Perspectives in Education for Sound and Music Computing

Federico Avanzini, Adriano Baratè, Goffredo Haus, Luca A. Ludovico, Stavros Ntalampiras, and Giorgio Presti

LIM – Laboratorio di Informatica Musicale
Dipartimento di Informatica "Giovanni degli Antoni"
Universita` degli Studi di Milano
Via Comelico 39, 20135 Milano, Italy
{federico.avanzini, adriano.barate, goffredo.haus, luca.ludovico, stavros.ntalampiras, giorgio.presti}@unimi.it

Abstract. In 2007, a document entitled "A Roadmap for Sound and Music Computing", authored by internationally renowned experts, aimed to identify, characterize and propose strategies for tackling the key research challenges that this growing and diversified domain was expected to be facing in the next ten to fifteen years. The original idea was to establish a common agenda and ensure consolidation, integration and exploitation of research results from European initiatives and projects. Ten years later, we can reconsider those forecasts and check the achievement of the expected results. Besides, considering recent technological innovations and approaches (such as machine learning, artificial intelligence, etc.), we take a fresh look at sound and music computing education. The goal is to outline the characteristics required by academia and industries to future domain experts, who have to be multifaceted persons with interdisciplinary expertise including music, musicology, math, physics, and computer science.

Keywords. education, music, sound, teaching.

1.1 Introduction

The idea of the present work originates from "A Roadmap for Sound and Music Computing" [1] ("Roadmap" hereafter), a detailed report released in 2007 and authored by internationally renowned experts aiming to identify, characterize and propose strategies for tackling the key research challenges that sound and music computing (SMC) was expected to be facing in the next ten to fifteen years. The Roadmap foresaw that, by 2020, music would have become a commodity as ubiquitous as water or electricity, and its content and the related activities would have promoted new business ventures, which in turn would have bolstered the music and cultural/creative industries. One of its major research goals was to stimulate a fruitful interaction among culture, science and industry. Actually, many of the forecasts cited in the document have been realized thanks to the technological advancement and the collaboration between research and development groups.

The document had the merit of arousing interest within the scientific community, also eliciting some critical reactions. For example, [2] invited to consider additional topics concerning study and involvement in economic and policy analysis and stressed the importance of collaboration and tool building to support the discipline.

The idea of declaring short, medium, and long-term SMC goals is at the base of other relevant initiatives, such as the 2007–2010 IRCAM research plan [3]. Published in response to the Roadmap, it aimed to deliver an extensive and synthetic vision of the identified directions for the research at IRCAM.

The subjects addressed in the mentioned reports included sound analysis/synthesis, physical models, sound spatialization, computer-aided composition, and interdisciplinary transversal themes concerning different levels of music representations, the renewal of concepts underlying the management of time and interaction, and the elicitation and mediation of the musical knowledge.

In the meanwhile, advances in technology have significantly impacted the way in which we produce and consume music, and SMC frontiers are rapidly evolving. For instance, sound design has been shifting and enlarging its scope to those contexts and applications where interactivity is of primary importance, thus originating a new discipline known as sonic interaction design [4]. The novel field of Unconventional Computing (UC), that aims to develop new types of computers, such as harnessing biological media to implement new kinds of processors, offers new possibilities to SMC [5]. In this context, it is worth citing the musical experiments with Cellular Automata modelling and in-vitro neural networks under development at Plymouth University's Interdisciplinary Centre for Computer Music Research, that are paving the way for interactive musical biocomputers.

In this context, the research questions we will try to answer are: What are the characteristics that higher education should present in order to train the future experts in the SMC field? What are the competences and skills that the industry and the academia are expecting from them? How should an "ideal" university curriculum be conceived to foster these goals?

We will try to answer these questions from two different perspectives: first, by analyzing the evolution of music and sound computing courses in the last 10 years, thus pointing out the different trends followed by international institutions such as universities and research centers; second, by listing current approaches towards open problems and delineating the characteristics of the expert who should deal with them.

1.2 Context

The Roadmap emphasized the need for a tight link between SMC education and research, and called for a major effort in developing higher-education programs. Meanwhile, the implementation of the Bologna process in European countries has progressed, and the EHEA (EU Higher Education Area) has evolved towards a common structure of degrees [6]. This context and the related trends should be acknowledged in shaping SMC education.

Employability of graduates should be one of the concerns of any highereducation program. Training should enhance employability of graduates by providing them with complementary and transferable skills aimed at facilitating their flow into the job market. SMC research has wide applicative implications, thus strong links with industry should naturally arise in training programs. Industrial partners should be involved in training programs, with the aim of broadening as much as possible skills related to technology transfer and entrepreneurship, thus widening the career prospects of students. Although the Roadmap acknowledges this, it does not contain a thorough need analysis for training programs in SMC.

It should be kept in mind that the SMC research community is a small one. Therefore the issue of SMC training must be situated in the broader context of neighboring and/or larger and/or more established academic disciplines. The latter question is relevant at all degree levels. Within more basic and general undergraduate degrees, topics related to SMC can still be successfully employed for educational purposes in foundational courses (e.g., insert elements of audio programming on a mobile device in a first-year java programming course). Within more specialized Master (or even PhD) degrees, SMC topics can be relevant to a number of related disciplines (HCI, robotics, etc.).

1.3 Defining a Body-of-Knowledge

The Roadmap contained preliminary work finalized at defining a set of "content areas", meant to constitute core academic topics on which courses (or course modules) in SMC may be built. This work should now be revised in the light of recent trends, and should be expanded and consolidated by the scientific community in order to define the Body-of-Knowledge (BoK) needed for an undergraduate or graduate SMC curriculum, course exemplars, and other guidelines.

