Digital Signal Processing for Audio Applications: Then, Now and the Future

Francesco Piazza, Stefano Squartini, Stefania Cecchi, Simone Fiori, Simone Orcioni, Susanna Spinsante and Stefano Pirani

Abstract In the last fifty years, the development of new technologies has enabled machines to sustain the ever increasing computational load, thus providing the implementation capability requested by real time applications. In this context, digital signal processing played an important role especially with relation to audio systems. Several approaches have been proposed to solve the main issues of the audio field in complex scenarios, including advanced audio rendering applications and acoustic monitoring systems exploiting multirate adaptive algorithms, machine learning techniques and deep neural circuits. Following this trend and based on our experience, the future will witness the joint use of these techniques to design applications able to improve quality and comfort of people's daily life. Among them, in this contribution we want to focus on the employment of advanced audio augmented reality solutions, involving both virtual audio sensors and transducers, to design enhanced spatial hearing experiences in diverse application contexts, spanning from entertainment to safety.

F. Piazza · S. Squartini · S. Cecchi (✉) · S. Fiori · S. Orcioni · S. Spinsante · S. Pirani
Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, Italy
e-mail: s.cecchi@univpm.it

F. Piazza
e-mail: f.piazza@univpm.it

S. Squartini
e-mail: s.squartini@univpm.it

S. Fiori
e-mail: s.fiori@univpm.it

S. Orcioni
e-mail: s.orcioni@univpm.it

S. Spinsante
e-mail: s.spinsante@univpm.it

S. Pirani
e-mail: stefano.pirani@univpm.it

S. Longhi et al. (eds.), *The First Outstanding 50 Years of "Università Politecnica delle Marche"*, https://doi.org/10.1007/978-3-030-32762-0_3

1 Introduction

In the last fifty years, the development of new technologies has enabled machines to sustain the ever increasing computational load, thus providing the implementation capability requested by real time applications. In this context, digital signal processing played an important role especially with relation to audio systems that are present in different applications and scenarios of the daily life. In particular, the growing presence and adoption of electronic equipment in the everyday life, has led to the need of audio processing procedures to make the ambient intelligent, such as the automatic recognition of commands and activities performed by subjects in their living environment, to help and support them. Furthermore, the pervasiveness of technology has raised a great attention on the aspects related to a comfortable living, including also the acoustic comfort. In this context, several audio processing algorithms can be used to enhance the audio reproduction systems, exploiting audio equalization and analyzing the non-linear behaviour of audio devices. All these applications can be implemented introducing multirate adaptive algorithms and machine learning techniques. Multirate adaptive algorithms are based on the use of adaptive filtering with subband structure allowing a real time identification of the analyzed system with fast convergence and low computational complexity. Machine learning techniques, including deep neural circuits which have recently encountered a remarkable success, are aimed at extracting useful representation knowledge from acquired audio signals, to pilot the execution of automatized services. Taking a look at the future, the use of these techniques will increase since they guarantee good results in terms of computational complexity, achieved quality and easy adoption in those applications that address people's comfort in their daily life.

In Sect. 2, a brief overview of our contributions is reported. In particular, Sect. 2.1 describes our contribution on the use of audio signals for ambient intelligence while Sect. 2.2 reports our contribution on the use of audio algorithms for the improvement of the audio reproduction. New trends and future research directions are presented in Sect. 3, where the employment of advanced audio augmented reality solutions, involving both virtual audio sensors and transducers, is introduced to design enhanced spatial hearing experiences in diverse application contexts.

2 Our Contribution in the Audio Field

2.1 Audio-Based System for Ambient Intelligence

Systems and solutions based on audio processing are of great interest for Ambient Intelligence, to allow the automatic recognition of commands, the identification of either so-called Activities of Daily Living (ADLs) as well as anomalous events or potentially dangerous situations, like human falls. Several research studies addressed

the possibility to exploit the audio signals acquired by means of suitable microphones deployed in the monitored living environment [2, 17].

In this context, the Audio-based System for Ambient Intelligence [5, 51, 52] has been recently developed by some of the authors at UnivPM. It addresses the automatic detection of emergencies and the recognition of commands in home automation contexts. An emergency here is represented by a situation of distress for the user, where he/she intentionally asks for help, or by an abnormal acoustic event. Recalling the classification proposed in [1] the application scenario is thus monitoring for emergency detection and ambient assisted living. The system operates in two modalities that are chosen by the user to monitor different situations.

