Journal of New Music Research
2007, Vol. 36, No. 3, pp. 169 – 184

Routledge
Taylor & Francis Group

Sound and Music Computing: Research Trends and Some Key Issues

Gerhard Widmer[1,2], Davide Rocchesso[3], Vesa Välimäki[4], Cumhur Erkut[4], Fabien Gouyon[5], Daniel Pressnitzer[6], Henri Penttinen[4,7], Pietro Polotti[8,9] and Gualtiero Volpe[10]

[1]Johannes Kepler University Linz, Austria, [2]Austrian Research Institute for Artificial Intelligence (OFAI), Austria, [3]IUAV University of Venice, Italy, [4]Helsinki University of Technology, Finland, [5]INESC Porto, Portugal, [6]CNRS-Université Paris Descartes & DEC, France, [7]Stanford University, USA, [8]University of Verona, Italy, [9]Conservatory of Music, Italy, [10]University of Genova, Italy

Abstract

This contribution attempts to give an overview of current research trends and open research problems in the rich field of Sound and Music Computing (SMC). To that end, the field is roughly divided into three large areas related to Sound, Music, and Interaction, respectively, and within each of these, major research trends are briefly described. In addition, for each sub-field a small number of open research (or research strategy) issues are identified that should be addressed in order to further advance the SMC field.

1. Introduction

In Bernardini and De Poli (2007), an attempt was made to define the field of Sound and Music Computing (henceforth SMC), trying to delineate its core areas and boundaries. The aim of the present article is to give an overview of current research trends (we deliberately refrain from trying to summarize the state of the art, as that would go far beyond what can be done in a short article like this), with a special emphasis on the open issues that wait to be addressed, or are currently being worked on. Faced with the great variety of research topics within SMC, we have tried to give our summary a coherent structure by grouping the topics into three major areas – Sound, Interaction and Music – which are further divided into sub-areas.

Figure 1 depicts the relationships between the different research areas and sub-areas as we see them. We make a basic distinction between research that focuses on sound (left-hand side of the figure) and research that focuses on music (right-hand side of the figure). For each research field, there is an analytic and a synthetic approach. The analytic approach goes from encoded physical (sound) energy to meaning (sense), whereas the synthetic approach goes in the opposite direction, from meaning (sense) to encoded physical (sound) energy. Accordingly, analytic approaches to sound and music pertain to analysis and understanding, whereas synthetic approaches pertain to generation and processing. In between sound and music, there are multi-faceted research fields that focus on interactional aspects. These are performance modelling and control, music interfaces, and sound interaction design.

The following sections identify and discuss some of the major current research trends in these areas, and for each sub-field a small number of open research issues are identified that should be addressed in order to further advance the SMC field. A discussion of more general strategies (beyond research) to further SMC can be found in Serra et al. (2007).

2. Sound

In this section we review the research on sound that is being carried out within the boundaries identified in Bernardini and De Poli (2007). From a sound to sense point of view, we include the analysis, understanding and description of all musical and non-musical sounds except speech. Then, in the sense to sound direction, we include

Correspondence: Gerhard Widmer, Department of Computational Perception, Johannes Kepler University Linz, Austria.
E-mail: gerhard.winder@jku.at

DOI: 10.1080/09298210701859222

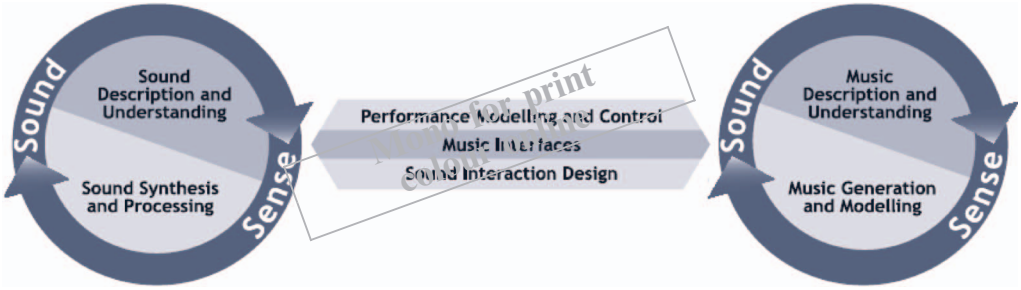

Fig. 1. Relations between the different SMC research areas.

the research that is more related to sound synthesis and processing.

2.1 Sound description and understanding

One of the basic aims of SMC research is to understand the different facets of sound from a computational point of view, or by using computational means and models. We want to understand and model not only the properties of sound waves but also the mechanisms of their generation, transmission and perception by humans. Even more, we want to understand sound as the basic communication channel for music and a fundamental element in our interaction with the environment. Sound serves as one of the main signals for human communication, and its understanding and description requires a notably multidisciplinary approach.

Traditionally, the main interest of SMC researchers has been musical sounds and thus the understanding of the sound generated by musical instruments and the specific transmission and perception mechanisms involved in the music communication chain. In recent years, this focus has been broadened and there is currently an increased interest in non-musical sounds and aspects of communication beyond music. A number of the methodologies and technologies developed for music are starting to be used for human communication and interaction through sound in general (e.g. ecological sounds) and there is increasing cross-fertilization between the various sound-related disciplines.

There has been a great deal of research work on the analysis and description of sound by means of *signal processing techniques*, extracting features at different abstraction levels and developing source-specific and application-dependent technologies. Most of the current research in this domain starts from frequency domain techniques as a step towards developing sound models that might be used for recognition, retrieval, or synthesis applications. Other approaches consider sparse atomic signal representations such as matching pursuit, the analytical counterpart to granular synthesis (Sturm et al., 2006).

Also of importance has been the study of sound-producing physical objects. The aim of such study is to understand the acoustic characteristics of musical instruments and other physical objects which produce sounds relevant to human communication. Its main application has been the development of *physical models* of these objects for synthesis applications (Rocchesso & Fontana, 2003; Smith, 2006; Välimäki, et al., 2006), so that the user can produce sound by interacting with the models in a physically meaningful way.

However, beyond the physical aspect, sound is a communication channel that carries information. We are therefore interested in identifying and representing this information. Signal processing techniques can only go so far in extracting the meaningful content of a sound. Thus, in the past few years there has been an exponential increase in research activity which aims to generate semantic descriptions automatically from audio signals. Statistical Modelling, Machine Learning, Music Theory and Web Mining technologies have been used to raise the semantic level of sound descriptors. MPEG-7 (Kim et al., 2005) has been created to establish a framework for effective management of multimedia materials, standardizing the description of sources, perceptual aspects and other relevant descriptors of a sound or any multimedia asset.

Most research approaches to sound description are essentially bottom-up, starting from the audio signal and trying to reach the highest possible semantic level. There is a general consensus that this approach has clear limitations and does not allow us to bridge what is known as the "*semantic gap*" – that is, the discrepancy between what can currently be extracted from audio signals and the kinds of high-level, semantically meaningful concepts that human listeners associate with sounds and music. The current trend is towards multimodal processing methods and top-down approaches based on ontologies, reasoning rules, and cognition models. Also, in practical applications (e.g. in web-based digital music services), collaborative tagging by users is being increasingly used to gain semantic information that would be hard or impossible to extract with current computational methods.

2.1.1 Sound description and understanding: key issues

The above synopsis of research in sound description and understanding has already revealed a number of current limitations and open problems. Below, we present some selected research questions that should be addressed, or issues that should be taken into account in future research.

2.1.1.1 Perceptually informed models of acoustic information processing. There is an active field of research in neuroscience that tries to relate behavioural and physiological observations, by means of computational models. There is a wide variety of approaches in the computational neuroscience field, from models based on accurate simulations of single neurons to systems-based models relying heavily on information theory. SMC has already benefitted in the past from auditory models as signal processing tools. For instance, audio compression schemes such as MP3 are heavily based on models of perceptual masking. This trend is set to continue as the models become more robust and computationally efficient. In the future, the interaction between auditory models and SMC could also be on a conceptual level. For instance, the sensory-motor theory suggests that the study of sound perception and production should be intimately related.

2.1 1.2 Sound source recognition and classification. The ability of a normal human listener to recognize objects in the environment from only the sounds they produce is extraordinarily robust. In contrast, computer systems designed to recognize sound sources function precariously, breaking down whenever the target sound is degraded by reverberation, noise, or by competing sounds. Musical signals present a real challenge for existing systems as the three sources of difficulty are almost always present. SMC can thus contribute to the development of sound source recognition systems, by providing well-controlled test situations that retain an ecological value (Elhilali et al., 2007). In return, models of sound source recognition will have obvious applications in current and future application of SMC, such as score following (adding timbre cues to the pitch cues normally used) or music-information retrieval systems.

