MUSICÆ
SCIENTIÆ

Special issue 2010

Understanding musical structure and form: papers in honour of Irène Deliège

Contents – Sommaire – Inhalt

See the escom web site :
www.escom.org <http ://www.escom.org>

International Standard Serial Number
ISSN 1029-8649

MUSICÆ SCIENTIÆ is abstracted and indexed in the Arts & Humanities Citation Index(®) (A&HCI), ISI Alerting Services, Current Contents(®)/Arts & Humanities, Social Sciences Citation Index, Current Contents/Social & Behavioral Sciences, APA's Psychological Abstracts, PsycINFO, The Music Index, The International Repertoire for Music Literature (RILM), Elsevier Bibliographic Databases.

TRANSLATORS PANEL

Spanish:	Rafael Ferrer Flores, University of Jyväskylä (Finland)
Italian:	Elvira Brattico, University of Jyväskylä (Finland)
French:	Olivier Lartillot, University of Jyväskylä (Finland)
German:	Reinhard Kopiez, University of Hannover (Germany)
English:	Geraint Wiggins (French - English translations), University of London (UK)

University of Jyväskylä, Department of Music
P.O. Box 35 (M)
40014 University of Jyväskylä, Finland.

Graphic design: Agence À3, Herstal (Belgium)
Cover illustration: Claude Ledoux
Page layout: Société Interligne, Liège
Printed by P.U.B., Presses Universitaires de Bruxelles

Musicæ Scientiæ
Special issue 2010, 131-147

The Musical Surface: Challenging Basic Assumptions

EMILIOS CAMBOUROPOULOS
Department of Music Studies, Aristotle University of Thessaloniki

• Abstract

This paper addresses problems and misconceptions pertaining to the notion of the *musical surface*, a notion that is commonly thought to be relatively straight-forward and is often taken as a given in computational and cognitive research. It is suggested that the musical surface is comprised of (complex) musical events perceived as wholes within coherent musical streams - the musical surface is not merely an unstructured sequence of atomic note events, such as score notes or a piano-roll representation. Additionally, it is maintained that the emergence of the musical surface involves rather complex mechanisms that require, not only multi-pitch extraction from the acoustic signal, but, the employment of cognitive processes such as beat-tracking, metre induction, chord identification and stream/voice separation. Such processes do not come into play *after* the surface has been formed, but are, rather, an integral part of the formation of the musical surface *per se*.

Keywords: musical surface, automatic transcription, beat-tracking, stream separation

Introduction

Research in music cognition has studied the emergence of different types of musical structures when listening to music or music-like stimuli. Irène Deliège has presented over the years a compelling theoretical framework, supported by a series of empirical studies, that explores how a listener may organize a musical surface into musically pertinent structures forming a mental representation. Cue abstraction, similarity and categorization are key concepts in her inquiry (see, for instance, Deliège 1987, 1996, 1997, 2001, 2007). Her work is of special importance as she has taken an ecologically valid road to studying music perception by means of employing pieces of real music in her experiments and by examining responses not only of music experts but of everyday listeners as well. My own research explores similar topics from a computational perspective - a direct comparison with Deliège's empirical work can be found in (Cambouropoulos 2001). In the current paper, however, I will focus on an issue that has not been extensively discussed in our work.

Much research in computational musicology and music cognition assumes the 'elementary' concept of the *musical surface, i.e.,* a minimal discrete representation of the musical sound continuum in terms of notes (each note described in terms of pitch, onset, duration, and possibly dynamic markings and timbre/instrumentation). Taking as a starting point the musical surface, abstract structures may be established, such as grouping/segmentation, metre, motivic categories. In this paper I will discuss aspects of the musical surface that 'challenge' the 'standard' understanding of musical surface as the note level of a musical piece and I will stress the role 'higher' level structures, such as beat/metre, parallelism, streaming, and, even, grouping and harmony, play on the construction of the surface *per se.*

In the first section below, I will present basic concepts underlying the 'standard' use of musical surface in cognitive and computational music research. In the next section, I will discuss issues relating to automated music transcription. Research in this field sheds light on various aspects of the musical surface, in the sense that music transcription can be seen as musical surface extraction; difficulties encountered in automated music transcription can help understand and resolve common misconceptions about the characteristics of the musical surface. In the final section, I will suggest that some processes, that are considered 'higher-level' (*e.g.* beat, metre, tonality, parallelism, voicing), are actually an integral part of the musical surface. I will claim that beat structure, chord simultaneities, voices/streams are internal 'primitive' constituent elements of the musical surface.

Musical surface and categorical perception

Our perceptual mechanisms actively try to organise the infinite variety and nuances of an input musical signal into manageable perceptual events. '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). Discontinuities and changes in the acoustic signal, for instance, in the temporal domain (*e.g.* onset detection) and the domain of frequencies (*e.g.* spectrum peaks), signify points at which the continuously evolving sound continuum can be broken down into smaller constituent parts.

The elementary events perceived as constituent units of an acoustic continuum are further grouped together into elementary categories. Research in categorical perception has investigated various facets of musical perception, especially those of musical pitch and time perception - see overviews and discussion in (Sloboda 1985; 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).