The first modality, speech monitoring, is enabled when the user is inside the home and consists in recognizing home automation commands and distress calls. Commands are automatically interpreted to control the appliances and the devices connected to the home automation system. Distress calls are employed to provide tele-assistance to the users. In particular, a distress call triggers an automatic phone call to a relative or a care center that then can provide assistance to the user. The acoustic environment is constantly monitored to detect speech signals by means of a Voice Activity Detector (VAD), and a speech recognizer based on PocketSphinx [23] captures distress calls and voice commands. Robustness against noise and reverberation is increased by integrating Power Normalized Cepstral Coefficients (PNCC) [26] and Multichannel Histogram Equalization (MHEQ) [57]. In addition, the sounds from a television or a radio are reduced by means of an interference cancellation module. The recognition performance has been assessed on ITAAL [52] a corpus of home automation commands and distress calls in Italian. The experiments have been conducted both with and without the presence of a radio show that represents an interference audio signal that can be present in the everyday use of the system.

The second modality is activated for surveillance purposes, e.g., when the user is outside the house. The system now monitors the acoustic environment to detect events that deviate from normality. As in the case of distress calls, their detection triggers an automatic phone call towards a user-defined phone number. The novelty detector [33, 34, 47] is based on the approach proposed in [39] which consists in extracting a set of features from the audio signal, and in modeling normal sounds by means of a statistical generative model. PNCCs, MFCCs, critical band-based TEO autocorrelation envelope [64], MPEG-7 features [25] and their combinations have been evaluated in order to determine the feature set with the overall best performance/computational cost ratio. In regard to the normality model, Gaussian Mixture Models and Hidden Markov Models have been both considered in order to find the best performing technique for the application scenario. Differently from [39] in the recognition phase, the decision is performed on a chunk-based analysis. The effectiveness of the approach has been assessed on a newly developed corpus for novelty detection, named A3Novelty, which contains more than 56 hours of recordings comprising both normal abnormal sounds.

Summarizing, it can be said that the system combines active and pro-active operation modes for emergency detection, since the user can explicitly ask for assistance by uttering a distress call, or have the system detect an emergency by analyzing

abnormal sounds. Moreover, the latter mode is performed by means of a novelty detector algorithm that does not require the explicit modeling of abnormal sounds representing an emergency. In regard to the sensors employed, the use of microphone does not require the user to wear a specific device for emergency detection and allows a simple integration of a voice user interface. Finally, the proposed system comprises both the algorithms for emergency detection, and the ones for its management, i.e., for enabling the communication between the person asking for assistance and the relative or care center. The algorithms have been implemented on a low-consuming embedded platform, i.e., the BeagleBoard-xM, while state-of-the art alternatives are based on more costly and energy consuming PC hardware [46].

A specific subsystem for fall detection based on an innovative floor acoustic sensor, as described in [50], has been also developed and included in the overall audio-based system for Ambient Intelligence. The sensor is composed of a microphone embedded in a resonant enclosure whose bottom surface is in direct contact with the floor. In this way, the microphone captures the acoustic waves transmitted through the floor and it mainly captures the sound of falling objects, resulting in a minor sensitivity to the environmental noise. In addition, it is able to capture the subtle signal components transmitted through the floor, which are absent in the signal transmitted through the air. The fall signals are then processed to recognize from which kind of fall they are produced: for this purpose, a multiclass classifier is implemented. The algorithm is based on Mel-Frequency Cepstral Coefficients as low-level acoustic features and Gaussian means supervectors as features for a Support Vector Machine classifier. More in details, a background model is created from a large set of audio events signals, and for each audio event class taken into consideration, a set of supervectors is calculated by adapting the background model with the Maximum a Posteriori algorithm and extracting the means of the Gaussians. The supervectors are then employed for training the Support Vector Machine classifier. The performance of the system has been assessed by creating a corpus of fall events acquired using the audio sensor in a realistic scenario. The obtained results showed that the proposed approach is able to discriminate persons' falls with values of recall and precision higher than 98%. A recent update of the subsystem includes a semi-supervised approach, relying on advanced template-matching solutions, for automatic human fall detection [15].