2.1.1.3 Sound search and retrieval based on content. Audio content analysis and description enables various new and advanced audiovisual applications and services. Search engines or specific filters could use the extracted description to help users navigate or browse through large collections of audio data. Digital analysis of an audio file may be able to discriminate between speech, music and other entities or identify how many speakers are contained in a speech segment, what gender they are, and even who exactly is speaking. Spoken content may be identified and converted to text. Music might be classified into categories, such as jazz, rock and classical (Tzanetakis & Cook, 2002) (although this is problematic because such categories are user-dependent and perhaps cannot be unequivocally defined). Finally, it may be possible to automatically identify and find particular sounds, such as explosions, gunshots, etc. (Cano, 2007). For such scenarios to become really useful, the necessary improvements in sound search and retrieval will call for a change of paradigm in the description of sounds – from descriptions constrained to a finite number of crisp labels, towards natural language descriptions, at a higher semantic level, similar to that used by humans. A step in this direction might be the inclusion of reasoning rules and knowledge bases (sound ontologies) encoding common sense knowledge about sound. Another key issue is the combination of information from complementary media, such as video or images.

2.2 Sound synthesis and processing

Sound synthesis and processing has been the most active research area in SMC for more than 40 years. Quite a number of the research results of the 1960s and 1970s are now standard components of many audio and music devices, and new technologies are continuously being developed and integrated into new products (Välimäki et al., 2007). The sounds of our age are digital. Most of them are generated, processed, and transcoded digitally. Given that these technologies have already become so common and that most recent developments represent only incremental improvements, research in this area has lost some of its prominence in comparison to others in SMC. Nonetheless, there remain a number of open issues to be worked on, and some of the new trends have the potential for huge industrial impact (see also Leman et al., 2007).

With respect to sound synthesis, most of the abstract algorithms that were the focus of work in the 1970s and 1980s – e.g. FM and waveshaping – were not directly related to a sound source or its perception (though some of the research was informed by knowledge of musical acoustics and source physics). The 1990s saw the emergence of computational modelling approaches to sound synthesis. These aimed either at capturing the characteristics of a sound source, known as *physical models* (Cadoz et al., 1993; Smith, 2006; Välimäki et al., 2006), or at capturing the perceptual characteristics of the sound signal, generally referred to as *spectral or signal models* (Serra, 1997).

The technology transfer expectations of the physical models of musical instruments have not been completely fulfilled. Their expressiveness and intuitive control – advantages originally attributed to this kind of model – did not help commercial music products to succeed in the market place. Meanwhile, synthesis techniques based on

spectral modelling have met with competitive success in voice synthesizers, both for speech and singing voices (Bonada and Serra, 2007), but to a lesser extent in the synthesis of all other musical instruments. A recent and promising trend is the combination of physical and spectral models, such as *physically informed sonic modelling* (Cook, 1997) and *commuted synthesis* (Smith, 2006; Välimäki et al., 2006). Another recent trend is to simulate traditional analog electronics used in music synthesizers of the 1960s and 1970s (Lane et al., 1997; Välimäki and Huovilainen, 2006) and in amplifiers used by electric guitar and bass players (Karjalainen et al., 2006; Yeh and Smith, 2006).

As an evolution of granular synthesis techniques (e.g. Roads, 2001), new corpus-based concatenative methods for musical sound synthesis, also known as *mosaicing*, have attracted much attention recently (Schwarz, 2007). They make use of a variety of sound snippets in a database to assemble a desired sound or phrase according to a target specification given via sound descriptors or by an example sound. With ever-larger sound databases readily available, together with a pertinent description of their contents, these methods are increasingly used for composition, high-level instrument synthesis, interactive exploration of sound corpora, and other applications (Lindeman, 2007).

In sound processing, there are a large number of active research topics. Probably the most well-established are audio compression and sound spatialization, both of which have clear industrial contexts and quite well defined research agendas. Digital audio compression techniques allow the efficient storage and transmission of audio data, offering various degrees of complexity, compressed audio quality and degree of compression. With the widespread uptake of mp3, audio compression technology has spread to mainstream audio and is being incorporated into most sound devices (Mock, 2004). These recent advances have resulted from an understanding of the human auditory system and the implementation of efficient algorithms in advanced DSP processors. Improvements to the state of the art will not be easy, but there is a trend towards trying to make use of our new understanding of human cognition and of the sound sources to be coded.

Sound spatialization effects attempt to widen the stereo image produced by two loudspeakers or stereo headphones, or to create the illusion of sound sources placed anywhere in three-dimensional space, including behind, above or below the listener. Some techniques, such as ambisonics, vector base amplitude panning and wave-field synthesis, are readily available, and new models are being worked on that combine signal-driven bottom-up processing with hypothesis-driven top-down processing (Blauert, 2005). Auditory models and listening tests currently help us to understand the mechanisms of binaural hearing and exploit them in transcoding and spatialization. Recent promising examples include the Binaural Cue Coding method (Faller, 2006) and Spatial Impulse Response Rendering (Pulkki and Merimaa, 2006).

Digital sound processing also includes techniques for audio post-production and other creative uses in music and multimedia applications (Zölzer, 2002). Time and frequency domain techniques have been developed for transforming sounds in different ways. But the current trend is to move from signal processing to content processing; that is, to move towards higher levels of representation for describing and processing audio material.

There is a strong trend towards the use of all these signal processing techniques in the general field of interactive sound design. Sound generation techniques have been integrated in various multimedia and entertainment applications (e.g. sound effects and background music for gaming), sound product design (ring tones for mobile phones) and interactive sound generation for virtual reality or other multimodal systems. Old sound synthesis technologies have been brought back to life and adapted to the needs of these new interactive situations. The importance of control has been emphasized, and source-centred and perception-centred modelling approaches have been expanded towards *interactive sonification* (Hermann & Ritter, 2005).

2.2.1 Sound synthesis and processing: key issues

2.2.1.1 Interaction-centred sound modelling. The interactive aspects of music and sound generation should be given greater weight in the design of future sound synthesis techniques. A challenge is how to make controllability and interactivity central design principles in sound modelling. It is widely believed that the main missing element in existing synthesis techniques is adequate control. The extraction of expressive content from human gestures, from haptics (e.g. pressure, impacts or friction-like interactions on tangible interfaces), from movement (motion capture and analysis) or voice (extraction of expressive content from the voice or breath of the performer), should become a focus of new research in sound generation. This will also open the field to multisensory and cross-modal interaction research. The next problem then concerns how to exploit the extracted contents in order to model sound. Effective sound generation needs to achieve a perceptually robust link between gesture and sound. The mapping problem is in this sense crucial both in musical instruments and in any other device/artefact involving sound as one of its interactive elements.

2.2.1.2 Modular sound generation. Sound synthesis by physical modelling has, so far, mainly focused on accurate reproduction of the behaviour of musical

instruments. Some other efforts have been devoted to everyday sounds (Rocchesso et al., 2003; Rocchesso and Fontana, 2003; Peltola et al., 2007) or to the application of sophisticated numerical methods for solving wave propagation problems (Trautmann et al., 2005; Bilbao, 2007). A classic dream is to be able to build or alter the structure of a musical instrument on the computer and listen to it before it is actually built. By generalizing this thought, the dream changes to the idea of having a toolkit for constructing sounding objects from elementary blocks such as waveguides, resonators and nonlinear functions (Rabenstein et al., 2007). This goal has faced a number of intrinsic limitations in block-based descriptions of musical instruments. In general, it is difficult to predict the sonic outcome of an untested connection of blocks. However, by associating macro-blocks to salient phenomena, it should be possible to devise a constructivist approach to sound modelling. At the lowest level, blocks should correspond to fundamental interactions (impact, friction, air flow on edge, etc.). The sound quality of these blocks should be tunable, based on properties of both the interaction (e.g. pressure, force) and the interactants (e.g. size and material of resonating object). Higher-level, articulated phenomena should be modelled on top of lower-level blocks according to characteristic dynamic evolutions (e.g. bouncing, breaking). This higher level of sound modelling is suitable for tight coupling with emerging computer animation and haptic rendering techniques, as its time scale is compatible with the scale of visual motion and gestural/tactile manipulation. In this way, sound synthesis can become part of a more general constructivist, physics-based approach to multisensory interaction and display.