The differentiation between *surface* and *deep* structure in music is mostly associated with the work of musicologist H. Schenker (*e.g.* Schenker 1935). Lerdahl and Jackendoff (1983) bring together aspects of generative linguistics and Schenkerian analysis in their Generative Theory of Tonal Music.

Jackendoff (1987) describes the *musical surface* as being the 'lowest level of representation that has musical significance' (p. 219) and suggests a direct link/ analogy to 'the system of available phonemes in a language' (p. 218). As the acoustic signal in spoken language is discretized into minimal categorically-perceived phonetic units, so is the musical acoustic continuum perceived as minimal discrete categorically-perceived musical units. In relation to tonal music Jackendoff 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).' (218) Sloboda (1985) suggests that "the basic 'phoneme' of music is a 'note'" (p.24) and presents empirical evidence of categorical perception in the frequency and time domains (pitch and duration). Wiggins (2007) states 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 maintains that 'a piano roll is actually a more accurate approximation' (p.326) of the musical surface than the musical score. The musical surface is commonly understood as being roughly equivalent to the notes of a musical score (or, even, a quantised piano-roll).

'The study of the complex processes by which the brain constructs a heard musical surface from auditory input belongs to the fields of acoustics and psychoacoustics. The musical surface, basically a sequence of notes, is only the first stage of musical cognition.' (Jackendoff and Lerdahl, 2006:37). Assuming these processes, *i.e.*, taking the musical surface as a given, Lerdahl and Jackendoff propose four levels of musical structure (namely, grouping structure, metrical structure, time-span reduction and prolongation reduction) that are 'derived ultimately from the musical surface' (Jackendoff, 1987:219; also Lerdahl and Jackendoff, 1983, fig.1.1). Following this framework, much computational and, also, cognitive research takes as a starting point the musical score (essentially a sequence of plain notes and note simultaneities) or a piano roll, and, then, examines the derivation of various 'higher' level structures, such as beat, metre, grouping, harmonic analysis, reduction and so on, that are assumed to lie *above* the musical surface.

It is suggested, in this paper, that the musical surface is not 'the notes' *per se* (the surface may contain both notes and other more complex events and features), and that 'higher' level musical knowledge is necessary for the derivation of the musical surface from the acoustic input, meaning that aspects of musical processes such as metre induction, grouping, pattern extraction, elementary harmonic analysis, lie *at* or, even, *below* the musical surface (acoustics and psychoacoustics are not necessarily sufficient for the derivation of the musical surface). It should be noted that an expressive piano-roll representation cannot be considered in any case as the musical surface as pitch events are not quantised (musical time is perceived categorically, not in milliseconds).

Scheirer (2000) introduces a very different view to the notion of musical surface. He defines ‘ the *musical surface* to be the set of representations and processes that result from immediate, preconscious, perceptual organization of a musical stimulus and enable a behavioral response.’ (p.61) Scheirer’s understanding of the musical surface is rather dismissive of the ‘standard’ linguistic-inspired notion of the musical surface. In his model ‘no notes or other reified entities of sound perception are used.’ (p.156) His research relies on the notion of *understanding without separation* which refers to ‘the idea that it is possible (and usually desirable) to construct theories and computer models that analyze sounds holistically and directly, without first separating them into notes, voices, tracks, or other entities. …’ (p.220)

Scheirer’s view that note or other music symbols are essentially irrelevant to music perception seems to be exaggerated. It is common understanding and there is sufficient research that shows that listeners spontaneously memorise, recognize, compare and classify things such as tunes, melodic motives, rhythmic patterns, harmonic progressions extracted from musical input; in my view, it is plausible that such processes take place at a symbolic level which is cognitively more efficient and ecomonic, than at a lower sub-symbolic acoustic/psychoacoustic level. His critique, however, on aspects of what he calls ‘the structuralist approach to music perception’ (p.219) contains many interesting insights that may have far-reaching ramifications for cognitive and computational musicology. I agree with his view that the musical surface does not boil down to a mere symbolic note representation and that the human mind does not make full transcriptions of musical works prior to higher-level musical processing.

Automated music transcription

‘One of the hard problems in musical scene analysis is automatic music transcription, that is, the extraction of a human readable and interpretable description from a recording of a music performance.’ (Cemgil *et al.*, 2006). Usually, the ‘human readable and interpretable description’ is taken to be the musical score and, more specifically, the musical notes of a score (*e.g.*, Cemgil *et al.*, 2006; Klapuri, 2004).

During the last decade there has been significant progress in music transcription from audio; transcription accuracy for polyphonic music has exceeded 60% in some cases. However, there is still a long way to go before approaching human transcription levels.

Music transcription involves extracting a musically relevant symbolic representation from audio. What should this symbolic representation be? Should it be the notes (onsets, durations and pitches) of the original composed musical score (*when* a score exists), *i.e.* the symbolic representation written by the composer to be performed?

Should it be a sequence of elementary musical events perceived by a skilled musician? Should it be an acoustically and psycho-acoustically relevant abstraction of the signal that need not coincide neither with a compositional / performative abstraction, nor a perceptual representation, *i.e.*, a kind of preliminary symbolic abstraction that requires further processing to reach the original score or a perceptually plausible description?