Similar guidelines are available for more established fields. A relevant example is provided by the curricula recommendations by ACM which, starting in the 1960's [7], has collaborated with leading professional and scientific computing societies in various efforts to establish international curricular guidelines for undergraduate and graduate programs in Computer Science [8], Computer Engineering, Information Systems, Information Technology, and Software Engineering.

Defining such a BoK for SMC curricula is a long-term goal which requires a coordinated effort by the scientific community. The BoK should answer such questions as: what are core skills that an SMC graduate can exploit in the job market? What are the core topics that need to be present in a degree in our discipline? How can these be

mapped into degree structures? How may SMC topics be applied to neighboring and more established academic disciplines? What are SMC referential textbooks?
Related to this latter point, it should be noted that almost no foundational textbooks in SMC were available in 2007. A milestone dating back to 1996 and covering all aspects of computer music is the Computer Music Tutorial book [9]. Another notable exception, and a reference example, was the Digital Audio Effects book [10], now in its second edition (the first edition was released as early as 2002): it presents the state of the art in the field, involves leading researchers, includes extensive code examples (in MATLAB), has been and still is widely used for teaching.
Ten years later the situation has changed to some extent. Many excellent textbooks have been published and are used for teaching. With no claim of being exhaustive, we can cite some examples regarding physically-based sound modeling [11], sound design [12], sonification [13] and sonic interaction design [14], machine learning for audio [15], music processing [16, 17].

1.4 Higher Education Courses for Sound and Music Computing

Appendix A of the 2007 Roadmap presented a survey of existing courses and curricula in SMC around the EU, with the aim of analyzing trends in SMC education. The document collected relevant data for both single courses and entire curricula centered on SMC, covering a total of 170 courses and 40 curricula across 15 European countries. It is worth analyzing how the situation has changed in recent times. For this reason, we decided to conduct a new survey in order to provide an updated picture of higher education courses for SMC. Our analysis takes into consideration 22 undergraduate and graduate courses implemented by leading institutions worldwide. Between 2007 and 2017, many subjects not only changed their name, but – more importantly – scope and goal. Thus, the original clustering needs to be slightly adjusted to compare the two corpora. Since these differences may lead the reader to biased observations, a 10% confidence bar was added over the plot in Figure 1.1 to minimize this effect.
The identified clusters, followed by examples of the corresponding subjects, are:
- C1. Acoustics – Acoustics of musical instruments, room acoustics, acoustic physics;
- C2. Audio signal processing and modeling – Systems, sampling and quantization, spectral and time-spectral representations, digital filters, models for sound synthesis, physics-based modeling, digital audio effects, spatial sound and virtual acoustics;
- C3. Hardware and software systems – Sensors and actuators, real-time systems, output devices, software platforms, software engineering;
- C4. Interaction and design of multimodal interfaces – Performance analysis, emotion and expression in music performance, computational models and control of music performance, multimodal perception and action, gesture and multisensory analysis and synthesis, representations of multisensory data, control mappings and interaction strategies, evaluation of interaction models, digital and virtual musical instruments, interactive performing arts, interactive installations, education, entertainment, multimedia and new media, therapy and rehabilitation;
- C5. Music information retrieval and sound analysis – Including: auditorybased

audio signal processing, perceptual coding, content-based audio processing and audio descriptors, content description/transmission languages, content-based transformation, feature extraction/classification, automatic transcription, music information retrieval, computer assisted composition;

- C6. Systematic musicology – Music semiotics, score analysis, computational models for music analysis;
- C7. Music perception and cognition – Psychoacoustics, music perception, computational approaches and models, sound-based cognition, music cognition, artificial intelligence;
- C8. Sound design and auditory display – Auditory warnings, sound in interaction design, sonification, sound design.

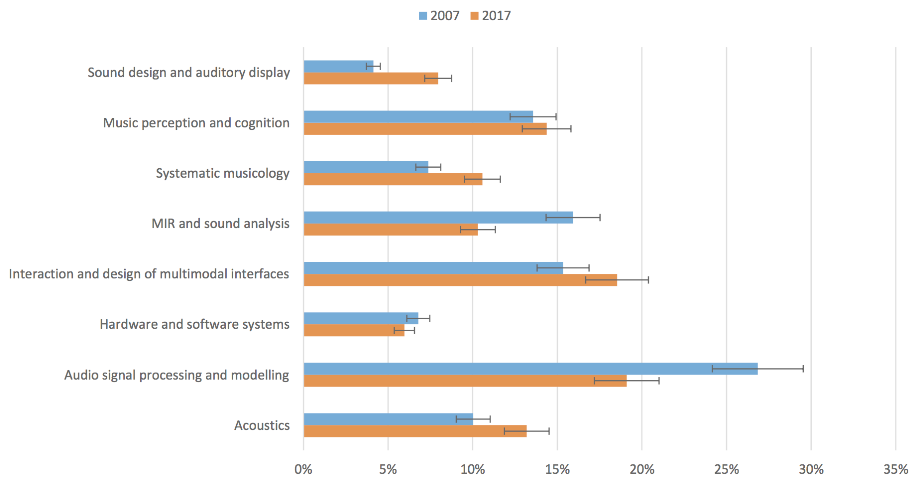

Fig. 1.1. Current balance between core SMC topics and background disciplines.