2.2.1.3 Physical modelling based on data analysis. To date, physical models of sound and voice have been appreciated for their desirable properties in terms of synthesis, control and expressiveness. However, it is also widely recognized that they are very difficult to fit onto real observed data due to the high number of parameters involved, the fact that control parameters are not related to the produced sound signal in an intuitive way and, in some cases, the radical non-linearities in the numerical schemes. All these issues make the parametric identification of physics-based models a formidable problem. Future research in physical voice and sound modelling should thus take into account the importance of models fitting real data, in terms of both system structure design and parametric identification. Co-design of numerical structures and identification procedures may also be a possible path to complexity reduction. It is also desirable that from the audio-based physical modelling paradigm, new model structures emerge which will be general enough to capture the main sound features of broad families of sounds (e.g. sustained tones from wind and string instruments, percussive sounds) and to be trained to reproduce the peculiarities of a given instrument from recorded data.

2.2.1.4 Audio content processing. Currently, a very active field of research is Auditory Scene analysis (Bregman, 1990), which is conducted both from perceptual and computational points of view. This research is conducted mostly within the cognitive neurosciences community. But a multidisciplinary approach would allow the translation of its fundamental research advances to many practical applications. For instance, as soon as robust results emerge from this field, it will be possible to approach (re)synthesis from a higher-level sound-object perspective, permitting us to identify, isolate, transform and recombine sound objects in a flexible way. Sound synthesis and manipulation using spectral models is based on *features* emerging from audio analysis. The use of auditory scene representations for sound manipulation and synthesis could be based on *sound objects* captured from the analysis. This possibility offers great prospects for music, sound and media production. With the current work on audio content analysis, we can start identifying and processing higher-level elements in an audio signal. For example, by identifying the rhythm of a song, a time-stretching technique can become a rhythm-changing system, and by identifying chords, a pitch shifter might be able to transpose the key of the song.

3. Interaction

In this section we review a variety of research issues that address interaction with sound and music. Three main topics are considered: Music Interfaces, Performance Modelling and Control, and Sound Interaction Design. Music interfaces is quite a well-established topic which deals with the design of controllers for music performance. Performance modelling and control is an area that has been quite active in the last decade. It has focused on the study of the performance of classical music but more recently is opening up to new challenges. The last topic covered under the interaction heading is sound interaction design. This is a brand new area that opens up many new research problems not previously addressed within the SMC research community.

3.1 Music interfaces

Digital technologies have revolutionized the development of new musical instruments, not only because of the sound generation possibilities of the digital systems, but also because the concept of ''musical instrument'' has changed with the use of these technologies. In most acoustic instruments, the separation between the control interface and the sound-generating subsystems is fuzzy and unclear. In the new digital instruments, the gesture

controller (or input device) that takes the control information from the performer(s) is always separate from the sound generator. For exact and repeatable control of a synthesizer, or a piece of music, a computer-based notation program gives a stable environment (see, e.g. Kuuskankare & Laurson, 2006). For real-time control the controlling component can be a simple computer mouse, a computer keyboard or a MIDI keyboard, but with the use of sensors and appropriate analogue-to-digital converters, any signal coming from the outside can be converted into control messages intelligible to the digital system. A recent example is music interfaces enabling control through expressive full-body movement and gesture (Camurri et al., 2005). The broad accessibility of devices, such as video cameras and analog-to-MIDI interfaces, provides a straightforward means for the computer to access sensor data. The elimination of the physical dependencies has meant that all previous construction constraints in the design of digital instruments have been relaxed (Jordà, 2005).

A computer-augmented instrument takes an existing instrument as its base and uses sensors and other instrumentation to pick up as much information as possible from the performer's motions. The computer uses both the original sound of the instrument and the feedback from the sensor array to create and/or modify new sounds. Augmented instruments are often called hyper-instruments after the work done at MIT's Media Lab (Paradiso, 1997), which aimed at providing virtuoso performers with controllable means of amplifying their gestures, suggesting coherent extensions to instrumental playing techniques.

One of the new paradigms of digital instruments is the idea of collaborative performance and of instruments that can be performed by multiple players. In this type of instrument, performers can take an active role in determining and influencing not only their own musical output but also that of their collaborators. These music collaborations can be achieved over networks such as the Internet, and the study of network or distributed musical systems is a new topic on which much research is being carried out (Barbosa, 2006).

Most current electronic music is being created and performed with laptops, turntables and controllers that were not really designed to be used as music interfaces. The mouse has become the most common music interface, and several of the more radical and innovative approaches to real-time performance are currently found in the apparently more conservative area of screen-based and mouse-controlled software interfaces. Graphical interfaces may be historically freer and better suited to unveiling concurrent, complex and unrelated musical processes. Moreover, interest in gestural interaction with sound and music content and in gestural control of digital music instruments is emerging as part of a more general trend towards research on gesture analysis, processing and synthesis. This growing importance is demonstrated by the fact that the Gesture Workshop series of conferences recently included sessions on gesture in music and the performing arts. Research on gesture not only enables a deeper investigation of the mechanisms of human–human communication, but may also open up unexplored frontiers in the design of a novel generation of multimodal interactive (music) systems.

A recent trend around new music interfaces and digital instruments is that they are more and more designed for interaction with non-professional users. The concepts of *active experience* and *active listening* are emerging, referring to the opportunity for beginners, naïve and inexperienced users, in a collaborative framework, to interactively operate on music content, by modifying and moulding it in real-time while listening. The integration of research on active listening, context-awareness, gestural control is leading to new creative forms of interactive music experience in context-aware (mobile) scenarios, resulting in an embodiment and control of music content by user behaviour, e.g. gestures and actions (for a recent example see Camurri et al., 2007).

3.1.1 Music interfaces: key issues

3.1.1.1 Design of innovative multimodal music interfaces. A key target for designers of future interactive music systems is to endow them with natural, intelligent and adaptive multimodal interfaces which exploit the ease and naturalness of ordinary physical gestures in everyday contexts and actions. Examples are tangible interfaces (e.g. Ishii & Ulmer, 1997) and their technological realization as Tangible Acoustic Interfaces (TAIs), which exploit the propagation of sound in physical objects in order to locate touching positions. TAIs are a very promising interface for future interactive music systems. They have recently been enhanced with algorithms for multimodal high-level analysis of touching gestures so that information can be obtained about how the interface is touched (e.g. forcefully or gently). Despite such progress, currently available multimodal interfaces still need improvements. A key issue is to develop interfaces that can grab subtler high-level information. For example, research has been devoted to multimodal analysis of basic emotions (e.g. happiness, fear, sadness, anger), but we are still far from modelling more complex phenomena such as engagement, empathy, entrainment. Moreover, current multimodal interfaces usually are not context-aware, i.e. they analyse users' gestures and their expressiveness, but they do not take into account the context in which the gestures are performed. Another key issue is related to scalability. Current multimodal interfaces often require special purpose set-ups including positioning of video cameras and careful preparation of objects e.g. for TAIs. Such systems are often not scalable

and difficult to port in the home and in the personal environment. A major research challenge is to exploit future mobile devices, the sensors they will be endowed with, and their significantly increased computational power and wireless communication abilities.

3.1.1.2 Integration of control with sound generation. The separation between gesture controllers and output generators has some significant negative consequences, the most obvious being the reduction of the "feel" associated with producing a certain kind of sound. Another frequent criticism is the inherent limitations of MIDI, the protocol that connects these two components of the instrument chain. A serious attempt to overcome these limitations is provided by the UDP-based Open Sound Control (OSC) protocol (Wright, 2005). However, there is a more basic drawback concerning the conceptual and practical separation of new digital instruments into two separated components: it becomes hard – or even impossible – to design highly sophisticated control interfaces without a profound prior knowledge of how the sound or music generators will work. Generic, non-specific music controllers tend to be either too simple, mimetic (imitating traditional instruments), or too technologically biased. They can be inventive and adventurous, but their coherence cannot be guaranteed if they cannot anticipate what they are going to control (Jordà, 2005).

3.1.1.3 Feedback systems. When musicians play instruments, they perform certain actions with the expectation of achieving a certain result. As they play, they monitor the behaviour of their instrument and, if the sound is not quite what they expect, they will adjust their actions to change it. In other words, they have effectively become part of a control loop, constantly monitoring the output from their instrument and subtly adjusting bow pressure, breath pressure or whatever control parameter is appropriate. The challenge is to provide the performer of a digital instrument with the appropriate feedback to control the input parameters better than that provided by mere auditory feedback. One proposed solution is to make use of the musician's existing sensitivity to the relationship between an instrument's "feel" and its sound with both haptic and auditory feedback (O'Modhrain, 2000). Other solutions may rely on visual and auditory feedback (Jordà, 2005).