The most common view employed in automated music transcription, is that a system should extract the notes of the original score (performance accuracy is measured against score representations). Underlying this view is an assumption that transcription involves discovering the original code that generated a specific music audio signal. Klapuri states that: written music is primarily a *performance instruction*, rather than a representation of music. It describes music in a language that a musician understands and can use to produce musical sound. From this point of view, music transcription can be viewed as discovering the recipe, or, reverse-engineering the source code of a music signal. (Klapuri 2004: 1)

I would claim that the aim of music transcription is not necessarily to reconstruct the 'original score.' Often, there exists no score at all, as is the case for many traditional musics. When a score does exist, it is not necessarily the case that a listener is meant to be able to perceive the full detail of the score – on the contrary, sometimes composers strive to create new timbres and textures that are perceived at a holistic level that is more than the constituent parts. For instance, orchestration techniques often aim at creating new timbres using standard instruments (a piano transcription of a complex symphonic work may be a more realistic/manageable task than full transcription of each instrument), or composers often aim at creating textures that are meant to be perceived as rich harmonic/contapunctal textures rather than sequences of independent elementary events.

Human transcribers do not necessarily try to extract the exact score. Hainsworth and Macloed (2003) report an informal study regarding transcription. They report that most interviewed musicians state that they start from global characteristics and move gradually to the finer detail of a transcription. Transcription starts with style recognition, instrumentation, large scale repetitions, rhythmic features, then, melodies, base line and harmonic content identification and, finally, transcription of inner parts. At least half of the interviewees state that they tried to produce a faithful transcription of the *important* features of the piece, and allowed differences in the fine detail that did not alter the identity of the piece. It is clear that transcription targets, in general, the generation of a symbolic abstraction that represents as faithfully as possible the heard musical performance/recording, and, only in some cases, the extraction of the exact notes and other details of the actual score/ performance.

Recent research in automated music transcription has shown that a purely bottom-up approach (from audio to score) is not possible; higher-level music processing is necessary to assist basic multi-pitch and onset extraction techniques so as to reach acceptable transcription results. 'Nowadays the concept of automatic music transcription includes several topics such as multipitch analysis, beat tracking and rhythm analysis, transcription of percussive instruments, instrument recognition, harmonic analysis and chord transcription, and music structure analysis.' (Ryynänen and Klapuri, 2008: 73). Some of the 'higher' level musical processes that are necessary for transcription will be addressed in more detail in the sections below.

Musical surface events and processes

A common underlying assumption in much cognitive and computational modelling of musical understanding is that musical structural processing starts at the musical surface and proceeds towards higher structural levels, such as metre, rhythmic patterns, melodic motives, harmonic structure and so on. Lerdahl and Jackendoff's influential theory is grounded on this assumption: 'The musical surface, basically a sequence of notes, is only the first stage of musical cognition. Beyond the musical surface, structure is built out of the confluence of two independent hierarchical dimensions of organization: rhythm and pitch. …' (Jackendoff and Lerdahl, 2006: 37). Most researchers agree that the formation of the musical surface from audio requires a large amount of processing; once, however, the surface is formed, it is assumed that the road to higher-level processing is open.

I would like to challenge this view. I would like to suggest, not only that higher-level processing influences the formation of the musical surface, but that some processes that are considered 'higher-level' are actually *necessary* for the formation of the surface, which means essentially that they are *at* or *below* the musical surface. I will claim that beat structure, chord simultaneities and voice separation are internal 'primitive' elements of the musical surface, that are necessary for the surface to exist.

It is well established that human perception fills in gaps in stimuli (*e.g.* phoneme restoration in language) and actively organises input data in a holistic top-down fashion. In this sense, knowledge of a specific musical idiom influences perception of the musical surface. However, we will not discuss further such influences in this paper, but we will focus on some aspects of musical structure that do not just influence perception but are necessary components for perceiving the musical surface at all.

Beat tracking

The first such aspect is beat structure (I would say, low- and mid-levels of metrical structure). Time quantization is linked to categorical perception of note onsets and durations. Musical time is not measured perceptually in milliseconds - we do not encode rhythmic patterns in absolute time intervals. The temporal time spans are rather organised in small integer values and ratios that are easy to encode and remember (see categorical perception of temporal units in Sloboda 1985). The problem, however, with the temporal domain (as opposed to the pitch domain) is that there exists no absolute musically relevant scale of time which enables rounding-up absolute time intervals to simpler integers (in the pitch domain, we have absolute pitch scales, such as the 12-tone equal-tempered scale in western music, against which actual performed frequencies may be quantized). Not only a 'relative clock' is necessary to measure musical time (Povel and Essens, 1985), but also a dynamically-changing relative clock that follows tempo changes.