Figure 1.1 shows some surprising effects, such as the significant reduction of topics like music information retrieval and digital signal processing. A possible explanation is that the in-depth theoretical investigation occurred in the past is now leaving room, in terms of resources, to engineering and implementation aspects. Additional considerations are postponed to the next section. Please note that some theoretical and technological evolutions, envisioned in the Roadmap just as research trends under development, matured in the last decade. These include novel methodologies for systematic musicology (C6), machine learning and deep learning (C5 and C7), interactive sound for virtual and augmented reality (C4). We discuss in the detail these trends in the next section. Other emerging trends, mentioned but not deeply investigated by the Roadmap, concern music-oriented biomolecular automata (C3), and intersections with neurosciences (C7). It is worth underlining that the 2007 survey deliberately ignored some areas supposedly not relevant in SMC, being either too general or too far from a strict vision of the domain. Nevertheless, nowadays they are considered key subjects for the comprehensive education of an SMC expert. Additional subject areas may include:

- C9. Music and sound technology, dealing with multimedia-oriented generalpurpose technologies, like MIDI, music coding approaches, and mobile app programming;
- C10. Audio production and post-production, investigating such processes with the aim of improving available software tools and being able to conduct a more informed analysis of audio signals;
- C11. Communication, multimedia publishing and law, providing a marketoriented vision of SMC activities;
- C12. Sociology of music, focusing on social aspects of musical behavior and the role of music in society;
- C13. Music theory, composition and instrument studies, providing musical training and knowledge;
- C14. Computer science core subjects, aiming to improve basic IT skills and knowledge;
- C15. Math, physics and statistics, strengthening the foundations of the STEM (Science, Technology, Engineering and Mathematics) area where students often exhibit gaps in their previous knowledge.

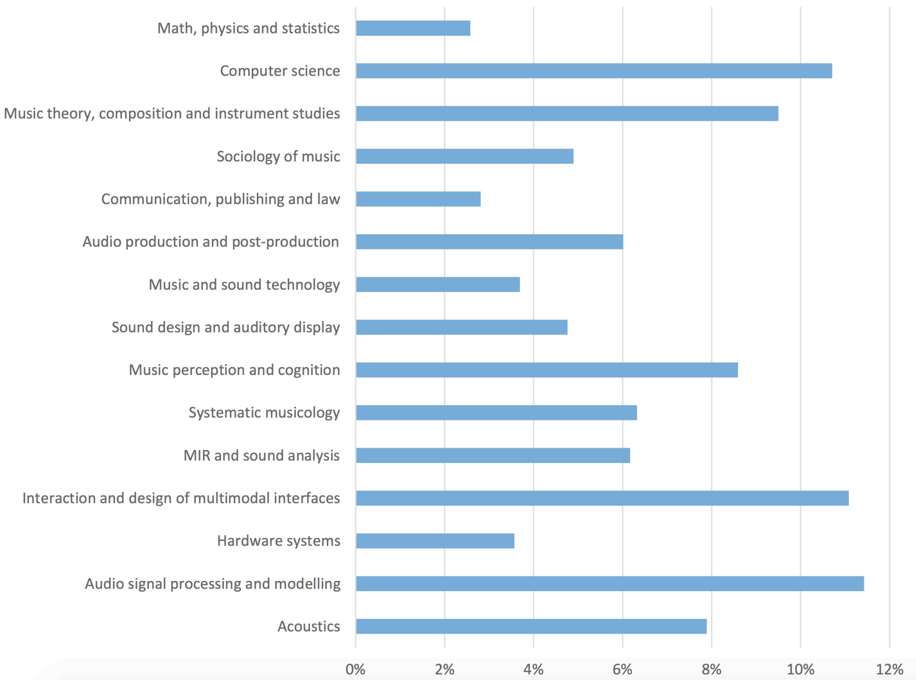

Fig. 1.2. Distribution of SMC core topics and background disciplines (including new content areas C9–C15) in the surveyed university programs.

From this revised list of skills and competences, the richness of the expected educa-

tion for an SMC expert clearly emerges. Figure 1.2 shows the balance of clustered subjects including the areas listed above. Nowadays, mastering only topics closely linked to SMC fields is not enough, rather it is necessary to have multiple skills and fluently speak the languages of music, mathematics, physics, and informatics.
Needless to say, it is very difficult to condense everything into a 5-year university degree, especially if the curriculum must be organized in a coherent 3+2 structure. Such a problem has been tackled in a number of ways by different institutions. For example, the Department of Signal Processing of the Helmut Schmidt University, Hamburg offers an educational program composed by 50% of acoustics and DSP subjects and 50% of scientific topics not directly related to music. Conversely, the University of Ghent focuses its Bachelor and Master degrees on creative disciplines, treating also cognitive music psychology, publishing-related and legal aspects. Finally, it is worth mentioning the case of the University of Milan, which offers one of the few undergraduate programs in Music Informatics.

1.5 Emerging Directions in SMC Research and Education

The comparison summarized in Fig. 1.1 shows that some content areas have been increasingly addressed in SMC curricula in the last 10 years. Courses in the area of systematic musicology (C6) have increased, similarly to those coming from the areas of sound design, interaction and multimodal interfaces, and – although to a lesser extent – perception and cognition.
One possible explanation of this trend is that emerging research directions related to these areas have gained momentum in the last decade. This correlation emphasizes the tight link between SMC education and research. In the remainder of this section we discuss recent and current developments of such research, and its implications on SMC education, emphasizing interconnections among content areas.