3.1.1.4 Designing effective interaction metaphors. Beyond the two previous issues, which concern the musical instrument paradigm, the design of structured and dynamic interaction metaphors, enabling users to exploit sophisticated gestural interfaces, has the potential to lead to a variety of music and multimedia applications beyond the musical instrument metaphor. The state-of-the-art practice mainly consists of direct and strictly causal gesture/sound associations, without any dynamics or evolutionary behaviour. However, research is now shifting toward higher-level indirect strategies (Visell and Cooperstock, 2007): these include reasoning and decision-making modules related to rational and cognitive processes, but they also take into account perceptual and emotional aspects. Music theory and artistic research in general can feed SMC research with further crucial issues. An interesting aspect, for instance, is the question of expressive autonomy (Camurri et al., 2000), that is, the degree of freedom an artist leaves to a performance involving an interactive music system.

3.1.1.5 Improving the acceptance of new interfaces. The possibilities offered by digital instruments and controllers are indeed endless. Almost anything can be done and much experimentation is going on. Yet the fact is that there are not that many professional musicians who use them as their main instrument. No recent electronic instrument has reached the (limited) popularity of the Theremin or the Ondes Martenot, invented in 1920 and 1928, respectively.[1] Successful new instruments exist, but they are not digital, not even electronic. The most recent successful instrument is the turntable, which became a real instrument in the early eighties when it started to be played in a radically unorthodox and unexpected manner. It has since then developed its own musical culture, techniques and virtuosi. For the success of new digital instruments, the continued study of sound control, mapping, ergonomics, interface design and related matters is vital. But beyond that, what is required is integral studies that consider not only ergonomic but also psychological, social and, above all, musical issues.

3.2 Performance modelling and control

A central activity in music is performance, that is, the act of interpreting, structuring, and physically realizing a work of music by playing a musical instrument. In many kinds of music – particularly so in Western art music – the performing musician acts as a kind of mediator: a mediator between musical idea and instrumental realization, between written score and musical sound, between composer and listener/audience. Music performance is a complex activity involving physical, acoustic, physiological, psychological, social and artistic issues. At the same time, it is also a deeply human activity, relating to emotional as well as cognitive and artistic categories.

[1]By "new electronic instrument", we here mean instruments that not only produce sound in an electronic way, but also offer some new kind of interface or way of interacting with the sound. Of course, instruments like the Korg M1 synthesizer have been very successful. But in a way, they are traditional music interfaces, with a computer-based sound-producing mechanism behind the scenes.

Understanding the emotional, cognitive and also (bio-)mechanical mechanisms and constraints governing this complex human activity is a prerequisite for the design of meaningful and useful music interfaces (see above) or more general interfaces for interaction with expressive media such as sound (see next section). Research in this field ranges from studies aimed at understanding expressive performance to attempts at modelling aspects of performance in a formal, quantitative and predictive way.

Quantitative, empirical research on expressive music performance dates all the way back to the 1930s, to the pioneering work by Seashore and colleagues in the US. After a period of neglect, the topic experienced a veritable renaissance in the 1970s, and music performance research is now thriving and highly productive (a comprehensive overview can be found in Gabrielsson, 2003).

Historically, research in (expressive) music performance has focused on finding general principles underlying the types of expressive "deviations" from the musical score (e.g. in terms of timing, dynamics and phrasing) that are a hallmark of expressive interpretation. Three different research strategies can be discerned (see De Poli, 2004; Widmer & Goebl, 2004, for recent overviews on expressive performance modelling): (1) acoustic and statistical analysis of performances by real musicians – the so-called analysis-by-measurement method; (2) making use of interviews with expert musicians to help translate their expertise into performance rules – the so-called analysis- by-synthesis method; and (3) inductive machine learning techniques applied to large databases of performances.

Studies along these lines by a number of research teams around the world have shown that there are significant regularities that can be uncovered in these ways, and computational models of expressive performance (of mostly classical music) have proved to be capable of producing truly musical results These achievements are currently inspiring a great deal of research into more comprehensive computational models of music performance and also ambitious application scenarios.

One such new trend is quantitative studies into the individual style of famous musicians. Such studies are difficult because the same professional musician can perform the same score in very different ways (cf. commercial recordings by Vladimir Horowitz and Glenn Gould). Recently, new methods have been developed for the recognition of music performers and their style, among them the fitting of performance parameters in rule-based performance models and the application of machine learning methods for the identification of the performance style of musicians. Recent results of specialized experiments show surprising artist recognition rates (e.g. Saunders et al., 2004).

So far, music performance research has been mainly concerned with describing detailed performance variations in relation to musical structure. However, there has recently been a shift towards high-level musical descriptors for characterizing and controlling music performance, especially with respect to emotional characteristics. For example, it has been shown that it is possible to generate different emotional expressions of the same score by manipulating rule parameters in systems for automatic music performance (Bresin & Friberg, 2000).

Interactive control of musical expressivity is traditionally the task of the conductor. Several attempts have been made to control the tempo and dynamics of a computer-played score with some kind of gesture input device. For example, Friberg (2006) describes a method for interactively controlling, in real time, a system of performance rules that contain models for phrasing, micro-level timing, articulation and intonation. With such systems, high-level expressive control can be achieved. Dynamically controlled music in computer games is another important future application.

Visualization of musical expressivity, though perhaps an unusual idea, also has a number of useful applications. In recent years, a number of efforts have been made in the direction of new display forms of expressive aspects of music performance. Langner and Goebl (2003) have developed a method for visualizing an expressive performance in a tempo-loudness space: expressive deviations leave a trace on the computer screen in the same way as a worm does when it wriggles over sand, producing a sort of "fingerprint" of the performance. This and other recent methods of visualization can be used for the development of new multi-modal interfaces for expressive communication, in which expressivity embedded in audio is converted into visual representation, facilitating new applications in music research, music education and HCI, as well as in artistic contexts. A visual display of expressive audio may also be desirable in environments where audio display is difficult or must be avoided, or in applications for hearing-impaired people.

For many years, research in Human–Computer Interaction in general and in sound and music computing in particular was devoted to the investigation of mainly "rational", abstract aspects. In the last ten years, however, a great number of studies have emerged which focus on emotional processes and social interaction in situated or ecological environments. Examples are the research on Affective Computing at MIT (Picard, 1997) and research on KANSEI Information Processing in Japan (Hashimoto, 1997). The broad concept of "expressive gesture", including music, human movement and visual (e.g. computer animated) gesture, is the object of much contemporary research.

3.2.1 Performance modelling and control: key issues

3.2.1.1 A deeper understanding of music performance. Despite some successes in computational performance modelling, current models are extremely limited and simplistic *vis-à-vis* the complex phenomenon of musical expression. It remains an intellectual and scientific challenge to probe the limits of formal modelling and rational characterization. Clearly, it is strictly impossible to arrive at complete predictive models of such complex human phenomena. Nevertheless, work towards this goal can advance our understanding and appreciation of the complexity of artistic behaviours. Understanding music performance will require a combination of approaches and disciplines – musicology, AI and machine learning, psychology and cognitive science.

For cognitive neuroscience, discovering the mechanisms that govern the understanding of music performance is a first-class problem. Different brain areas are involved in the recognition of different performance features. Knowledge of these can be an important aid to formal modelling and rational characterization of higher order processing, such as the perceptual differentiation between human-like and mechanical performances. Since music making and appreciation is found in all cultures, the results could be extended to the formalization of more general cognitive principles.

3.2.1.2 Computational models for artistic music performance. The use of computational music performance models in artistic contexts (e.g. interactive performances) raises a number of issues that have so far only partially been faced. The concept of a creative activity being predictable and the notion of a direct ''quasi-causal'' relation between the musical score and a performance are both problematic. The unpredictable intentionality of the artist and the expectations and reactions of listeners are neglected in current music performance models. Surprise and unpredictability are crucial aspects in an active experience such as a live performance. Models considering such aspects should take account of variables such as performance context, artistic intentions, personal experiences and listeners' expectations.

