closer to strings, *i.e.* sequences of musical events (such as sequence of notes or chords). For instance, the extract from the waltz in Figure 1 is perceived as two streams, namely a melodic sequence and harmonic accompaniment, not as a two-dimensional piano-roll representation of individual notes.

Let's suppose that a query, such as the one illustrated in Figure 1 is posed within a geometric framework; an exact match is found as indicated in Figure 1 (middle). However, this match is cognitively irrelevant as the recognised pattern is distributed across two different streams and, additionally, the lower stream is not a melodic stream at all — it is well-established that listeners have great difficulty in perceiving patterns distributed across different perceptual streams (Bregman, 1990). Essentially, a geometric-based algorithm may be looking for patterns in the wrong place. It is musically more meaningful to apply, firstly, a stream segregation algorithm on the two-dimensional representation and, then, to search for melodic patterns *within* melodic streams. Stream segregation is a rather complex problem (see discussion and proposed algorithm in Karydis *et al.*, 2007, and Cambouropoulos, 2008), but is necessary to extract a cognitively plausible musical surface. Then, pattern matching algorithms can be applied to the separate streams more effectively as, on the one hand, efficient string matching algorithms can be used (streams can often be represented by strings of symbols) and, on the other, the search space can be significantly reduced by disallowing certain matches (*e.g.* across streams) and by omitting certain streams altogether (*e.g.*, in the waltz of Figure 1 the accompanimental harmonic stream can be disregarded in melodic pattern matching tasks).

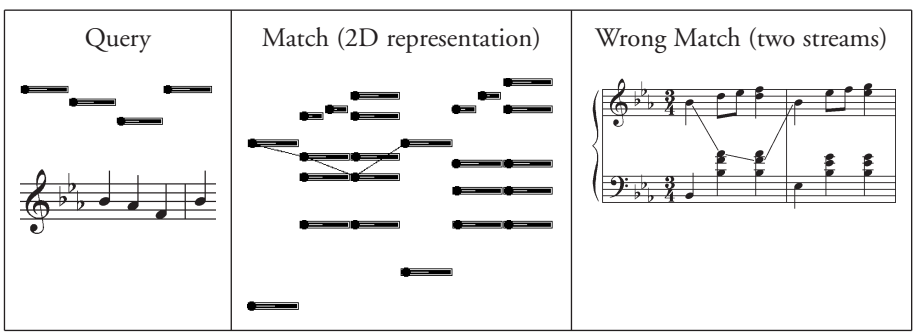

Figure 1.

The query melodic pattern (left) yields an exact match in the two-dimensional piano-roll representation (middle) — this match, however, is cognitively irrelevant as it is distributed across different musical streams (right).

A musical surface that is close to the way listeners perceive sequences of musical events (*i.e.* musical streams), can enhance higher-level processing and can enable the extraction of musically pertinent structures in a cognitively plausible manner. It is argued that the musical surface, *i.e.* the level of representation that is musically meaningful, lies above the individual note level. Research from auditory scene analysis and, more specifically, from musical stream segregation, can inform such a quest for an adequate level of musical representation. The present discussion aims at highlighting various facets of the problem and giving hints as to the kind of things that might be considered as part of the musical surface, rather than providing a definitive solution.

Setting a Similarity Threshold

Ahlbäck (2004) brings together, with amazing skill, diverse aspects of musical understanding in an integrated model of melody cognition. In his recent paper he "focuses on melodic similarity as a cue for melodic segmentation in metrical music." (Ahlbäck, 2007, p. 239). Ahlbäck is adamant in that "similarity is contextual and relative.… categorization by similarity is related to the perspective of the observer." (2007, p. 236). The author strongly links similarity with categorisation and emphasises the contextual nature of similarity: "… if we are looking for similarity, *i.e.*, forming a category by sameness, there is almost no limits as to how general similarity can be considered significant. It depends on the diversity within the sample we are looking at; the more diverse the context, the more general similarity may be recognized." (p. 237).