1.5.1 Music and Musicology Oriented education

Here we address the broad content area of systematic musicology and investigate the challenges that an SMC expert has to face when dealing with problems typical of music and musicology. Specifically, we will address the evolving fields of music notation, computational musicology, and computerassisted composition.
Concerning music notation, finding the most suitable way to encode score symbolic content in the digital domain is a fundamental goal, not only for transcription and printing purposes, but also to preserve, analyze, rework, and enjoy music information in novel ways. In the early 2000s, XML-based proposals began to emerge, coupling the advantages of plain text with the possibility to structure information in a hierarchical way, to make it easily readable by both humans and machines, and to extend supported elements when required [18]. In this sense, it is worth citing IEEE 1599, Music Encoding Initiative and MusicXML. In more recent times, the World Wide Web (W3C) consortium has launched the Music Notation Community Group[1], an

[1] https://www.w3.org/community/music-notation/

initiative that aims to unify formats syntactically and semantically different in order to establish the guidelines for a standardized approach over the Web. Ten years after the approval of the IEEE 1599 format as an international standard in 2008, the goal is still the creation of a commonly-accepted music representation format able not only to encode notation in all its forms, but also to integrate the other layers music information is made of [19], such as graphical and audio aspects. This idea, already contained in the original IEEE 1599 proposal, is the core of the MNX format currently under development by the W3C Music Notation Community Group[2].

An expert of music notation in the digital domain has to master multiple competences and skills, including the capability to read and understand different kinds of notation (music theory skills, cluster C13), transcribe them through already available computer programs (user skills, cluster C9), develop new software tools when the existing ones are not sufficient (programming skills, cluster C14), and above all thinking in a multi-layer way (openmindedness). Moreover, the concept of multi-layer representation of music information, taken to the extreme, can involve even more areas, such as cluster C4 to offer an effective user experience, C5 concerning music information retrieval, C6 to extract structural information through music analysis, and C11 to understand the impact of new multimedia products over the market [20].

Concerning computational musicology, many mathematical and algorithmic approaches could be mentioned. A recent trend is the adoption of the Tonnetz, a graph used in computational musicology to describe the harmonic relationships of notes in equal tuning [21]. Figure 1.3 shows a graphical representation of a Tonnetz, originally introduced as a 3×4 matrix, where values represent pitch classes, and the matrix itself shows their relationships. This model has been largely generalized to several formalisms, one of them interprets it as a simplicial complex.

In order to capture both the temporal and harmonic information encoded in a musical phrase, an associated Tonnetz can consider the played notes together with their duration and repetition, linking to each vertex of the Tonnetz a non-negative real number that represents how long the associated pitches have been played during the execution of the phrase. In this case a metric representation of music as a planar polyhedral surface is obtained [22]. The musical meaning of this topological representation of music can be applied to music analysis and classification [23].

As demonstrated by the example above, the expert who wants to master modern musicological approaches and techniques has to acquire knowledge in different fields, such as C5, C6 and C13, presenting also a strong background in the C15 area.

[2] https://www.w3.org/community/music-notation/2016/05/19/introducing-mnx/

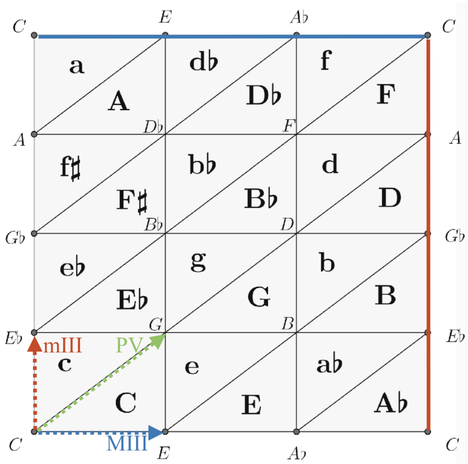

Fig. 1.3. A finite subcomplex of the Tonnetz T.

Even in more "traditional" contexts, such as the formalization of music structures and computer-assisted composition, multiple skills are required. An example of in-use formalism is represented by Petri nets, an abstract and formal approach that aims to capture the dynamic behavior of a system with asynchronous and concurrent activities [24]. This model can be viewed as a directed bipartite graph with two kinds of nodes: transitions (i.e. events that may occur), graphically represented through rectangles, and places (i.e. conditions), drawn as circles. Transitions are linked to places, and vice versa, by directed arcs, that describe which places are preand/or post-conditions for which transitions. Places in a Petri net contain a number of marks called tokens, while the upper limit of tokens that a given place can host represents its capacity. The current place marking and its capacity are conventionally indicated by an upper and a lower number in the circle, respectively. An example of graphical representation is shown in Fig. 1.4.

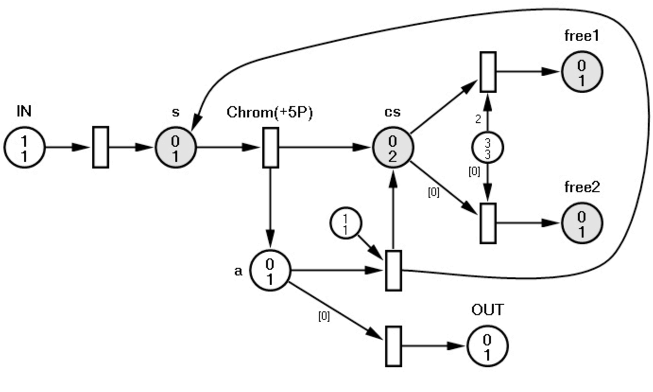

Fig. 1.4. An example of a Petri Net model.