Let us assume that we have a sequence of repeating tones that gradually accelerate and slow down as depicted in Figure 1a. For the purposes of this paper, such a sequence was constructed and an informal experiment was performed. This initial sequence comprised of a repeating E4 piano note (the actual time intervals, counted as multiples of 64th note durations, were: 15-17-15-13-11-12-13, at a tempo of 120 beats per minute). The sequence (repeated 4 times) was presented to 46 undergraduate music students; they were asked whether they perceived a metrical structure (Yes/No) – additionally, they were urged (if they wished) to indicate time-signature and to notate the sequence in standard music notation. As expected, the first sequence (Figure 1a) was perceived by most as a sequence with fluctuating tempo (no metrical

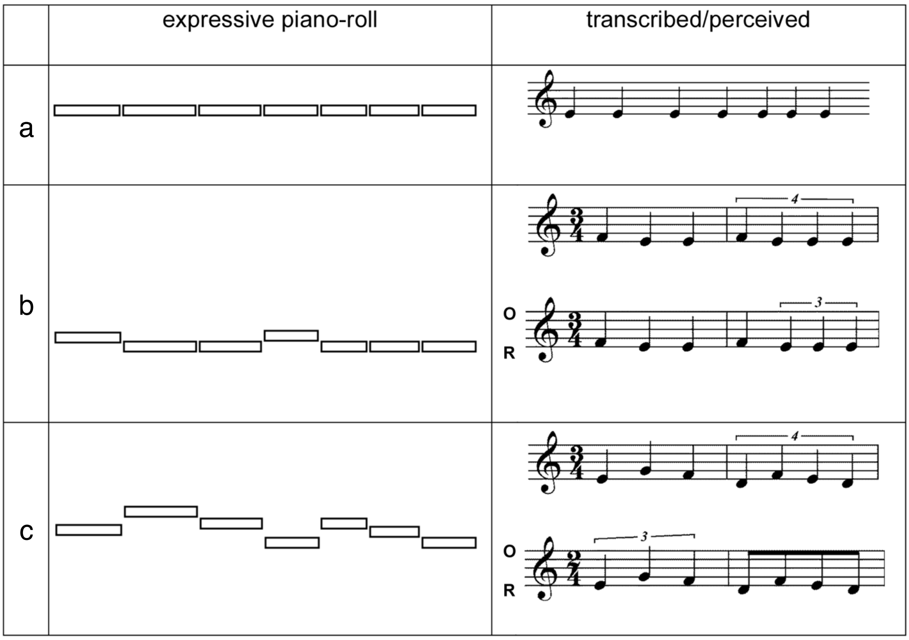

Figure 1.

A temporally fluctuating sequence of tones is perceived as a temporally fluctuating sequence of repeating tones in (a), and as metrically organised in three and four tones in (b, c) due to phenomenal accents (b) and parallelism (c).

structure) – see figure 2. The next two sequences (Figure 1b,c) were temporally identical to the first sequence, but some notes were altered in terms of pitch. Most listeners indicated that they perceived a metrical structure (Figure 2) and many notated the sequences as comprising of two equal length sub-sequences of 3 and 4 notes (instances of notation examples are depicted in Figure 1b,c). In the case of the sequence of Figure 1b, the F4 note was perceived as being accented (phenomenal accent) and it induced a metrical structure at the level of the accented notes. In the case of the sequence of Figure 1c, it was primarily parallelism (interval pattern: ascending third – descending second) that suggested a periodicity coinciding with the beginnings of the repeated patterns (see Cambouropoulos 2006 on the relation between parallelism and grouping).

As can be seen in this example, the musical surface corresponding to the same temporal pattern is different, depending on the induced metrical structure. In the first sequence, a fluctuating beat is perceived, and the sequence is essentially understood as an 'isochronous' sequence with varying tempo. In the next two sequences, cues are given that allow a listener to parse the sequence in two groups of approximately equal duration, allowing thus a metrical structure to emerge. In all cases, finding the beat or metre is a prerequisite to determining categorically-perceived durations. Without a dynamically evolving relative clock it is not possible to quantise meaningfully a temporal sequence. Of course, any sequence can be quantised according to an arbitrary minimal unit (*e.g.* 16th or 32nd duration), but such quantisation often does not have perceptual relevance (the inadequacy of such an approach can be seen when one tries to quantise expressively performed music without a beat-track in music notation programs).

In this sense, I would suggest that metrical structure (primarily the tactus level) 'lives' *at* or *below* the musical surface. The temporal facet of the musical surface, (*i.e.*,

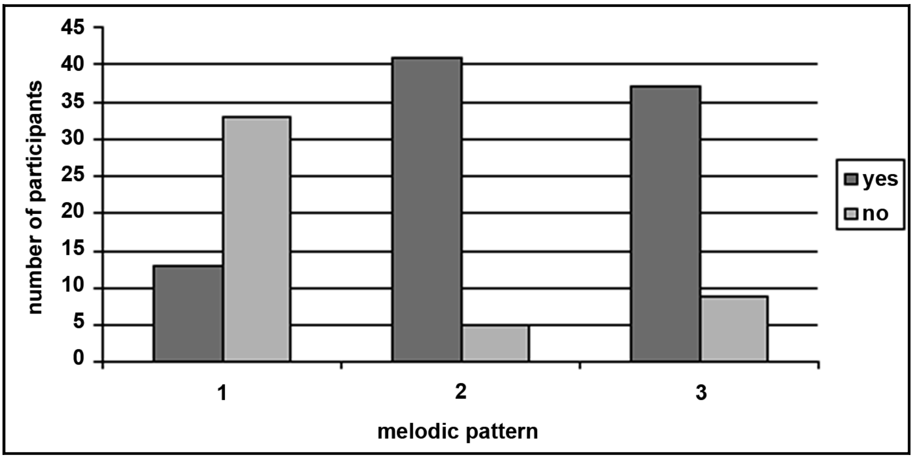

Figure 2.