It must be observed that falls of persons have been widely addressed by the scientific community, since they represent the primary cause of injury-related death for the elders [37]. Approaches to the problem are based either on wearable sensors (e.g., accelerometers) or on ambient sensors (e.g., microphones, cameras, floor vibration sensors). The first ones exploit the information of the falling body acceleration [14], from which a fall event is detected when the value exceeded the typical normal level. Instead, the ambient sensors reveal the falling activity from the observation of the environment in which they are positioned: those could be used individually [55, 65] or combining the information coming from heterogeneous sensors [63, 66] for improving the reliability. In the case of floor vibration sensors, the detection is performed by analysing the signal resulting from the fall event, which produces a characteristic vibration, while in systems based on videocameras, the detection is based on the deformation of the human shape. Recently, several approaches appeared

in the literature that are based exclusively on audio signals [31, 32]. The motivation is that microphones are perceived as less invasive compared to wearable sensors and cameras and they do not suffer from occlusions. A common approach is to install several microphones in the house, usually on the ceiling or near the walls. The problem with these approaches is their sensitivity to environmental noise, that usually requires the adoption of beamforming techniques to enhance the signal quality and thus achieve a sufficient detection accuracy [31]. The floor acoustic sensor is immune to this kind of limitations.

Concluding, it is interesting to underline that, besides a specific deployment of acoustic sensors in the environment to be monitored, also microphones mounted on mobile devices, like smartphones, can be effectively used to collect audio data, and integrated into ADL identification modules, in order to facilitate the correct classification of the ADL by sensor-fusion approaches. As shown in the recent review by Pires et al. [49], most of the studies exploiting microphones on-board smartphones apply audio fingerprinting techniques, aiming to find a match between the signal collected by the microphone during ADL execution, and a database of well-known audio fingerprints. Being the published methods very diverse, and having been tested over different data sets and different feature extraction techniques, it is quite difficult to provide a final evaluation about the best audio fingerprinting technique to be used for the aim of ADL identification. Different approaches exist, that avoid the need to collect big amounts of fingerprints, despite being anyway able to identify different acoustic events. This aspect will be faced in the next future and suitable algorithms need to be developed for integration within the current Audio-based Systems of Ambient Intelligence.

2.2 Advanced Systems for Audio Reproduction Enhancement

When sound is reproduced by one or more loudspeakers in a real scenario, the acoustic perception is modified by the characteristics of the listening environment such as a room or a car cockpit. A small quantity of reverberation is required since it adds spaciousness and depth to the sound, however excessive reflections or resonances may result in an undesired alteration of the auditory illusion, adding some artifacts (e.g., frequency band extension, nonlinearities) to the original sound. In this context, an audio equalization algorithm is required to contrast the detrimental effects of the room environment and of the reproduction system [13]. Equalization is realized taking into consideration the transfer function that represents the path from the sound reproduction system to the listener and then this function is modified with a suitably designed equalizer that can be realized in several manners. The basic idea is to measure the impulse response of the environment using a microphone, and then obtain the equalizer through its inversion. However, several issues influence this method, and thus a wide variety of techniques have been developed over the last 40 years to counteract them [13]. Approaches to the design of the equalizer can be divided in single-point and multi-point ones. A single position equalizer estimates

the equalization filter on the basis of the measurement of the impulse response in a single location [38]. This way, the filter is effective only on a reduced zone around the measurement point, the extension of which is proportional to a fraction of the acoustic wavelength. However, the impulse response varies significantly with the position of the microphones in the room or car environment [29, 35] and with time [22] as these environments can be considered as "weakly non-stationary" systems [36]. To enlarge the equalized zone and to contrast the room and car response variations, multi-point equalizers have been proposed [4, 7, 11]. A multi-point equalizer uses multiple measurements of the impulse responses at different locations in order to design the equalizer. These approaches can be used for fixed and adaptive equalization [12]. The former is based on measurements obtained with a microphone positioned in a fixed place, the latter is capable of tracking and adapting to environment variations that can occur due to the modifications of temperature, pressure, and movement of people or other obstacles within the enclosure. Different pre-processing techniques can be applied to contrast the audible distortions caused by equalization errors due to these environment variations [11], and different equalizer design techniques can also be adopted, taking into consideration minimum-phase or mixed-phase approaches.

In the context of audio reproduction enhancement, an important role is relative to the audio devices identification. The non-linear behaviour of some devices could be considered beneficial in some cases, such as guitar amplifier reproduction, or not beneficial in the case of impulse response measurement, where the amplifier can introduce its own non-linear behaviour. Several methods can be found in the literature about non-linear system identification.