3.2.1.3 Music interaction models in multimedia applications. There will be an increasing number of products which embed possibilities for interaction and expression in the rendering, manipulation and creation of music. In current multimedia products, graphical and musical objects are mainly used to enrich textual and visual information. Most commonly, developers focus more on the visual rather than the musical component, the latter being used merely as a realistic complement or comment to text and graphics. Improvements in the human–machine interaction field have largely been matched by improvements in the visual component, while the paradigm of the use of music has not changed adequately. The integration of music interaction models in the multimedia context requires further investigation, so that we can understand how users can interact with music in relation to other media. Two particular research issues that need to be addressed are models for the analysis and recognition of users' expressive gestures, and the communication of expressive content through one or more non-verbal communication channels mixed together.

3.3 Sound interaction design

Sound-based interactive systems can be considered from several points of view and several perspectives: content creators, producers, providers and consumers of various kinds, all in a variety of contexts. Sound is becoming more and more important in interaction design, in multimodal interactive systems, in novel multimedia technologies which allow broad, scalable and customized delivery and consumption of active content. In these scenarios, some relevant trends are emerging that are likely to have a deep impact on sound related scientific and technological research in the coming years. Thanks to research in Auditory Display, Interactive Sonification and Soundscape Design, sound is becoming an increasingly important part of Interaction Design and Human–Computer Interaction.

Auditory Display is a field that has already reached some kind of consolidated state. A strong community in this field has been operating for more than twenty years (see http://www.icad.org/). Auditory Display and Sonification are about giving audible representation to information, events and processes. Sound design for conveying information is, thus, a crucial issue in the field of Auditory Display. The main task of the sound designer is to find an effective mapping between the data and the auditory objects that are supposed to represent them in a way that is perceptually and cognitively meaningful. Auditory warnings are perhaps the only kind of auditory displays that have been thoroughly studied and for which solid guidelines and best design practices have been formulated. A milestone publication summarizing the multifaceted contributions to this subdiscipline is the book edited by Stanton and Edworthy (1999).

If Sonification is the use of non-speech audio to perceptualize information, *Interactive Sonification* is a more recent specialization that takes advantage of the increasing diffusion of sensing and actuating technologies. The listener is actively involved in a perception/action loop, and the main objective is to generate a sonic feedback which is coherent with physical interactions performed with sonically-augmented artifacts. This allows active exploration of information spaces and

more engaging experiences. A promising approach is Model Based Sonification (Hermann & Ritter, 2005) which uses sound modelling techniques in such a way that sound emerges as an organic product of interactions among modelling blocks and external agents. Often, interaction and sound feedback are enabled by physically-based models. For example, the user controls the inclination of a stick, and a virtual ball rolls over it producing a sound that reveals the surface roughness and situations of equilibrium (Rath & Rocchesso, 2005). While building these interactive objects for sonification, it is soon realized that fidelity to the physical phenomena is not necessarily desirable. Sound models are often more effective if they are "minimal yet veridical" (Rocchesso et al., 2003), or if they exaggerate some traits as is done by cartoonists.

A third emerging area of research with strong implications for social life, whose importance is astonishingly underestimated, is that of sound in the environment – on different scales, from architectonic spaces to urban contexts and even to truly geographical dimensions. *Soundscape Design* as the auditory counterpart of landscape design is the discipline that studies sound in its environmental context, from both naturalistic and cultural viewpoints. It is going to become more and more important in the context of the acoustically saturated scenarios of our everyday life. Concepts such as "clear hearing" and hi-fi versus lo-fi soundscapes, introduced by Murray Schafer (1994), are becoming crucial as ways of tackling the "composition" of our acoustic environment in terms of appropriate sound design.

3.3.1 Sound interaction design: key issues

3.3.1.1 Evaluation methodologies for sound design. Before sound interaction design, there is sound design. And it is worth asking whether this latter is a mature discipline in the sense that design itself is. Is there anybody designing sounds with the same attitude that Philippe Starck designs a lemon squeezer? What kind of instruments do we have at our disposal for the objective evaluation of the quality and the effectiveness of sound products in the context, for example, of industrial design? As a particular case, sound product design is rapidly acquiring a more and more relevant place in the loop of product implementation and evaluation. Various definitions of sound quality have been proposed and different evaluation parameters have been put forward for deriving quantitative predictions from sound signals (Lyon, 2000). The most commonly used parameters (among others) are loudness, sharpness, roughness and fluctuation strength. Loudness is often found to be the dominant measurable factor that adversely affects sound quality. However, more effective and refined measurement tools for defining and evaluating the aesthetic contents and the functionality of a sound have not yet been devised. The development of appropriate methodologies of this kind is an urgent task for the growth of Sound Design as a mature discipline.

3.3.1.2 Everyday listening and interactive systems. In the field of human–computer interaction, auditory icons have been defined as "natural" audio messages that convey information and feedback about events in an intuitive way. The concepts of auditory icons and "Everyday Listening", as opposed to "Musical Listening", were introduced by William Gaver (1994). The notion of auditory icons is situated within a more general philosophy of an ecological approach to perception. The concept of auditory icons is to use natural and everyday sounds to represent actions and sounds within an interface. In this context, a relevant consideration emerges: a lot of research effort has been devoted to the study of musical perception, while our auditory system is first of all a tool for interacting with the outer world in everyday life. When we consciously listen to or more or less unconsciously hear "something" in our daily experience, we do not really perceive and recognize sounds but rather events and sound sources. Both from a perceptual point of view (sound to sense) and from a modelling/generation point of view (sense to sound), a great effort is still required to achieve the ability to use sound in artificial environments in the same way that we use sound feedback to interact with our everyday environment.

3.3.1.3 Sonification as art, science, and practice. Sonification, in its very generic sense of information representation by means of sound, is still an open research field. Although a lot of work has been done, clear strategies and examples of how to design sound in order to convey information in an optimal way have only partially emerged. Sonification remains an open issue which involves communication theory, sound design, cognitive psychology, psychoacoustics and possibly other disciplines. A specific question that naturally emerges is whether the expertise of composers, who are accustomed to organizing sound in time and polyphonic density, could be helpful in developing more "pleasant" (and thus effective) auditory display design. Would it be possible to define the practice of sonification in terms that are informed by the practice of musical composition? Or, more generally, is an art-technology collaboration a positive, and perhaps vital, element in the successful design of auditory displays?

Another inescapable issue is the active use of auditory displays. Sonification is especially effective with all those kinds of information that have a strong temporal basis, and it is also natural to expect that the active involvement of the receiver may lead to better understanding, discoveries and aesthetic involvement.

In interactive sonification, the user may play the role of the performer in music production. In this sense, the interpreter of a precisely prescribed music score, adding expressive nuances, or the jazz improviser jiggling here and there within a harmonic sieve could be two good metaphors for an interactive sonification process.

3.3.1.4 Sound and multimodality. Recently, Auditory Display and Sonification research has also entered the field of multimodal and multi-sensory interaction, exploiting the fact that synchronization with other sensory channels (e.g. visual, tactile) provides improved feedback. An effective research approach to the kinds of problems that this enterprise brings up is the study of sensorial substitutions. For example, a number of sensory illusions can be used to ''fool'' the user via cross-modal interaction. This is possible because everyday experience is intrinsically multimodal and properties such as stiffness, weight, texture, curvature and material are usually determined via cues coming from more than one channel.

3.3.1.5 Soundscape design. A soundscape is not an accidental by-product of a society. On the contrary, it is a construction, a more or less conscious ''composition'' of the acoustic environment in which we live. Hearing is an intimate sense similar to touch: the acoustic waves are a mechanical phenomenon and they ''touch'' our hearing apparatus. Unlike eyes, the ears do not have lids. It is thus a delicate and extremely important task to take care of the sounds that form the soundscape of our daily life. However, the importance of the soundscape remains generally unrecognized and a process of education which would lead to more widespread awareness is urgently needed.

4. Music

This section reviews research aimed at understanding, describing and generating music. This area includes several very difficult problems which are a long way from being solved and will definitely require multidisciplinary approaches. All the disciplines involved in SMC have something to say here. Humanities and engineering approaches are required and scientific and artistic methodologies are also needed.

4.1 Music description and understanding

Music is central to all human societies. Moreover, there is an increasing belief that interaction with musical environments and the use of music as a very expressive medium for communication helped the evolution of cognitive abilities specific to humans (Zatorre, 2005). Despite the ubiquity of music in our lives, we still do not fully understand, and cannot completely describe, the musical communication chain that goes from the generation of physical energy (sound) to the formation of meaningful entities in our minds via the physiology of the auditory system.