Number of participants that perceived a metrical structure in the three different melodic patterns (that share the same temporal pattern).

categorically perceived durations), is not possible without access to a significant part of metrical structure. Beat structure is not extracted after the musical surface is formed, but, rather, is a precursor to the formation of the musical surface itself.

Finding an appropriate relative clock to measure musical time is equivalent to the computational notion of beat induction (finding an appropriate beat level – essentially the tactus level), whereas a dynamically-changing clock can be thought as being equivalent to beat tracking (following the beat in music with changing tempo) – see (Dixon and Cambouropoulos, 2001). Research in beat induction and beat tracking in recent years has produced a number of relatively stable beat tracking systems that operate directly on audio or expressively performed MIDI data (see recent review in Guyon and Dixon, 2007). Such systems illustrate the fact that beat tracking does not require a musical surface in advance; on the contrary beat/tempo can be thought of as a facet of the musical surface itself. Scheirer (that has developed a tempo tracking system) maintains that 'the beat of a piece of music is a constituent of the perceptual musical surface' (Scheirer, 2000: 82).

Beat trackers are nowadays almost always incorporated in automated music transcription systems that attempt to extract a symbolic representation from audio or expressive MIDI (as discussed above). This shows again that beat structure is necessary for the extraction of the musical surface and therefore is an integral part of the musical surface.

Chords - multi-notes simultaneities

Should the note be considered as the lowest level of representation that has musical significance and perceptual relevance? There is evidence that things such as 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. It seems plausible that listeners perceive such larger entities in a holistic manner prior to perceiving their constituent parts (such parts in some cases might not be registered/encoded at all). Vernon (1934), a gestalt psychologist, suggests that ordinary listeners frequently perceive a more or less complex figure (*e.g.* a complex tone or a chord) without knowing or being able to analyse its actual constituent elements.

Empirical research has shown that listeners perceive musical intervals categorically (see Handel, 1989; Smith *et al.*, 1994; Burns 1999). Categorical perception applies to chords, as well, as has been shown by Locke and Kellar (1973). In their study, a series of chords composed of two steady outer tones a perfect fifth apart and a middle tone, that took values in small steps from minor third to major third, were used as stimuli in two experiments; categorical perception for major and minor chords was clearly evidenced in the responses by musicians (non-musician listeners showed weaker signs of categorical perception). It is suggested that pitch intervals or chords, rather than isolated notes, are actually 'equivalent' to the categorically perceived phonemic units of language (depending on streaming – see below).

Due to octave equivalence and transpositional pitch interval equivalence, harmonic pitch intervals and chords are considered 'equivalent' even though their constituent pitches may be placed in different octaves. Empirical research has shown that listeners confuse pairs of tones that are related by inversion (Deutsch, 1999) and that chord positioning does not affect recognising components of a chord, *i.e.* that chords in different positions are essentially equivalent (Hubbard and Datteri, 2001).

Parncutt (1989) provides a psychoacoustic explanation of how chords are heard.[1] As our perceptual mechanisms analyse a complex periodic sound into partials and then reintegrate them into a single percept, so can chords be heard as a single entity with a single perceived root rather than three or more individual co-sounding tones. He suggests that the same factors that govern the perception of individual pitches, govern the perception of chords. Parncutt's approach to chord perception is in line with the suggestion in this paper, that chords are perceived at the surface level as single integrated entities rather than sets of constituent individual notes.

Identifying a set of co-sounding partials/notes as, for instance, a major or minor or diminished chord, involves additionally culture-specific knowledge that is acquired via exposure to a certain idiom. 'Chord recognition is the result of a successful memory search in which a tone series is recognized as a pattern stored in long-term memory.' (Povel and Jansen 2001:185). Template matching models (Parncutt 1994) are integrated in cognitive mechanisms of musical listening and are responsible for the extraction of musically pertinent entities from sound at the surface level (the musical surface is music-idiom specific in a similar way that phonological structure is language specific).

Research in automated transcription has recently focused on problems other than full transcription of all the notes of a piece. Multi-pitch extraction is a difficult problem, and it may be unnecessary for a number of tasks. For instance, Ryynänen and Klapuri (2008) propose a system that transcribes the melody, the bass line and chords in polyphonic music (tested on a large collection of Beatle songs). Learned pitch-class profiles for major and minor chords (along with chord-transition probabilities) are used in the chord transcription process. Chord transcription may be a more realistic computational problem to solve, and it probably is more cognitively plausible than the extraction of every individual note of inner voices.