When there are sufficient tokens in all of input places of a transition and there is sufficient room in all of its output places to host newly generated tokens, the transition may fire, subtracting tokens from the input places and creating tokens in its output places. Thanks to their characteristics, Petri nets are well suited for modeling the behavior of distributed systems. The modeled processes can include choice, iteration, and concurrent execution [25].

A specific extension of Petri nets has been created for music applications. In Music Petri Nets (MPNs), music objects are hosted by places, and are played when a token is received; transitions can contain musical operators that alter the music objects in input places and put these modified objects into output places. A music object can be seen as anything with musical meaning, e.g. a single note, a fragment of music, a control signal, etc. Music operators apply transformational algorithms such as transpositions, inversions, and time stretching. MPNs have been applied both to the analysis of existing music [26] and to composition and musical expression [27].

In order to master Petri nets and other formal models aiming to both represent and manipulate music structures, skills in cluster C13 are not sufficient; rather, a mathematical (C15) and computer-science (C14) background are required to understand how these formalisms work and to fully exploit their potential.

1.5.2 Education in Computational Auditory Scene Analysis

Submit your manuscript electronically for review.

One more trending direction in SMC research is connected to the area of music perception and cognition (C7), although the scope has been more and more broadened to include non-musical sound. Specifically, an expected outcome of SMC education will be a thorough understanding of the scientific domain often called Computational Auditory Scene Analysis (CASA). It aims at a complete description of the space of interest based solely on the acoustic modality. To this end, CASA entails the following four components: a) localization, b) enumeration, c) separation, and d) recognition of the encountered acoustic sources. These are interconnected since the efficacy of one component might be directly or indirectly related (either boosted or degraded) to that of anothers.

CASA has a wide range of applications, which may include the following:

- Voice Activity Detection, where the main goal is to segment the audio flow into speech and non-speech chunks for boosting a speech/speaker recognition system;
- Processing of musical signals, such as music transcription, identification of music genre, recognition of performer, indexing and retrieval of musical data, etc. [28, 29, 30];
- Processing of bioacoustic signals, where animal vocalizations are used towards tasks such as tracking of animals, monitoring of endangered species, biodiversity indexing [31] etc.;
- Machine acoustics, which elaborates on acoustic signals emitted by solids (e.g. metal, rock, ceramic etc.) when they are subjected to stress. Potential applications include non-destructive testing, fault detection and function control, maintenance services [32] etc.;
- Context recognition, which encompasses the recognition of the physical environ-

ment around a device including detection and identification of relevant sound events as well as recognition of the activity of the user [33, 34].

The underlying assumption is that every sound source exhibits a consistent acoustic pattern, thus a characteristic energy distribution of its frequency content. Consequently, the aim is to represent and model such distributions as accurately as possible. To this end, one needs to employ feature extraction and pattern recognition algorithms. Both processing parts are significant and have an immediate effect on the final recognition framework. On one hand, features able to capture the structure of the signals at hand are essential towards pattern discovery. On the other hand, algorithms able to model the extracted features and subsequently identify them on novel data are of paramount relevance for reliable sound recognition. Naturally, such systems need operating under real-world conditions; nonetheless there are several difficulties along this path. The system needs to deal with a large number of different sound sources, where usually the overall performance degrades. Some sound sources might not be known a-priori and as such the respective model is inexistent. Such situations need to be dealt in an online manner, thus increasing the level of difficulty. Another problem concerns the categorization of sounds into distinct classes, since sometimes it is not uniquely leading to one class overlapping with one or more. Finally, complex sound scenes where many sound sources are active simultaneously could be extremely hard to analyze. Due to these reasons, the approaches existing in the literature have targeted constrained problems and a system with generic applicability remains an open research subject.
Figure 1.5 depicts the standard structure of a sound recognition system able to identify N sound classes. Typically, the audio signal passes through a pre-processing step (mean removal and gain normalization) before it is parameterized. There, the signal is framed into small parts where the feature extraction methodology operates. The classifier concludes the sound recognition systems. Currently employed classifiers can be divided to two categories: discriminative and non-discriminative [35].

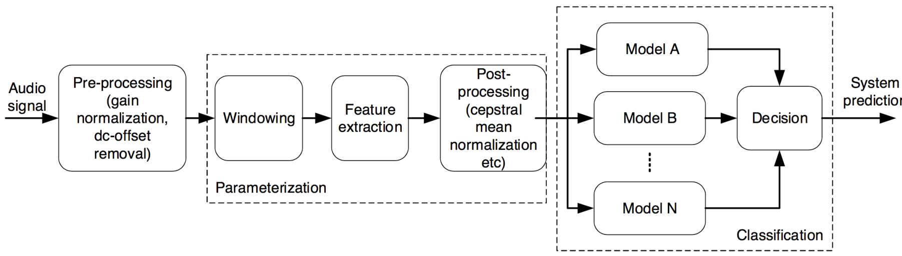

Fig. 1.5. A sound recognition system as regards to classification of N sound categories.