I will focus on one small detail of Ahlbäck's model that seems to contradict his view on the contextual nature of similarity outlined above. In discussing "how much needs to be similar in order for it to be structurally significant" (Ahlbäck, 2007, p. 253), the author suggests that "the basic threshold for a significant temporal structure defined by sequence repetition…, is interpreted in the model as a situation where the sequences are not more dissimilar than similar, in terms of structural equivalence at corresponding temporal positions. This implies a basic repetition to be regarded as structurally significant if the equivalent part between the original and the repetition is at least as great as the difference, which means a basic threshold of 50% similarity." (p. 253) In order to illustrate the 50% threshold, Ahlbäck gives an example in which the repeating part of melodic segments A1 and A2 is smaller than 50% and, thus, this sequence is not significant — A3 is a significant sequence in this example as similarity is 100% (see Figure 2).

In the example of Figure 2, Ahlbäck claims that the structure A1-A2 is not significant as only 40% of sequence A1 is common with sequence A2. I believe this is not true in general. Of course, an empirical study would be necessary to justify this claim, but a few indications that structure A1-A2 may be significant are: the first

four-note repetition is more "characteristic" in terms of intervals and more appropriate to initiate a pattern, this initial repetition is identical in terms of absolute pitch, and the second half of the two sequences is also common (the two sequences only differ in one interval that has same size but different direction). A similar situation appears in the theme from Schubert's *Symphony in B minor, D.759, "Unfinished"* (Figure 3). In analogy to the example in Figure 2, sequence A3-A4 should be preferred over sequence A1-A2 as the latter is non-significant (20% similarity). However, it is clear that the structure of this theme supports the A1-A2 structure (actually, my own *Pattern Boundary Detection Model* described in Cambouropoulos 2006 fails to detect the beginning of the repetition at the beginning of A2).

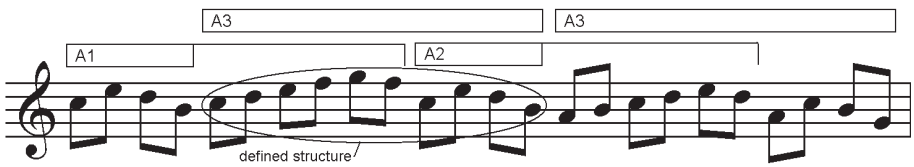

Figure 2.
Illustration of Ahlbäck's 50% threshold — "only 40% of A1 is repeated in A2, which makes the A1-A2 sequence non-significant. Thus, the overlapping sequence repetition A3-A3 will be significant." (Ahlbäck, 2007, p. 254, figure 14a).

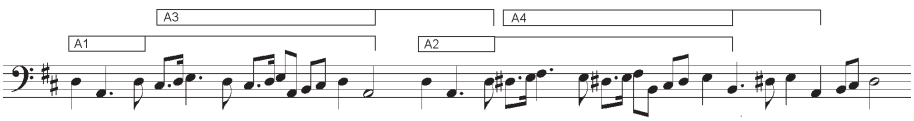

Figure 3.
Theme from Schubert's Symphony in B minor, D.759, "Unfinished". *In analogy to the example in Figure 2, sequence A3-A4 should be preferred over sequence A1-A2 as the latter is non-significant (20% similarity). This result is counter-intuitive as a boundary is placed at the beginning of A2.*

The problem with Ahlbäck's similarity threshold is not the threshold itself (a certain similarity threshold may give intuitive similarity ratings in a particular music context), but its purported "universality". Ahlbäck claims that the 50% similarity threshold is justified because it "guarantees" that two sequences "are not more dissimilar than similar" (Ahlbäck, 2007, p. 253). This means that there exists a "natural" threshold of similarity above which two things are more similar than dissimilar, and below which they are more dissimilar than similar. This claim contradicts the earlier claim of the contextual and relative nature of similarity. Two entities may share few common features and still be considered very similar if they

appear in a very diverse context, or two entities with many common features may be dissimilar if encountered within a very homogeneous context (*e.g.*, two musical patterns may be very different in the context of some minimal music or the same two patterns may be very similar in the context of other contemporary music). If there was a natural split point between similarity and dissimilarity (*e.g.* 50%), that is, if it was possible to say when two entities are more similar than dissimilar, then the alleged relative and contextual nature of similarity would find itself abolished.

Similarity and Categorisation in Conflict?