An understanding of what music is and how it functions is of more than just academic interest. In our society, music is a commercial commodity and a social phenomenon. Understanding how music is perceived, experienced, categorized and enjoyed by people would be of great practical importance in many contexts. Equally useful would be computers that can ''understand'' (perceive, categorize, rate, etc.) music in ways similar to humans.

In the widest sense, then, the basic goal of SMC in this context is to develop veridical and effective computational models of the whole music understanding chain, from sound and structure perception to the kinds of high-level concepts that humans associate with music – in short, models that relate the physical substrate of music (the sound) to mental concepts invoked by music in people (the ''sense''). In this pursuit, SMC draws on research results from many diverse fields which are related either to the sound itself (physics, acoustics), to human perception and cognition (psycho-acoustics, empirical psychology, cognitive science), or to the technical/algorithmic foundations of computational modelling (signal processing, pattern recognition, computer science, Artificial Intelligence). Neurophysiology and the brain sciences are also displaying increasing interest in music (Zatorre, 2005), as part of their attempts to identify the brain modules involved in the perception of musical stimuli, and the coordination between them.

With respect to computational models, we currently have a relatively good understanding of the automatic identification of common aspects of musical structure (beat, rhythm, harmony, melody and segment structure) at the symbolic level (i.e. when the input to be analysed is musical scores or atomic notes) (Temperley, 2004). Research is now increasingly focusing on how musically relevant structures are identified directly from the audio signal. This research on musically relevant audio descriptors is driven mainly by the new application field of Music Information Retrieval (MIR) (Orio, 2006). Currently available methods fall short as veridical models of music perception (even of isolated structural dimensions), but they are already proving useful in practical applications (e.g. music recommendation systems).

In contrast to these bottom-up and reductionist approaches to music perception modelling, we can also observe renewed interest in more ''holistic'' views of music perception which stress the importance of considering music as a whole instead of the sum of simple

structural features (see, e.g. Serafine (1988), who argues that purely structural features, such as rhythm or harmony, may have their roots in music theory rather than in any psychological reality). Current research also tries to understand music perception and action not as abstract capacities, but as "embodied" phenomena that happen in, and can only be explained with reference to, the human body (Leman, 2008). Generally, many researchers feel that music understanding should address higher levels of musical description related, for example, to kinaesthetic/synaesthetic and emotive/affective aspects. A full understanding of music would also have to include the subjective and cultural contexts of music perception, which means going beyond an individual piece of music and describing it through its relation to other music and even extra-musical contexts (e.g. personal, social, political and economic). Clearly, computational models at that level of comprehensiveness are still far in the future.

4.1.1 Music description and understanding: key issues

4.1.1.1 "Narrow" SMC versus multidisciplinarity research. As noted above, many different disciplines are accumulating knowledge about aspects of music perception and understanding, at different levels (physics, signal, structure, "meaning"), from different angles (abstract, physiological, cognitive, social), and often with different terminologies and goals. For computational models to truly capture and reproduce human-level music understanding in all (or many) of its facets, SMC researchers will have to learn to acquaint themselves with this very diverse literature (more so than they currently do) and actively seek alliances with scholars from these other fields – in particular from the humanities, which often seem far distant from the technology-oriented field of SMC.

4.1.1.2 Reductionist versus multi-dimensional models. Quantitative-analytical research like SMC tends to be essentially reductionist, cutting up a phenomenon into individual parts and dimensions, and studying these more or less in isolation. In SMC-type music perception modelling manifests itself in isolated computational models of, for example, rhythm parsing, melody identification and harmony extraction, with rather severe limitations. This approach neglects, and fails to take advantage of, the interactions between different musical dimensions (e.g. the relation between sound and timbre, rhythm, melody, harmony, harmonic rhythm and perceived segment structure). It is likely that a "quantum leap" in computational music perception will only be possible if SMC research manages to transcend this approach and move towards multi-dimensional models which at least begin to address the complex interplay of the many facets of music.

4.1.1.3 Bottom-up versus top-down modelling. There is still a wide gap between what can currently be recognized and extracted from music audio signals and the kinds of high-level, semantically meaningful concepts that human listeners (with or without musical training or knowledge of theoretical music vocabulary) associate with music. Current attempts at narrowing this "semantic gap" via, for example, machine learning, are producing sobering results. One of the fundamental reasons for this lack of progress seems to be the more or less strict bottom-up approach currently being taken, in which features are extracted from audio signals and ever higher-level features or labels are then computed by analysing and aggregating these features. This may be sufficient for associating broad labels like genre to pieces of music (as, e.g. in Tzanetakis & Cook, 2002), but already fails when it comes to correctly interpreting the high-level structure of a piece, and definitely falls short as an adequate model of higher-level cognitive music processing. This inadequacy is increasingly being recognized by SMC researchers, and the coming years are likely to see an increasing trend towards the integration of high-level expectation (e.g. Huron, 2006) and (musical) knowledge in music perception models. This, in turn, may constitute a fruitful opportunity for musicologists, psychologists and others to enter the SMC arena and contribute their valuable knowledge.

4.1.1.4 Understanding the music signal versus understanding music in its full complexity. Related to the previous issue is the observation that music perception takes place in a rich context. "Making sense of" music is much more than decoding and parsing an incoming stream of sound waves into higher-level objects such as onsets, notes, melodies and harmonies. Music is embedded in a rich web of cultural, historical, commercial and social contexts that influence how it is interpreted and categorized. That is, many qualities or categorizations attributed to a piece by listeners cannot solely be explained by the content of the audio signal itself. It is thus clear that high-quality automatic music description and understanding can only be achieved by also taking into account information sources that are external to the music. Current research in Music Information Retrieval is taking the first cautious steps in that direction by trying to use the Internet as a source of "social" information about music ("community meta-data"). Much more thorough research into studying and modelling these contextual aspects is to be expected. Again, this will lead to intensified and larger scale cooperation between SMC proper and the human and social sciences.

4.2 Music generation modelling

Due to its symbolic nature – close to the natural computation mechanisms available on digital computers –

music generation was among the earliest tasks assigned to a computer, possibly pre-dating any sound generation attempts (which are related to signal processing). The first well-known work generated by a computer, Lejaren Hiller's *Illiac Suite* for string quartet, was created by the author (with the help of Leonard Isaacson) in 1955–1956 and premiered in 1957. At the time, digital sound generation was no more than embryonic (and for that matter, analog sound generation was very much in its infancy, too). Since these pioneering experiences, the computer science research field of Artificial Intelligence has been particularly active in investigating the mechanisms of music creation.

Soon after its early beginnings, Music Generation Modelling split into two major research directions, embracing compositional research on one side and musicological research on the other. While related to each other, these two sub-domains pursue fundamentally different goals. In more recent times, the importance of a third direction, mathematical research on music creation modelling, has grown considerably, perhaps providing the necessary tools and techniques to fill in the gap between the above disciplines.

Music generation modelling has enjoyed a wide variety of results of very different kinds in the compositional domain. These results obviously include art music, but they certainly do not confine themselves to that realm. Research has included algorithmic improvisation, installations and even algorithmic *Muzak* creation. Algorithmic composition applications can be divided into three broad modelling categories: modelling traditional compositional structures, modelling new compositional procedures, and selecting algorithms from extramusical disciplines (Supper, 2001). Some strategies of this last type have been used very proficiently by composers to create specific works. These algorithms are generally related to self-similarity (a characteristic that is closely related to that of "thematic development", which seems to be central to many types of music) and they range from genetic algorithms to fractal systems, from cellular automata to swarm models and co-evolution. In this same category, a persistent trend towards using biological data to generate compositional structures has developed since the 1960s. Using brain activity (through EEG measurements), hormonal activity, human body dynamics and the like, there has been a constant attempt to equate biological data with musical structures (Miranda et al., 2003). Another use of computers for music generation has been in "computer-assisted composition". In this case, computers do not generate complete scores. Rather, they provide mediation tools to help composers manage and control some aspects of musical creation. Such aspects may range, according to the composers' wishes, from high-level decision-making processes to minuscule details. While computer assistance may be a more practical and less "generative" use of computers in musical composition, it is currently enjoying a much wider uptake among composers.

The pioneering era of music generation modelling has also had a strong impact on musicological research. Ever since Hiller's investigations and works, the idea that computers could model and possibly re-create musical works in a given style has become widely diffused through contemporary musicology. Early ideas were based on generative grammars applied to music. Other systems, largely based on AI techniques, have included knowledge based systems, neural networks and hybrid approaches (Papadopoulos & Wiggins, 1999; Cope, 2005).