The first class aims at estimating the boundaries between the categories in the high dimensional space of the features. Some examples are the Polynomial Classifier [36], Multi-Layer Perceptron [37], Support Vector Machines [38], and more recently, Deep Learning [39]. On the contrary, generative approaches try to approximate the underlying distribution of the training data. These include Gaussian mixture models (GMM) [40], hidden Markov models (HMM) [41] and probabilistic neural networks (PNN)

[42]. Other non-discriminative approaches are the k-nearest neighbors (KNN) [43] and the learning vector quantization [44]. Additionally, several hybrid classification schemes have been reported in the literature [45, 46].
A promising future research directions is transfer learning for audio analysis: a typical assumption is that the training and future data must lie within the same feature space and have the same distribution. However, in many realworld applications, this assumption may not hold. For example, we sometimes have a classification task in one domain of interest, but we only have sufficient training data in another domain of interest. In such cases, knowledge transfer, if done successfully, would greatly improve the performance of learning by avoiding much expensive data labeling efforts. In recent years, transfer learning has emerged as a new learning framework to address this problem, however is not yet exploited in the field of computational audio analysis.
One further research direction is concerned with learning in non-stationary environments, where the underlying phenomena change over time [47]. Examples of these applications include making inferences or predictions based on acoustic sensor networks, monitoring of biodiversity, acoustic surveillance, etc. In non-stationary environments the probability density function of the datagenerating process may change (drift) over time. Therefore, the fundamental and rather naive assumption made by most computational intelligence approaches – that the training and testing data are sampled from the same fixed, albeit unknown, probability distribution – is simply not true. Learning in non-stationary environments requires adaptive or evolving approaches that can monitor and track the underlying changes, and adapt a model to accommodate those changes accordingly.
The above discussion shows that an expert of computational auditory scene analysis has to master multiple competences and skills, including substantial fluency in audio signal processing and modeling (cluster C2) and in related problems of acoustics (e.g., room acoustics, cluster C1), as well as audo processing techniques specifically focused on feature extraction (cluster C5). General issues related to machine learning call for a solid background in core computer science subjects (cluster C14), as well as in maths and statistics (cluster C15).

1.5.3 Education on Interactive Sound and Auditory VR/AR

Here we address the broad content area of interaction and design of multimodal interfaces, which intersects with clusters C4 and – to a minor extent – C8. Specifically we show how current developments in Virtual Reality (VR) and Augmented Reality (AR) are providing a substantial boost to SMC research related to interactive and immersive 3D sound.
Research on VR [48] and AR [49] have been going on for decades. Ten years after the Roadmap, we now are at a turning point where this research is finally reaching out to real-world applications. This is mainly due to the advent of low-cost, helmet-style VR systems that are capable of rendering complex stereoscopic 3D visual scenes (both dedicated helmets like the Oculus Rift and, more recently, smartphone-based VR systems). Big players (Google, Samsung, etc.) are developing their own VR ecosystems, and it is easy to predict that AR will be next (the Microsoft HoloLens techno-

logy is a good example of this trend). Application scenarios span a wide range, including gaming and entertainment (personal cinema, multi-channel downmix over headphones), real world human interaction, sensory substitution devices for visually-impaired users, human-robot interaction.

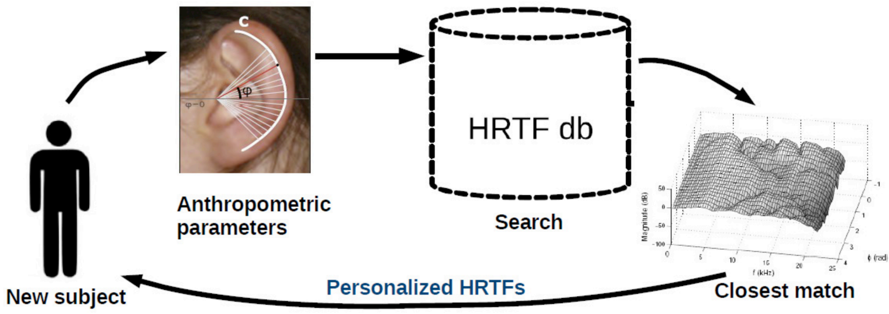

Fig. 1.6. A general scheme of a "HRTF selection" approach.

Immersive VR/AR call for multimodality, since properly designed and synchronized auditory and haptic displays are likely to provide a much greater sense of immersion in a virtual environment than a high-fidelity visual display alone. Immersive 3D audio rendering has huge potential both for VR [50] and AR [51], and can exploit off-the-shelf hardware such as acoustically transparent (hear-through) headphones [52], head-tracking sensors, 3D cameras that simultaneously capture spherical video and surround sound.

Most of current 3D sound rendering techniques over headphones rely on the use of so-called Head-Related Transfer Functions (HRTFs, or their timedomain counterparts Head-Related Impulse Responses, HRIRs) [53], i.e. filters that capture the acoustic effects of the human head and ears and allow simulation of the audio signal at the entrance of the ear canal as a function of the sound source's spatial position. These filters can be combined with other environmental acoustic effects into the so-called Binaural Room Impulse Responses (BRIRs), for given source and listener positions. Convolving an appropriate BRIR (for left and right ear) with a monaural virtual sound recreates the listening experience of the same sound as emitted from a real source at the position defined by the BRIR.

One of the main current limitations of binaural audio lies in the lack of individualization of the rendering process. Since the recording of individual HRTFs is both timeand resource-expensive, different and more convenient ways to obtain or simulate them are desirable. A common practice amounts to employing a single predefined HRTF set for any possible listener (i.e. recorded on dummy heads shaped with average anthropometric data, such as the KEMAR mannequin [54]). However, individual anthropometric features of the human body affect substantially the acoustic features of the HRTFs, therefore personal functions should be used in order to achieve high-quality rendering.

One recent trend concerns the investigation of innovative approaches for personalization of HRTFs, i.e. approaches that allow to provide a listener with a HRTF set that matches as closely as possible the perceptual characteristics of his/her own individual

HRTFs. Personalized HRTFs can be derived from computational models, which generate synthetic responses from a physical or structural interpretation of the acoustic contribution of head, pinna, shoulders and torso [55]. In alternative, personalization can be also achieved through "HRTF selection" (see Fig. 1.6). In this case, personalized HRTFs are chosen among the HRTF sets available in a database, by finding the "best" match between the listener and one of the subjects in the database based on individual anthropometric features. One the main problems to be tackled is the lack of large amount of data to train statistical models: available HRTF datasets are characterized by sparsity and heterogeneity [56]. However the recent advent of new common formats for data representation mitigates this problem and facilitates access to heterogeneous HRTF data.