Ziv and Eitan (2007) report an experiment that explores the relationship between similarity and categorisation judgements in musical contexts. Ziv and Eitan employed musical materials extracted from piano pieces by Beethoven and Schoenberg, and participants were asked to rate for each piece, to what degree extracts "belong" to each of the two main themes (categorisation task). The same materials were employed in an earlier empirical study by Lamont and Dibben (2001), in which participants rated the similarity of pairs of musical extracts taken from the same piece (similarity judgement task). Ziv and Eitan compare the results of the two experiments and observe a number of potential differences/dissociations between similarity and categorisation tasks on the same data.

According to Ziv and Eitan, a comparison of the two experiments' results shows that "categorizations concurred with similarity ratings for Beethoven, and differed for Schoenberg.… in both pieces categorizations, unlike similarity ratings, were negatively correlated… This effect resulted in some dramatic disparities between similarities and corresponding categorizations." (Ziv & Eitan, 2007, p. 99).

How does the dissociation observed by Ziv and Eitan relate to the assertion made in this paper that similarity and categorisation are always strongly linked together? Isn't this dissociation refuting the above claim?

My opinion is that Ziv and Eitan are comparing things that are not directly comparable. Of course, the same musical materials are used, but they are conducting a different experiment (from Lamont and Dibben's experiment) that sets a different context in which categorisation judgements are investigated. Assuming that similarity is by definition context-dependent (as suggested above), changing the experimental set-up changes in essence the context and, therefore, it is expected that there will be differences in similarity ratings and categorisation judgements. If one wants to be certain that context does not change, one could simply replicate the L&D experiment, and change only the question posed to listeners — *i.e.*, instead of asking listeners to rate the degree of similarity of pairs of extracts, one could ask listeners to rate the degree of "belongingness together" or "belongingness to the same category" for the same pairs of extracts.

In Lamont and Dibben's experiment, participants listen to the whole piano piece

once (Beethoven or Schoenberg) and, then, rate the similarity between all the pair extracts that can be formed out of 9 extracts. In Ziv and Eitan's experiments, participants, first, listen to the whole piece once (Beethoven or Schoenberg), then a brief explanation of the term "musical theme" is read to them, then the two principal themes of the piece are presented three times each, then the whole piece is played again, then the two themes are played again twice each, and, finally, the participants rate the degree to which each of the 9 extracts "belongs" to the two themes (A and B).

It is clear, that the two listening contexts are not the same. In L&D's experiment participants listen to each piece of music only once and we can assume that there is some preliminary implicit abstraction of potential cues, thematic materials and musical patterns; however, a single listening of a piece does not give the opportunity to listeners to establish more sophisticated structural relationships between the musical materials. In contrast, in the Z&E experiment, listeners are "guided" (if not "forced") to abstract certain cues, and to make associations of musical materials to the given prototypes, that is, listeners are assisted to obtain a more elaborate understanding of a specific structural organisation of the pieces. Judging similarity or categorisation relationships in the two different settings, is expected, according to the earlier discussion, to yield different results, simply because different "respects" of similarity/categorisation may be more pertinent in the different contexts.

It should be noted that the dissociations found in Z&E's study between similarity and categorisation judgements in the case of the Schoenberg piano piece, may be additionally due to the fact that the prototypes (two primary themes) presented to listeners are determined not perceptually but in music theoretic terms (*i.e.* by music theorists). It is possible, that the themes established by the composer and analysed by music theorists may be different from the thematic categories (and corresponding prototypes) abstracted by listeners. This may be especially true for Schoenberg's atonal piece where thematic materials rely on the twelve-tone row — there exists significant empirical evidence against the perceptibility of the tone row (*e.g.* Bruner, 1984; Krumhansl *et al.*, 1987; Imberty, 1993). In other words, in L&D's experiment listeners may be abstracting different cues and prototypes in the piece by Schoenberg than in the Z&E experiment where listeners are probed to organise the structure of the piece around the two musicologically established twelve-tone themes.