Early mathematical models for Music Generation Modelling included stochastic processes (with a special accent on Markov chains). These were followed by chaotic non-linear systems and by systems based on the mathematical theory of communication. All these models have been used for both creative and musicological purposes. In the last 20 years, mathematical modelling of music generation and analysis has developed considerably, going some way to providing the missing link between compositional and musicological research. Several models following different mathematical approaches have been developed. They involve "enumeration combinatorics, group and module theory, algebraic geometry and topology, vector fields and numerical solutions of differential equations, Grothendieck topologies, topos theory, and statistics. The results lead to good simulations of classical results of music and performance theory. There is a number of classification theorems of determined categories of musical structures" (Mazzola, 2001).

A relevant result of mathematical modelling has been to provide a field of potential theories where the specific peculiarities of existing ones can be investigated against non-existing variants. This result creates the possibility of the elaboration of an "anthropic principle" in the historical evolution of music similar to that created in cosmology (that is: understanding whether and why existing music theories are the best possible choices or at least good ones) (Mazzola, 2001).

4.2.1 Music generation modelling: key issues

4.2.1.1 Computational models. The main issue of computational models in both the "creative" and the "problem solving" sides of Music Generation Modelling seems to relate to the failure to produce "meaningful" musical results. "... computers do not have feelings, moods or intentions, they do not try to describe something with their music as humans do. Most of human music is referential or descriptive. The reference can be something abstract like an emotion, or something more objective such as a picture or a landscape." (Papadopoulos & Wiggins, 1999). Since "meaning" in music can be

expressed – at least in part – as "planned deviation from the norm", future developments in this field will need to find a way to formalize such deviations in order to get closer to the cognitive processes that lie behind musical composition (and possibly also improvisation). In addition, "multiple, flexible, dynamic, even expandable representations [are needed] because this will more closely simulate human behaviour" (Papadopoulos & Wiggins, 1999). Furthermore, while mathematicians and computer scientists evaluate algorithms and techniques in terms of some form of efficiency – be it theoretical or computational – efficiency is only a minor concern, if any, in music composition. The attention of composers and musicians is geared towards the "quality of interaction they have with the algorithm. (...) For example, Markov chains offer global statistical control, while deterministic grammars let composers test different combinations of predefined sequences" (Roads, 1996).

4.2.1.2 Mathematical models. In a similar vein, the mathematical coherence of current compositional modelling can help understanding the internal coherence of some musical works, but it can hardly constitute, at present, an indication of musical quality at large. Mathematical coherence is only one (possibly minor) aspect of musical form, while music continues to be deeply rooted in auditory perception and psychology. The issue becomes then to merge distant disciplines (mathematics, psychology and auditory perception, to name the most relevant ones) in order to arrive at a better, but still formalized, notion of music creation.

4.2.1.3 Computer-assisted composition tools. Currently, composers who want to use computers to compose music are confronted, by and large, with two possible solutions. The first is to rely on prepackaged existing software which presents itself as a "computer-assisted composition" tool. The second is to write small or not-so-small applications that will satisfy the specific demands of a given compositional task. Solutions that integrate these approaches have yet to be found. On the one hand, composers will have to become more proficient than at present in integrating their own programming snippets into generalized frameworks. On the other, a long overdue investigation of the "transparency" (or lack thereof) of computer-assisted composition tools (Bernardini, 1985) is in order. Possibly, the current trend that considers good technology as technology that creates the illusion of non-mediation could provide appropriate solutions to this problem. In this case, however, the task will be to discover the multimodal primitives of action and perception that should be taken into consideration when creating proper mediation technologies in computer-assisted composition.

4.2.1.4 Notation and multiple interfaces. The composing environment has radically changed in the last 20 years. Today, notation devices and compositional tools inevitably involve the use of computer technology. However, the early research on new notation applications which integrated multimedia content (sound, video, etc.), expressive sound playback, graphic notation for electronic music and advanced tasks such as automatic orchestration and score reduction (Roads, 1982), remains to be exploited by composers and musicians at large. Also, little investigation has been conducted into the taxonomy of composing environments today. A related question is whether composing is still a one-(wo)man endeavour, or whether it is moving towards some more elaborate teamwork paradigm (as in films or architecture). Where do mobility, information, participation and networking technologies come into play? These questions require in-depth multidisciplinary research whose full scope is yet to be designed.

5. Concluding remarks

This article has attempted to give a rough overview of the current state of the research field of Sound and Music Computing (SMC). Efforts were made to make the survey comprehensive and informative, while keeping it reasonably compact. If the result still looks (and is!) cursory and incomplete, we ask the reader to attribute this to the extreme complexity and diversity of the field, which is exacerbated by the fact that many different scientific disciplines are – or should be – involved in SMC research. Sound and Music Computing is a thriving, energetic research field whose practical application potential and possible impact on our everyday lives are just beginning to be realized. We hope that the present article can provide some orientation to new (young) researchers entering in this exciting field.

Acknowledgments

The research that led to this document has been generously supported by the EU project (Coordination Action) Sound to Sense, Sense to Sound ($S2S^2$). The authors would like to thank Claude Cadoz, Chris Chafe and Curtis Roads for many constructive comments on earlier versions of this text.

References

Barbosa, A. (2006). *Computer-supported cooperative work for music applications*. PhD thesis, Pompeu Fabra University, Barcelona, Spain.

Bernardini, N. (1985). Semiotics and computer music composition. In *Proceedings of the International Computer Music Conference 1985*, San Francisco.

Bernardini, N. & De Poli, G. (2007). The sound and music computing field: present and future. *Journal of New Music Research*, *36*(3), 143–148.

Bilbao, S. (2007). Robust physical modeling sound synthesis for nonlinear systems. *IEEE Signal Processing Magazine*, *24*(2), 32–41.

Blauert, J. (2005). *Communication acoustics (signals and communication technology)*. Berlin: Springer.

Bonada, J. & Serra, X. (2007). Synthesis of the singing voice by performance sampling and spectral models. *IEEE Signal Processing Magazine*, *24*(2), 67–79.

Bregman, A.S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: The MIT Press.

Bresin, R. & Friberg, A. (2000). Emotional coloring of computer-controlled music performances. *Computer Music Journal*, *24*(4), 44–63.

Cadoz, C., Luciani, A. & Florens, J.-L. (1993). CORDIS-ANIMA: a modeling and simulation system for sound and image synthesis – the general formalism. *Computer Music Journal*, *17*(1), 19–29.

Camurri, A., Canepa, C. & Volpe, G. (2007). Active listening to a virtual orchestra through an expressive gestural interface: the Orchestra Explorer. In *Proceedings NIME-07 International Conference on New Interfaces for Musical Expression*, New York.

Camurri, A., Coletta, P., Ricchetti, M. & Volpe, G. (2000). Expressiveness and physicality in interaction. *Journal of New Music Research*, *29*(3), 187–198.

Camurri, A., De Poli, G., Leman, M. & Volpe, G. (2005). Toward communicating expressiveness and affect in multimodal interactive systems for performing art and cultural applications. *IEEE Multimedia Magazine*, *12*(1), 43–53.

Cano, P. (2007). *Content-based audio search: from fingerprinting to semantic audio retrieval*. PhD thesis, Pompeu Fabra University, Barcelona, Spain.

Cook, P.R. (1997). Physically informed sonic modeling (PhISM): synthesis of percussive sounds. *Computer Music Journal*, *21*(3), 38–49.

Cope, D. (2005). *Computer models of musical creativity*. Cambridge, Mass.: MIT Press.

De Poli, G. (2004). Methodologies for expressiveness modeling of and for music performance. *Journal of New Music Research*, *33*(3), 189–202.

Elhilali, M., Shamma, S., Thorpe, S. & Pressnitzer, D. (2007). Models of timbre using spectro-temporal receptive fields: investigation of coding strategies. *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain.

Faller, C. (2006). Parametric multichannel audio coding: synthesis of coherence cues. *IEEE Transactions on Audio, Speech and Language Processing*, *14*(1), 299–310.

Friberg, A. (2006). pDM: an expressive sequencer with real-time control of the KTH music performance rules. *Computer Music Journal*, *30*(1), 37–48.

Gabrielsson, A. (2003). Music performance research at the millennium. *Psychology of Music*, *31*(3), 221–272.