A related applied research direction is concerned with the design of novel musical interfaces. In this context, one of the key ingredients to achieve a truly immersive and multimodal interaction with the virtual instrument is seamless 3D audio, that simulates the spatial characteristics of the instrumental sound as if it were immersed in the real environment. In particular, an emerging buzzword is that of Virtual Reality Musical Instruments – VRMIs [57]. This scenario calls for a set of enabling technologies (personalized binaural rendering, transparent AR headsets), and for additional research in terms of implementation on mobile architectures, measurements and collection of acoustic data from real instruments, evaluation of user experience.

An expert in interactive and immersive sound has to master core subjects in clusters C1 and C2. Perceptual and cognitive aspects (area C7) are paramount too, and notably include the development and use of computational models of the human auditory system, which can be used to develop perceptually motivated metrics of validation [58, 59]. Finally, the tools and the methodologies typical of this research have a large intersection with other fields of computer science (cluster C14), namely computer graphics and computer vision, as well as HCI.

References

[1] Bernardini, N., Serra, X., Leman, M., Widmer, G.: A roadmap for sound and music computing. The S2S2 Consortium (2007)

[2] Dannenberg, R.B.: Impressions from the SMC roadmap. Journal of New Music Research 36(3) (2007) 191–196

[3] Vinet, H.: Science and technology of music and sound: The ircam roadmap. Journal of New Music Research 36(3) (2007) 207–226

[4] Rocchesso, D.: Sounding objects in europe. The New Soundtrack 4(2) (2014) 157–164

[5] Miranda, E.R., Kirke, A., Braund, E., Antoine, A.: On unconventional computing for sound and music. In: Guide to Unconventional Computing for Music. Springer (2017) 23–61

[6] European Commission/EACEA/Eurydice: The European Higher Education Area in 2015: Bologna Process Implementation Report. Technical report, Luxembourg: Publications Office of the European Union (2015)

[7] Atchison, W.F., Conte, S.D., Hamblen, J.W., Hull, T.E., Keenan, T.A., Kehl, W.B., McCluskey, E.J., Navarro, S.O., Rheinboldt, W.C., Schweppe, E.J., et

al.: Curriculum 68: Recommendations for academic programs in computer science: a report of the ACM curriculum committee on computer science. Communications of the ACM 11(3) (1968) 151–197

[8] ACM/IEEE-CS Joint Task Force on Computing Curricula: Computer science curricula 2013. Technical report, ACM Press and IEEE Computer Society Press (December 2013)

[9] Roads, C.: The computer music tutorial. MIT press (1996)

[10] Zölzer, U.: DAFX: digital audio effects. Second edn. John Wiley & Sons (2011)

[11] Bilbao, S.: Numerical sound synthesis: finite difference schemes and simulation in musical acoustics. John Wiley & Sons (2009)

[12] Farnell, A.: Designing sound. Mit Press (2010)

[13] Hermann, T., Hunt, A., Neuhoff, J.G.: The sonification handbook. Logos Verlag Berlin (2011)

[14] Franinović, K., Serafin, S.: Sonic interaction design. Mit Press (2013)

[15] Lyon, R.F.: Human and machine hearing. Cambridge University Press (2017)

[16] Lerch, A.: An introduction to audio content analysis: Applications in signal processing and music informatics. John Wiley & Sons (2012)

[17] Müller, M.: Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. Springer (2015)

[18] Steyn, J.: Framework for a music markup language. In: Proceeding of the First International IEEE Conference on Musical Application using XML. (2002) 22– 29

[19] Baratè, A., Ludovico, L.A.: Local and global semantic networks for the representation of music information. Journal of e-Learning and Knowledge Society 12(4) (2016) 109–123

[20] Baratè, A., Haus, G., Ludovico, L.A., Perlasca, P.: Managing intellectual property in a music fruition environment the ieee 1599 case study. IEEE MultiMedia 23(2) (2016) 84–94

[21] Cohn, R.: Introduction to neo-riemannian theory: a survey and a historical perspective. Journal of Music Theory (1998) 167–180

[22] Edelsbrunner, H., Harer, J.: Persistent homology-a survey. Contemporary mathematics 453 (2008) 257–282

[23] Bergomi, M.G., Baratè, A., Di Fabio, B.: Towards a topological fingerprint of music. In: International Workshop on Computational Topology in Image Context, Springer (2016) 88–100

[24] Petri, C.A.: Introduction to general net theory. In Brauer, W., ed.: Net Theory and Applications, Berlin, Heidelberg, Springer Berlin Heidelberg (1980) 1–19

[25] Best, E., Devillers, R.: Sequential and concurrent behaviour in Petri net theory. Theoretical Computer Science 55(1) (1987) 87–136

[26] Baratè, A., Haus, G., Ludovico, L.A.: Music analysis and modeling through Petri nets. In: International Symposium on Computer Music Modeling and Retrieval, Springer (2005) 201–218

[27] Barate, A., Haus, G., Ludovico, L.A., et al.: Real-time music composition through P-timed Petri nets. In: Proceedings of ICMC—SMC—2014. (2014)