More complex configurations of tones such as tone clusters, tremolos, trills, glissandi can also be perceived by listeners as wholes rather than combinations of atomic lower-level components. 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). A glissando is perceived and can be represented as a single

[1] Parncutt (1997) proposes an extended model that calculates the perceptual root of a chord from its pitch classes, voicing, and the prevailing tonality.

entity with parameters such as start- and end-pitch, slope of transition, duration and intensity. 'For example, suppose there is a glide in frequency, bounded by a rise and fall in intensity. Between these boundaries, the change in frequency may be measured by the auditory system and assigned to the unit as one of its properties. This frequency-gliding unit will prefer to group with other ones whose frequency change has the same slope and which are in the same frequency region.' (Bregman, 1990:644).

The above discussion supports the idea that chords (and other multinote entities) tend to be perceived at the surface level as single integrated entities rather than agglomerates of independent atomic notes.

Musical streams

Listeners break down the acoustic continuum into *musical streams*; they perceive streams of musical events such as streams of notes (*e.g.*, melodic lines) or streams of chords (*e.g.*, accompaniment). A number of perceptual factors studied in the domain of auditory scene analysis (Bregman 1990) enable a listener to integrate or fuse co-modulating components (*e.g.*, partials or notes moving in parallel) into coherent events and sequences of events, and, at the same time, to segregate then from other independent musical sequences. The notion of musical stream relates in certain ways to the musicological notion of voice, as discussed by Cambouropoulos (2008).

Voice separation refers to the task of separating a musical work, consisting of multi-note sonorities, into independent constituent voices. Recently, there have been a number of attempts to model computationally the segregation of music into separate voices (see brief overview of research in the symbolic domain in Cambouropoulos 2008). Much of this research takes as input a symbolic representation of the music (*e.g.* MIDI or other symbolic notation) and outputs a number of discrete monophonic voices (few models allow non-monophonic voices/streams – Kilian and Hoos 2002; Rafailidis *et al.*, 2008, 2009).

It is herein suggested that stream segregation into sequences of musical events takes place at the surface level, in the sense that streaming is necessary for the categorical perception of note and chord events - it is not true that first, notes/chords are formed and, afterwards, they are organised into streams/voices. The reason for this is that pitch intervals and note durations are defined in relation to *two* pitches and *two* onsets belonging to the same stream. A melodic interval requires two successive pitches to be defined. A note duration usually requires two successive onsets to be defined, as note offsets often cannot be determined with accuracy (especially for sounds that gradually fade out such a piano or plucked string sounds). To know which music event (pitch/onset) follows which event (pitch/onset), musical streams have to be determined in advance. For instance, the first melody C5 note in Figure 3 is a quarter-note as it belongs to the same stream with the following Bb4 note, and not an eighth-note as the following C4 note belongs to a different stream (the accompaniment). In a score extraction system (from expressive MIDI performances) an elementary stream separation algorithm had to be introduced in order to obtain reasonable quantised note durations (Cambouropoulos 2000). The main point is that melodic intervals, note durations and, even, chord distances/relations, cannot exist without knowing note/chord successions within musical streams.

The relation between musical streaming and segmentation is explored in (Rafailidis *et al.* 2008), and a computational model is presented that detects *stream segments* in symbolic musical data. A stream segment is taken to be a relatively small number of tones grouped together into a coherent 'whole' perceived independently from other adjacent tones (co-sounding tones, and preceding and following tones) – see example in Figure 3.

It is suggested that such an approach may be applied in the audio domain as well; it is proposed that stream segments might be extracted from audio prior to individual pitches. A system can be developed to detect frequency components that co-modulate and evolve 'together' forming independent musical stream segments, and, then, attempt to analyse some streams further (*e.g.*, detect fundamental frequencies in monophonic streams, or determine key/harmonic evolution in chordal streams). Stream segments can be viewed as unified sound events at some level of the musical surface.

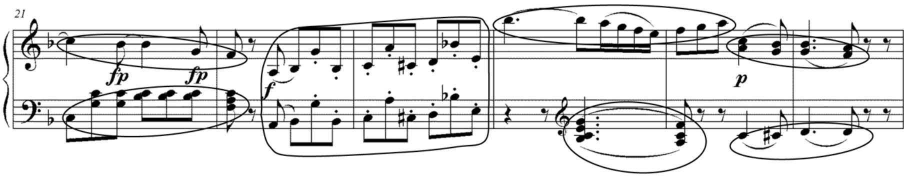

Figure 3.

Excerpt from Mozart's Sonata K332, Allegro Assai. Potential stream segments are circled - alternative segmentations are also possible : third segment can be broken into two pseudopolyphonic voices, and last two segments can be seen as a single homophonic stream segment (Rafailidis et al. *2008).*

The notion of stream segments is a novel concept that aims to describe in a more systematic manner the domain that lies in between pure homophony and pure contrapunctal polyphonic music. Musicians and music theorists have a good understanding of homophony and polyphony, and can detect homophonic and contrapunctal elements in a given musical piece, but there exists, to my knowledge, no systematic music theoretic concept that describes the 'uncharted' territory that lies between the two 'extremes' of pure homophony and polyphony. It is hoped that further research on stream segments can provide a new systematic way to examine such aspects of musical texture. Additionally, stream segments may suggest new ways to approach issues pertaining to the perception of the musical surface and structure, and, also, to develop music transcription systems and music analytic tools.