Ziv and Eitan (2007) explain the discrepancies between similarity and categorisation hypothesising that "categorizations will emphasize deeper-level structural features (like harmony, voice-leading, or intervallic structures), while similarity ratings will emphasize surface perceptual features (like dynamics or textural density)" (p. 106). Essentially, they are supporting that similarity applies to surface features of the music, whereas categorisation takes in account deeper structural elements. This view seems to endorse a very restricting and narrow view on similarity (*i.e.* similarity applying to appearance or surface features). However, as suggested above, there is no reason to restrict similarity in such a way. Two musical extracts can be similar in terms

of dynamics or texture, or indeed, in terms of pitch intervals, harmonic functions, contour, rhythm and so on. Even if we assume that categorization relies of "deeper-level structural features", as Ziv and Eitan suggest, then it is *similarity* in respect to these features that allows categorisation (similar extracts in terms of "harmony, voice-leading, or intervallic structures" are grouped together into categories). On the other hand, similar musical extracts in terms of "surface perceptual features" can be organised into categories (the initial categorisation may be replaced by more sophisticated categories after repeated hearing that enables extraction of "deeper-level" features).

Conclusions

In this paper the notion of similarity has been discussed in relation to the notions of categorisation and cue abstraction. An attempt was made to present different and, even, opposing views on these notions that have generated heated debate in recent years. The aim of the paper was to show the necessity for researchers to state clearly what view and what definitions of similarity/categorisation they endorse in order to avoid unnecessary conflict and confusion.

Re-examination of a number of recent musical research studies has unveiled potential problems in the way the concepts of similarity and categorisation are understood. A clearer description of these notions can lead to a more consistent interpretation of empirical results and better design of computational models.

It has been maintained, in this paper, that similarity, categorisation and cue abstraction are strongly inter-related and contextually-defined. This is not necessarily *the* correct way to define these concepts, but, I believe, it is intuitive in many ways and, also, appropriate for the study of music, as musical meaning is to a large extent self-referential and dependent on the context of one or more works. Music seems to be a privileged domain for the study of similarity relations, category formation and cue abstraction, and, hopefully, further research will contribute to a better understanding of these indispensable cognitive capacities and their role in music cognition.

Address for correspondence:
Emilios Cambouropoulos
Department of Music Studies
Aristotle University of Thessaloniki
54124, Thessaloniki
Greece
e-mail: emilios@mus.auth.gr

• References

Ahlbäck, S. (2004). Melody Beyond Notes: A Study of Melody Cognition. PhD Thesis, Department of Musicology, Götenberg University, Sweden.

Ahlbäck, S. (2007). Melodic Similarity as a Determinant of Melody Structure. *Musicæ Scientiæ, Discussion Forum 4A*, 235-80.

Bregman, A. S. (1990). *Auditory Scene Analysis*. The MIT Press, Cambridge (Ma).

Bruner, C.L. (1984). The perception of contemporary pitch structures. *Music Perception*, *2*, 25-9.

Cambouropoulos, E. (2008). Voice and Stream: Perceptual and Computational Modelling of Voice Separation. *Music Perception 26(1)*, 75-94.

Cambouropoulos E. (2006). Musical Parallelism and Melodic Segmentation: A Computational Approach. *Music Perception 23(3)*, 249-69.

Cambouropoulos E. (2001). Melodic Cue Abstraction, Similarity and Category Formation: A Formal Model. *Music Perception 18(3)*, 347-70.

Cambouropoulos, E., Crawford, T. and Iliopoulos, C.S. (2001). Pattern Processing in Melodic Sequences: Challenges, Caveats and Prospects. *Computers and the Humanities, 35(1)*, 9-21.

Cambouropoulos E. and Widmer, G. (2000). Automated Motivic Analysis via Melodic Clustering. *Journal of New Music Research 29(4)*, 347-70.

Cambouropoulos, E. (1998). *Towards a General Computational Theory of Musical Structure.* PhD Thesis, Faculty of Music and Department of Artificial Intelligence, University of Edinburgh. http://users.auth.gr/~emilios/englishpage/phd.html

Clause, M., Engelbrecht, R., Meyer, D., & Schmitz, J. (2000). PROMS: A web-based tool for searching in polyphonic music. In *Proceedings of the First International Symposium on Music Information Retrieval (ISMIR 2000)*. Plymouth, Massachusetts.

Cook, N. (1990). *Music, Imagination and Culture.* Oxford University Press, Oxford.

Deliège, I. (2001). Introduction: Similarity Perception ↔ Categorisation ↔ Cue Abstraction. *Music Perception*, *18(3)*, 233-43.

Deliège, I. (2007). Similarity Relations in Listening to Music: How do they Come into Play? *Musicæ Scientiæ*, *Discussion Forum 4A*, 9-37.