Volterra series is a linear-in-the-parameters (LIP) [8] nonlinear filter used for non-linear signal processing and non-linear system identification. It was actively used in the audio field from audio effect emulation [59, 60], to nonlinear acoustic echo cancellation [3, 6] or nonlinear active noise control [18, 58]. The identification of Volterra series can be carried out by searching the minimum of the mean square error (MMSE) between the outputs of the series and the target system. If the input is taken from an independent identically distributed (i.i.d.) sequence, it is well known that the cross-correlation method due to Lee–Schetzen [56] gives the optimal solution in the MMSE sense. This method needs the output to be expressed as a sum of orthogonal functional as those proposed by Wiener [62], since Volterra functional are not orthogonal to each other.

The Lee–Schetzen method undergoes many drawbacks: the central moments of a Gaussian input deviate from ideal values as the moment order increases [43]; the input non-idealities affect particularly the estimation of the kernels diagonal points [43, 44]; and the problem is worsened by the errors caused by a model order under-determination [40].

Effective solutions that overcome the problem of diagonal points identification have been proposed in the literature, for series up to the third order in [19], and for a generic order in [48], where a comparison between the two methods is also provided. While in analytical power series (infinite sum of elements), the identification with cross-correlation is independent of the input variance, this is no more true with truncated power series, where the approximation error depends on the variance used

in the identification: a Volterra series is optimal only for inputs with variances in a neighborhood of that used for identification, also called the problem of "locality" of solution [40]. An improved cross-correlation method to overcome the problem of the "locality" of solution, based on multiple-variances has been proposed for Wiener-Volterra series and Gaussian noise in [40]: low input variances are used to identify lower-order Wiener kernels, while the input variance is gradually increased for higher-order kernels. This allows a better identification of systems that have high dynamic inputs, like audio systems, and can be applied to amplifiers [41, 45] or loudspeaker systems. In [41] the multiple-variance approach was used for the identification of audio devices with deterministic periodic signals, called perfect periodic sequence (PPS), that guarantee the orthogonality of the basis functions on a finite period. In [45] the multiple-variance approach for the identification of tube audio devices was completed with a method that drastically reduces the curse of Volterra series dimensionality, i.e. the exponential relashionship between the coefficient of the series and the order and memory of the system to be identified.

It is possible to make easier the nonlinear system identification with the use of Wiener nonlinear (WN) filters, which derive directly from the double truncation of the Wiener series. The WN filter is a first example of nonlinear filters with orthogonal basis functions, in particular, orthogonal for a white Gaussian input signal. The orthogonality of the basis functions allows the efficient identification of the filter coefficients with the cross-correlation method, as in [30]. PPS can also be developed for WN filters. Expressing the WN filter as a linear combination of basis functions and using a PPS input signal, problems in the estimation of the kernel diagonal points can be avoided [9, 10]. Also the multiple-variances method, that avoid the locality of the solution, can be applied to WN filters [42] with some advantages with respect to the use of a white Gaussian input, as originally proposed in [40].

All experiments have been realized taking advantage of the semianechoic chamber realized at the Department of Information Engineering, and shown in Fig. 1, that allows to perform several tests in a controlled environment.

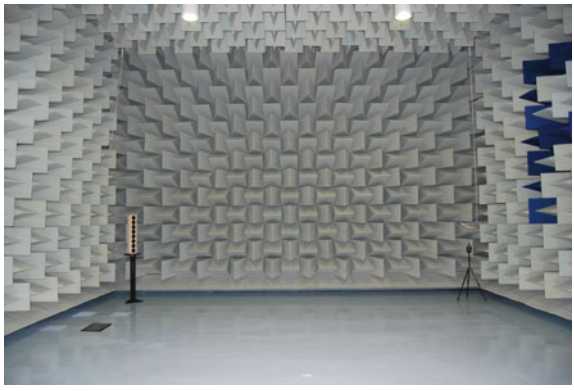

Fig. 1 Semianechoic chamber at the department of information engineering used for audio experiments

3 A Knowledgeable Vision on Digital Audio Applications

3.1 Vision

The overall future vision on Digital Audio applications sees the synergistic combination of methodologies for spatial audio processing aimed at augmented and virtual reality on headsets, and techniques of machine audition in the context of safety monitoring (Fig. 2). By analyzing and defining the acoustic scene through the use of microphones, the correct spatial information can be reproduced by means of headphones.