CONCLUSION

In this paper problems and misconceptions pertaining to the notion of the musical surface were discussed. It has been suggested that the note representation is probably not sufficient to be considered as the lowest level that is perceptually pertinent and that more complex musical events and processes, such as chords, beat/metre, stream segments, should be considered holistically as constituent elements of the surface. The aim of this discussion on the characteristics of the musical surface is not to give a definitive description of the nature of the musical surface, but, rather, to provide a critical re-examination of some commonly accepted assumptions leading, thus, to a better understanding of this fundamental notion. Further empirical research may be necessary to explore in more detail what listeners perceive at the surface level. A better understanding of what the musical surface is and how it is formed, may lead to cognitive models that describe better the perceptual processes involved, and computational systems that achieve improved performance in musical tasks that require a more sophisticated and smoother transition between the audio and the symbolic domain.

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

• References

Burns, E. M. (1999). Intervals, scales and tuning. In D. Deutsch (Ed.), *Psychology of Music* (2nd edition) (pp. 215-264). Academic Press, San Diego.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organisation of sound.* The MIT Press, Cambridge (Ma).

Cemgil, A. T., Kappen, H. J., & Barber, D. (2006). A generative model for music transcription. *IEEE Transactions on Audio, Speech, and Language Processing, 14*(2), 679- 694

Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception 26*(1), 75-94.

Cambouropoulos E. (2006). Musical parallelism and melodic segmentation: A computational approach. *Music Perception 23*(3), 249-269.

Cambouropoulos, E. (2000). From MIDI to traditional musical notation. *Proceedings of the AAAI Workshop on Artificial Intelligence and Music*, Austin, Texas.

Deliège, I. (2007). Similarity relations in listening to music: How do they come into play? *Musicae Scientiae*, Discussion Forum 4A:9-37.

Deliège, I. (2001). Prototype effects in music listening: An empirical approach to the notion of imprint. *Music Perception, 18*(3), 371-407.

Deliège, I. (1997). Similarity in processes of categorisation: Imprint formation as a prototype effect in music listening. In M. Ramscar, U. Hahn, E. Cambouropoulos and H. Pain (Eds.), *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 59-65). University of Edinburgh, Edinburgh, U.K.

Deliège, I. (1996). Cue abstraction as a component of categorisation processes in music listening. *Psychology of Music, 24*(2),131-156.

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

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The Psychology of Music* (revised version) (pp. 349-411). Academic Press, San Diego.

Dixon, S., & Cambouropoulos, E. (2000). Beat tracking with musical knowledge. In *ECAI 2000: Proceedings of the 14th European conference on artificial intelligence* (pp. 626-630). Amsterdam: IOS Press.

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

Gouyon, F., & Dixon, S. (2005). A review of automatic rhythm description systems. *Computer Music Journal, 29* (1), 34–54.

Hainsworth, S., & Macloed, M. (2003).*The automated music transcription problem.* http://wwwsigproc.eng.cam.ac.uk/ swh21/ontrans.pdf

Handel, S. (1989). *Listening. An introduction to the perception of auditory events.* The MIT Press, Cambridge (Ma).

Hubbard, T. L., & Datteri, D. L. (2001). Recognizing the component tones of a major chord. *American Journal of Psychology, 114*(4), 569-589.

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

Kilian, J., & Hoos, H. (2002). Voice separation: A local optimisation approach. In *Proceedings of ISMIR'02* (pp.39-46).

Klapuri, A. (2004). *Signal processing methods for the automatic transcription of music*. Ph.D. dissertation, Tampere University of Technology, Finland.

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

Locke, S. & Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex, 9*, 355-369.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Springer, Berlin.

Parncutt, R. (1994). Template-matching models of musical pitch and rhythm perception. *Journal of New Music Research, 23*, 145-167.

Parncutt, R. (1997). A model of the perceptual root(s) of a chord accounting for

voicing and prevailing tonality. In Marc Leman (Ed.) *Music, gestalt, and computing: Studies in cognitive and systematic musicology*. Springer, Berlin.

Povel, D.- J., & Jansen, E. (2001). Perceptual mechanisms in music processing. *Music Perception, 19*(2), 169–198

Rafailidis, D., Nanopoulos, A., Cambouropoulos, E., & Manolopoulos, Y. (2008).

Detection of stream segments in symbolic musical data. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR08)*, Philadelfia, Pennsylvania, U.S.A.

Rafailidis, D., Cambouropoulos, E., & Manolopoulos, Y. (2009). Musical voice integration/ segregation: VISA revisited. In *Proceedings of the 6th Sound and Music Computing Conference (SMC09),* Porto, Portugal.

Ryynänen M. P., & Klapuri A. P. (2008). Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal, 32*(3), 72-86

Schenker, H. (1935). *Der Freie Satz* (translated E. Oster, 1979). Longman, New York.