Dowling, W.J. and Harwood D.L. (1986). *Music Cognition.* Academic Press, New York.

Goodman, N. (1972). Seven Strictures on Similarity. In *Project and Problems*, by N. Goodman, The Bobbs-Merrill Company, Inc., New York.

Hahn, U. and Ramscar, M. (2001). *Similarity and Categorization*, Oxford University Press, Oxford.

Handel, S. (1989). *Listening. An Introduction to the Perception of Auditory Events.* The MIT Press, Cambridge (Ma).

Heit, E. (2001). Background Knowledge and Models of Categorisation. In *Similarity and Categorization*, Hahn, U. & Ramscar, M. (eds), Oxford University Press, Oxford.

Imberty, M. (1993). How Do We Perceive Atonal music? Suggestions for a Theoretical Approach. *Contemporary Music Review*, 9(1&2), 325-37.

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

Karydis, I., Nanopoulos, A., Papadopoulos, A.N. & Cambouropoulos, E., (2007). VISA: The

Voice Integration/Segregation Algorithm. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR07)*, Vienna, Austria.

Krumhansl, C., Sandell, G.J. & Sergeant, D.C. (1987). Tone hierarchies and mirror forms in serial music. *Music Perception*, *5*, 31-78.

Lamont, A. and Dibben, N. (2001). Motivic Structure and the Perception of Similarity. *Music Perception*, *18(3)*, 245-74.

Lemström, K. & Pienimäki, A. (2007). On Comparing Distance and Geometric Frameworks in Content-based Retrieval of Symbolically Encoded Polyphonic Music. *Musicæ Scientiæ, Discussion Forum 4A*, 135-52.

McAdams, S. & Bregman, A. (1979). Hearing Musical Streams. *Computer Music Journal*, *3(4)*, 26-44.

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

Müllensiefen, D. & Frieler, K. (2007). Modelling Experts' Notions of Melodic Similarity. *Musicæ Scientiæ, Discussion Forum 4A*, 183-210.

Murphy, G.L. and Medin, D.L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review*, *92(3)*, 289-316.

Piéron, H. (1951). *Vocabulaire de la Psychologie*. Presses Universitaires de France, Paris.

Rosch, E. (1978). Principles of Categorisation. In *Cognition and Categorisation.* E. Rosch and B.B. Lloyd (eds), Lawrence Erlbaum Associates, London.

Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, *104*, 192-232.

Tenney, J. (1961). *Meta+Hodos.* Frog Peak Music, Oakland (Ca).

Wiggins, G. (2007). Models of Musical Similarity. *Musicæ Scientiæ, Discussion Forum 4A*, 315-38.

Xenakis, I. (1989). Concerning Time. *Perspectives of New Music*, *27(1)*, 84-92.

Ziv, N. & Eitan, Z. (2007). Themes as Prototypes: Similarity Judgements and Categorization Tasks in Musical Contexts. *Musicæ Scientiæ, Discussion Forum 4A*, 99-133.

• ¿Cuán similar es lo similar?

En la primera parte del artículo se presenta un debate teórico sobre el concepto fundamental de la similitud y su relación con la abstracción y la categorización de índices. Se sostiene que la similitud es, por definición, dependiente del contexto y está estrechamente relacionada con la abstracción y la categorización de índices. Se insiste en la determinación de la "superficie musical", que puede actuar como el nivel más bajo de abstracción musicalmente pertinente en el cual puede ser medida la similitud entre entidades musicales. Cada uno de estos conceptos es examinado a continuación con mayor detalle respecto a algunos estudios presentados en un reciente número especial de *Musicæ Scientiæ* dedicado a la similitud musical (Foro de Discusión 4A, 2007). Las opiniones según las cuales el modo más adecuado de representación de una fórmula polifónica es en forma de rollo perforado de piano mecánico, o aquéllas según los cuales la repetición musical es estructuralmente significativa si al menos el cincuenta por ciento de dos piezas musicales son equivalentes (esto es, si son más similares que disimilares), o aquéllas otras según las cuales en los estudios empíricos pueden encontrarse "disparidades fuertes" entre similitudes musicales y categorías correspondientes, son reexaminadas de forma crítica, desde el punto de vista de clarificar el concepto fundamental de similitud.

• Quanto simile è la similarità?