408–415

[28] Eronen, A., Klapuri, A.: Musical instrument recognition using cepstral coefficients and temporal features. In: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100). Volume 2. (2000) II753–II756 vol.2

[29] Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE 96(4) (April 2008) 668–696

[30] Ntalampiras, S.: Directed acyclic graphs for content based sound, musical genre, and speech emotion classification. Journal of New Music Research 43(2) (2014) 173–182

[31] Potamitis, I., Ntalampiras, S., Jahn, O., Riede, K.: Automatic bird sound detection in long real-field recordings: Applications and tools. Applied Acoustics 80 (2014) 1 – 9

[32] Gu, D., kim, J., An, Y., Choi, B.: Detection of faults in gearboxes using acoustic emission signal. Journal of Mechanical Science and Technology 25(5) (Aug 2011) 1279

[33] Chu, S., Narayanan, S., c. J. Kuo, C., Mataric, M.J.: Where am i? scene recognition for mobile robots using audio features. In: 2006 IEEE International Conference on Multimedia and Expo. (July 2006) 885–888

[34] Ntalampiras, S., Potamitis, I., Fakotakis, N.: Acoustic detection of human activities in natural environments. J. Audio Eng. Soc 60(9) (2012) 686–695

[35] Mesaros, A., Heittola, T., Benetos, E., Foster, P., Lagrange, M., Virtanen, T., Plumbley, M.D.: Detection and classification of acoustic scenes and events: Outcome of the dcase 2016 challenge. IEEE/ACM Transactions on Audio, Speech, and Language Processing 26(2) (Feb 2018) 379–393

[36] Specht,D.F.:Generation of polynomial discriminant functions for pattern recog nition. IEEE Transactions on Electronic Computers EC-16(3) (June 1967) 308–319

[37] Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review (1958) 65–386

[38] Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer-Verlag New York, Inc., New York, NY, USA (1995)

[39] Hertel, L., Phan, H., Mertins, A.: Comparing time and frequency domain for audio event recognition using deep learning. CoRR abs/1603.05824 (2016)

[40] Peltonen, V., Tuomi, J., Klapuri, A., Huopaniemi, J., Sorsa, T.: Computational auditory scene recognition. In: 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 2. (May 2002) II–1941–II–1944

[41] Ntalampiras, S.: A novel holistic modeling approach for generalized sound recognition. IEEE Signal Processing Letters 20(2) (Feb 2013) 185–188

[42] Bolat, B., Ku¨c¸u¨k, U¨.: Musical sound recognition by active learning pnn. In Gunsel, B., Jain, A.K., Tekalp, A.M., Sankur, B., eds.: Multimedia Content Representation, Classification and Security, Berlin, Heidelberg, Springer Berlin Heidelberg (2006) 474–481

[43] Priya, T.L., Raajan, N., Raju, N., Preethi, P., Mathini, S.: Speech and non-speech identification and classification using knn algorithm. Procedia Engineer ing 38 (2012) 952 – 958
[44] S. Yella, N.G., Dougherty, M.: Pattern recognition approach for the automatic classification of data from impact acoustics. In: Proc. of the AISC2006. (2006) 144–149
[45] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3) (Mar 1998) 226–239
[46] Ntalampiras, S.: Hybrid framework for categorising sounds of mysticete whales. IET Signal Processing 11(4) (2017) 349–355
[47] Ntalampiras, S.: Automatic analysis of audiostreams in the concept drift environment. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). (Sept 2016) 1–6
[48] Lanier, J., Heilbrun, A.: A portrait of the young visionary. Whole Earth Review (1989) 108–19
[49] Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators & Virtual Environments 6(4) (1997) 355–385
[50] Serafin, S., Nordahl, R., De Go ̈tzen, A., Erkut, C., Geronazzo, M., Avanzini, F.: Sonic interaction in virtual environments. In: Proc. IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), IEEE (2015) 1–2
[51] Albrecht, R.: Methods and applications of mobile audio augmented reality. PhD thesis, Dept. of Computer Science, Aalto University, Italy (2016)
[52] Valimaki, V., Franck, A., Ramo, J., Gamper, H., Savioja, L.: Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments. IEEE Sig. Process. Magazine 32(2) (2015) 92–99
[53] Xie, B.: Head-related transfer function and virtual auditory display. J. Ross Publishing (2013)
[54] Gardner, W.G., Martin, K.D.: HRTF measurements of a KEMAR. J. Acoust. Soc. Am. 97(6) (1995) 3907–3908
[55] Algazi, V.R., Duda, R.O., Morrison, R.P., Thompson, D.M.: Structural composition and decomposition of HRTFs. In: Proc. IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE (2001) 103–106
[56] Geronazzo, M., Granza, F., Spagnol, S., Avanzini, F.: A standardized repository of Head-Related and Headphone Impulse Response data. In: Proc. of the Audio Engineering Society Convention 134, Rome, Audio Engineering Society (2013) Paper no. 8902.
[57] Serafin, S., Erkut, C., Kojs, J., Nilsson, N.C., Nordahl, R.: Virtual reality musical instruments: State of the art, design principles, and future directions. Computer Music J. 40(3) (2016) 22–40
[58] Baumgartner, R., Majdak, P., Laback, B.: Modeling sound-source localization in sagittal planes for human listeners. J. Acoust. Soc. Am. 136(2) (2014) 791–802
[59] Geronazzo, M., Spagnol, S., Avanzini, F.: Do we need individual Head-

Related Transfer Functions for vertical localization? the case study of a spectral notch distance metric. IEEE/ACM Trans. Audio, Speech, and Language Process. (2018) Accepted for publication.