Scheirer, E. D. (2000). *Music-listening systems*. Ph.D. thesis, Massachusetts Institute of Technology, June, 2000.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford University Press, Oxford.

Smith, L. B., Kemler Nelson, D., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition, 52*, 23–54.

Vernon, P. E. (1934). Auditory perception. I. The Gestalt approach. II. The evolutionary approach. *British Journal of Psychology, 25*, 123-139; 265-283.

• La superficie musical : desafiando suposiciones básicas

El presente artículo aborda los problemas e ideas erróneas relacionadas con la noción de *superficie musical*, una noción que se suele pensar relativamente sencilla y a menudo se da por hecho en la investigación computacional y cognitiva. Se sugiere que la superficie musical esta compuesta por eventos musicales (complejos) percibidos como totalidades coherentes en el flujo de la música - la superficie musical no es solamente una secuencia desestructurada de eventos atómicos (notas), como en una partitura o en el papel agujereado de una pianola. Además, se aduce que el surgimiento de la superficie musical involucra mecanismos complejos que requieren, no sólo de la extracción de varios tonos (*pitch*) de la señal acústica, sino del empleo de procesos cognitivos como el seguimiento del rítmo (*beat*), inducción métrica, identificación de acordes, y separación de voces (*stream/ voice separation*). Estos procesos no entran en juego hasta *después* de que la superficie se ha formado, y constituyen además una parte medular en la formación de la superficie musical *per se*.

• La superficie musicale : Oltre le concezioni tradizionali

Questo articolo tratta dei problemi e confusioni che riguardano la nozione di *superficie musicale*, una nozione che é comunemente considerata chiara ed é spesso data per scontata nella ricerca computazionale e cognitiva. Si ritiene che la superficie musicale comprenda eventi musicali (complessi) percepiti come unitá all'interno di "stream" musicali coerenti – la superficie musicale non é semplicemente una sequenza astrutturata di eventi tonali atomici, come possono essere le note di uno spartito o la rappresentazione "piano-roll". Inoltre, si ritiene che l'emergenza della superficie musicale coinvolga meccanismi piuttosto complessi che richiedono non solo l'estrazione parallela di altezze tonali dal segnale acustico ma anche l'impiego di processi cognitivi come il "tracking" del battito, l'induzione del metro, l'identificazione dell'accordo e la separazione "stream"/voce. Tali processi non avvengono dopo la costituzione della superficie, ma sono piuttosto una parte integrale della formazione della superficie musicale *per se*.

• La surface musicale : Une remise en cause de principes de base

Cet article met en évidence des difficultés et des erreurs de jugement relatives à la notion de *surface musicale*, une notion qui est souvent vue comme relativement évidente, considérée comme un concept qui va de soi dans de nombreuses recherches informatiques et/ou cognitives. Il est suggéré que la surface musicale est constituée d'évènements musicaux (complexes) perçus comme des entités complètes au sein de flux musicaux cohérents : la surface musicale ne se borne pas à une séquence non-structurée d'évènements atomiques, de notes, à l'image de partitions ou de représentation en « *piano-roll* » de fichiers MIDI. En outre, nous affirmons que l'émergence de la surface musicale met en jeu des mécanismes assez complexes qui nécessitent non seulement l'extraction de hauteurs simultanées à partir du signal acoustique, mais aussi l'emploi de processus cognitifs tels que le suivi de pulsation, l'induction métrique, l'identification d'accords et la séparation en flux ou en voix. De tels processus n'entrent pas en jeu une fois la surface formée, mais jouent un rôle essentiel dans la formation de cette surface musicale *per se*.

Die musikalische Oberfläche : Fragwürdige Grundannahmen

Der vorliegende Aufsatz beschäftigt sich mit Problemen und Missverständnissen, welche die Bedeutung der *musikalischen Oberfläche* betreffen – ein Begriff, der allgemein für relativ offenkundig gehalten und in der computerbasierten kognitiven Forschung als gegeben vorausgesetzt wird. Es wird bei diesem Begriff angenommen, dass die musikalische Oberfläche aus (komplexen) musikalischen Ereignissen besteht, die innerhalb eines kohärenten Musikflusses als Ganzes wahrgenommen werden. Die musikalische Oberfläche ist folglich keine unstrukturierte Folge unzusammenhängender Notenereignisse, wie etwa in Form von Partiturnoten oder einer Musik-Repräsentation als „Klavierrollen"-Darstellung eines Sequenzers. Darüber hinaus wird behauptet, das Erscheinen der musikalischen Oberfläche beinhalte weitere komplexe Mechanismen, die nicht nur eine mehrstimmige Tonhöhen-Extrahierung aus dem akustischen Signals erfordern, sondern auch die Anwendung von kognitiven Prozessen wie das Verfolgen des Rhythmus' und des Metrums, die Identifikation der Akkorde und die Trennung der einzelnen Tonsatzstimmen. Diese Prozesse spielen *nach* der Entstehung der musikalischen Oberfläche keine Rolle mehr, sondern sind vielmehr ein wesentlicher Bestandteil der Entstehung der musikalischen Oberfläche an sich.