Nella prima parte del saggio è presentata un'indagine teorica sul concetto fondamentale di similarità e sul suo rapporto con la *cue abstraction* e con la categorizzazione. Si sostiene che la similarità dipenda per definizione dal contesto e che sia fortemente correlata alla *cue abstraction* e alla categorizzazione. L'enfasi è posta sulla determinazione della "superficie musicale" che può rappresentare il livello inferiore di astrazione musicalmente pertinente su cui è possibile misurare la similarità tra diverse entità musicali. Ciascuno di questi concetti è, poi, esaminato più dettagliatamente rispetto ad alcuni studi presentati nella recente pubblicazione di *Musicæ Scientiæ* sulla similarità in musica (Forum di discussione 4A, 2007). Le opinioni secondo cui la scelta più appropriata di rappresentazione di una formula polifonica è quella geometrica, quale la rappresentazione *piano roll*, oppure quelle secondo cui la ripetizione musicale è significativa dal punto di vista strutturale solo se almeno il 50% di una formula risulta equivalente (ossia, se è più simile che dissimile), o quelle secondo cui negli studi empirici possono ritrovarsi "forti disparità" tra similarità musicali e le categorie corrispondenti sono riesaminate criticamente con l'obiettivo di chiarire il concetto fondamentale di similarità.

• Jusqu'où va la similarité ?

La première partie de cet article est une discussion théorique sur le concept fondamental de similarité et sa relation avec l'abstraction et la catégorisation

d'indices. On soutient que la similarité est, par définition, dépendante du contexte, et en interrelation étroite avec l'abstraction et la catégorisation d'indices. On insiste sur la détermination de la « surface musicale », qui peut agir comme le niveau le plus bas d'abstraction musicalement pertinent et sur lequel on peut mesurer la similarité entre entités musicales. Chacun de ces concepts est ensuite examiné de façon plus approfondie, en lien avec un certain nombre de recherches déjà présentées dans un récent numéro spécial de *Musicæ Scientiæ* consacré à la similarité musicale (Discussion Forum 4A, 2007). Les points de vue selon lesquels une représentation géométrique en forme de rouleau perforé de piano mécanique est le choix le plus approprié pour un appariement de modèle polyphonique, ou ceux selon lesquels la répétition musicale est structurellement significative si au moins cinquante pourcent de deux pièces musicales sont équivalentes (*i.e.* si elles sont plus similaires que dissemblables) ; ou encore les points de vue selon lesquels des « disparités frappantes » entre similarités musicales et catégories correspondantes peuvent être trouvées dans des études empiriques, sont réexaminés de façon critique, dans l'optique de clarifier le concept fondamental de similarité.

• Wie ähnlich ist ähnlich?

Der erste Teil dieses Artikels stellt eine theoretische Diskussion zum fundamentalen Konzept der Ähnlichkeit und Beziehungen zu Abstraktion und Kategorisierung von Hinweisreizen dar. Ähnlichkeit scheint per Definition kontextabhängig und stark durch Interrelationen zu Abstraktion und Kategorisierung von Hinweisreizen geprägt zu sein. Besonderes Gewicht wird auf die Bestimmung der "musikalischen Oberflächenstruktur" gelegt, wo auf einem musikalisch relevanten, niedrigen Abstraktionsgrad Ähnlichkeit zwischen musikalischen Einheiten gemessen werden kann. Daraufhin wird jedes dieser Konzepte genauer untersucht unter Berücksichtigung von Forschungsarbeiten, die in der kürzlich erschienen Sonderausgabe von *Musicæ Scientiæ* über musikalische Ähnlichkeit vorgelegt wurden (Diskussionsforum 4A, 2007). Dabei werden Behauptungen kritisch überprüft, dass eine geometrische, Klavierrollen-ähnliche Repräsentation die beste Form der polyphonen Mustererkennung sei, oder dass musikalische Wiederholungen strukturell signifikant seien, sofern wenigstens fünfzig Prozent der Muster äquivalent sind (das heißt, wenn sie stärker ähnlich als unähnlich sind), oder dass „dramatische Disparitäten" zwischen musikalischen Ähnlichkeiten und korrespondierenden Kategorien in empirischen Studien gefunden werden können. Dies soll zu einer Klärung des fundamentalen Ähnlichkeitskonzepts beitragen.