Gaver, W.W. (1994). Using and creating auditory icons. In *Auditory display: sonification, audification and auditory interfaces* (pp. 417–446). Reading, MA: Addison Wesley.

Hashimoto, S. (1997). KANSEI as the third target of information processing and related topics in Japan. In *Proceedings of the International Workshop on KANSEI: the technology of emotion*. Genova: Italian Computer Music Association (AIMI).

Hermann, T. & Ritter, H. (2005). Model-based sonification revisited: authors' comments on Hermann and Ritter, ICAD 2002. *ACM Transactions on Applied Perception*, *4*(2), 559–563.

Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. Cambridge, MA: MIT Press/Bradford Books.

Ishii, H. & Ullmer, B. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of CHI '97*, pp. 22–27.

Jordà, S. (2005). *Digital Lutherie: crafting musical computers for new musics performance and improvisation*. PhD thesis, Pompeu Fabra University, Barcelona, Spain.

Karjalainen, M., Mäki-Patola, T., Kanerva, A. & Huovilainen, A. (2006). Virtual air guitar. *Journal of the Audio Engineering Society*, *54*(10), 964–980.

Kim, H.-G., Moreau, N. & Sikora, T. (2005). *MPEG-7 audio and beyond: audio content indexing and retrieval*. New York: Wiley.

Kuuskankare, M. & Laurson, M. (2006). Expressive notation package. *Computer Music Journal*, *30*, 67–79.

Lane, J., Hoory, D., Martinez, E. & Wang, P. (1997). Modeling analog synthesis with DSPs. *Computer Music Journal*, *21*, 23–41.

Langner, J. & Goebl, W. (2003). Visualizing expressive performance in tempo-loudness space. *Computer Music Journal*, *27*(4), 69–83.

Leman, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.

Leman, M., Avanzini, A., de Cheveigné, A. & Bigand, E. (2007). The societal contexts for sound and music computing: research, education, industry, and socio-culture. *Journal of New Music Research*, *36*(3), 149–167.

Lindemann, E. (2007). Music synthesis with reconstructive phrase modeling. *IEEE Signal Processing Magazine*, *24*(2), 80–91.

Lyon, R.H. (2000). *Designing for product sound quality*. New York: Marcell Dekker.

Martin, K.D. (1999). *Sound-source recognition: A theory and computational model*. PhD thesis, MIT, USA.

Mazzola, G. (2001). *Mathematical Music Theory – Status Quo 2000*. Available online at: http://www.ircam.fr/equipes/repmus/mamux/documents/status.pdf

Miranda, E., Sharman, K., Kilborn, K. & Duncan, A. (2003). On harnessing the electroencephalogram for the musical braincap. *Computer Music Journal*, *27*(2), 80–102.

Mock, T. (2004). Music everywhere. *IEEE Spectrum*, *41*(9), 42–47.

O'Modhrain, M.S. (2000). *Playing by feel: incorporating haptic feedback into computer-based musical instruments*. PhD thesis, Stanford University, USA.

Orio, N. (2006). Music retrieval: a tutorial and review. *Foundations and Trends in Information Retrieval*, *1*(1), 1–90.

(2) Papadopoulos, G. & Wiggins, G. (1999). AI methods for algorithmic composition: a survey, a critical view and future prospects. In *Proceedings of the AISB'99 Symposium on Musical Creativity*.

Paradiso, J.A. (1997). Electronic music: new ways to play. *IEEE Spectrum*, *34*(12), 18–30.

Peltola, L., Erkut, C., Cook, P.R. & Välimäki, V. (2007). Synthesis of hand clapping sounds. *IEEE Transactions on Audio, Speech and Language Processing*, *15*(3), 1021–1029.

Picard, R. (1997). *Affective computing*. Cambridge, MA: MIT Press.

Pulkki, V. & Merimaa, J. (2006). Spatial impulse response rendering II: reproduction of diffuse sound and listening tests. *Journal of the Audio Engineering Society*, *54*(1), 3–20.

Rabenstein, R., Petrausch, S., Sarti, A., De Sanctis, G., Erkut, C. & Karjalainen, M. (2007). Blocked-based physical modeling for digital sound synthesis. *IEEE Signal Processing Magazine*, *24*(2), 42–54.

Rath, M. & Rocchesso, D. (2005). Continuous sonic feedback from a rolling ball. *IEEE Multimedia*, *12*(2), 60–69.

Roads, C. (1982). Interactive orchestration based on score analysis. In J. Strawn and T. Blum, (Eds), *Proceedings of the 1982 International Computer Music Conference*, Venice, Italy. San Francisco: International Computer Music Association.

Roads, C. (1996). *The computer music tutorial*. Cambridge, MA: MIT Press.

Roads, C. (2001). *Microsound*. Cambridge, MA: MIT Press.

Rocchesso, D., Bresin, R. & Fernström, M. (2003). (4) Sounding objects. *IEEE Multimedia*, 42–52.

Rocchesso, D. & Fontana, F. (Eds) (2003). *The sounding* (5) *object*. ????: Edizioni di Mondo Estremo.

Saunders, C., Hardoon, D., Shawe-Taylor, J. & Widmer, G. (2004). Using string kernels to identify famous performers from their playing style. In *Proceedings of the 15th European Conference on Machine Learning (ECML'2004)*, Pisa, Italy.

Schafer, M. (1994). *Soundscape – our sonic environment and the tuning of the world*. Rochester, Vermont.: Destiny Books.

Schwarz, D. (2007). Corpus-Based Concatenative Synthesis. *IEEE Signal Processing Magazine*, *24*(2), 92–104.

Serafine, M.L. (1988). *Music as cognition: the development of thought in sound*. New York: Columbia University Press.

Serra, X. (1997). Musical sound modeling with sinusoids plus noise. In C. Roads, S. Pope, A. Piccialli, and G. De Poli (Eds), *Musical signal processing* (pp. 91–122). Lisse, the Netherlands: Swets & Zeitlinger Publishers.

Serra, X., Bresin, R. & Camurri, A. (2007). Sound and music computing: challenges and strategies. *Journal of New Music Research*, *36*(3), 185–190.

Smith, J.O. (2006). *Physical audio signal processing: for virtual musical instruments and digital audio effects*. Available online at: http://ccrma.stanford.edu/~jos/pasp/

Stanton, N.A. & Edworthy, J. (1999). *Human factors in auditory warnings*. Aldershot, UK: Ashgate.

Sturm, B., Daudet, L. & Roads, C. (2006). Pitch-shifting audio signals using sparse atomic approximations. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*, Santa Barbara, CA, USA, pp. 45–52.

Supper, M. (2001). A few remarks on algorithmic composition. *Computer Music Journal*, *25*(1), 48–53.

Temperley, D. (2004). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Trautmann, L., Petrausch, S. & Bauer, M. (2005). Simulations of string vibrations with boundary conditions of third kind using the functional transformation method. *Journal of the Acoustical Society of America*, *118*(3), 1763–1775.

Tzanetakis, G. & Cook, P. (2002). Musical genre classifcation of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302.

Välimäki, V. & Huovilainen, A. (2006). Oscillator and filter algorithms for virtual analog synthesis. *Computer Music Journal*, *30*(2), 19–31.

Välimäki, V., Pakarinen, J., Erkut, C. & Karjalainen, M. (2006). Discrete-time modelling of musical instruments. *Reports on the Progress in Physics*, *69*(1), 1–78.

Välimäki, V., Rabenstein, R., Rocchesso, D., Serra, X., Smith, J.O. (Eds) (2007). *IEEE Signal Processing Magazine*, *24*(2), Special Issue on Signal Processing for Sound Synthesis.

Visell, Y. & Cooperstock, J. (2007). Enabling gestural interaction by means of tracking dynamical systems models and assistive feedback. In *Proceedings of the IEEE Interantional Conference on Systems, Man, and Cybernetics*, Montreal.

Widmer, G. & Goebl, W. (2004). Computational models of expressive music performance: the state of the art. *Journal of New Music Research*, *33*(3), 203–216.

Wright, M. (2005). Open sound control: an enabling technology for musical networking. *Organised Sound*, *10*(3), 193–200.

Yeh, D.T. & Smith, J.O. (2006). Discretization of the '59 Fender Bassman tone stack. In *Proceedings of the International Conference on Digital Audio Effects*, Montreal, Quebec, Canada, pp. 1–6.

Zatorre, R. (2005). Music, the food of neuroscience? *Nature*, *434*, 312–315.

Zölzer, U. (Ed.) (2002). *DAFX: digital audio effects*. New York: Wiley.