Audio augmented reality (AAR) combines virtual sound sources with the real sonic environment of the user [21]. It can be realized by means of a device that a user could be wearing at all times, such as a headset [53]. This way, the user can at the same time, hear and interact with the real acoustic environment in a natural way, allowing ordinary speech communication with other people and permitting all those operations for which acoustic feedback is important [16, 61]. To generate and render

Fig. 2 Future vision on digital audio applications where augmented and virtual reality will be applied on headsets for spatial audio processing

virtual auditory events, it is necessary to consider the properties of a real auditory event and its perception by the human ears. The human perception is spatial and it is bounded to localization cues and to the brain capability of interpreting them. Therefore, it is possible to develop a spatial rendering system through the use of auralisation, i.e., the reproduction of virtual sound considering the human perception capability in such a way that it evokes the same listening experience of a real sound source at a specific point in the space [27]. This aspect is realized considering binaural audio approaches that recreate this sensation exploiting the Head Related Transfer Functions (HRTFs) [20]. HRTFs are unique to each person since they are related to the human head physiology, however several approaches capable of overcoming HRTF individualization can be found in the literature [20]. In this context, several approaches for the enhancement of spatialized sound reproduction can be developed. Adaptive equalization is another key point to be implemented since it is necessary to make headphones acoustically transparent allowing a real perception of the ambient sounds [53]. In [54], a natural augmented reality headset with two pairs of binaural microphones to achieve natural listening experience using online adaptive filtering was presented. If the ambient is extremely noisy, an Active Noise Cancellation (ANC) system can be developed for the headset. Active noise controllers (or cancellers) cancel the noise in a certain location by destructive interference with an anti-noise signal [28]. In this particular case of AAR, selective ANC is considered since just specific disturbances have to be cancelled, preserving other sound sources that are important for a pleasant fruition of the audio scene [24], or for the audio security of the user.

In the perspective of the future research activities, two relevant application scenarios are of interest: indoor (applicable to workers in a noisy factory) and outdoor (targeting pedestrians listening to music). In both scenarios, the computational time required by the overall processing chain (source localization and enhancement, acoustic scene understanding, and synthesis of the audio stream) must be kept at bay in order to give the user the proper time to react. At the authors' best knowledge, the idea of synergistic cooperating tools coming from machine audition and augmented reality on headset, featuring a low latency in order to increase the sense of presence in real environments, is innovative and not addressed before. This goal poses several challenges, which compel researchers to develop solutions that go beyond the state of the art. In order to keep under control the reaction of the system, it is important to consider the different nature of the scenarios of interest, especially for what concerns the interfering noises and the reverberation.

The overarching goal of the envisioned research is pursued into three main objectives, all of them pertaining to areas that are individually very active.

Audio Quality Enhancement. This objective aims for the enhancement of the sound field acquired by one or more microphone arrays, to obtain specific signals for the subsequent analysis of the acoustic scene. This is a very active area in the audio community, and many applications in the last years have gained commercial interest. The scenarios of interest mentioned above, however, require to develop functionalities with capabilities that go far beyond the state of the art. In particular, innovative algo-

rithms for source enhancement that can work in the presence of strong reverberations, multiple moving sources and interferers will be investigated and developed.

Acoustic scene analysis and understanding. Research on this topic has been receiving much interest within the audio research community. In this context, the problems of localizing sources and understanding their nature will be tackled by exploiting innovative techniques, such as plenacoustic methods, which are able to accommodate reverberant environments and multiple sources. In particular, advanced deep learning techniques and neural network architectures will be considered to overcome the main problems of the aforementioned scenarios.

Audio augmented and virtual reality. In this objective, an artificial sound spatialization system exploiting personal spatial audio technologies will be proposed. Starting from the study of the state of the art, innovative solutions based on HRTF will be presented, exploiting the possibility of using alternate and less expensive solutions for spatial audio rendering, based on spatial cues such as Interaural Time Difference (ITD) and Interaural Level Difference (ILD). Furthermore, some techniques capable of improving the sound perception will also be investigated, such as adaptive equalization and selective ANC.

3.2 Proposed Methodology

The intended research direction envisioned for the future is both innovative and challenging, being based on the effective combination and fusion of different powerful methodologies. In particular, it mainly draws on the digital signal processing and machine learning theories applied to audio and acoustic data. The proposed methodology aims at providing effective solutions to the tasks identified, while meeting some constraints posed by a specific application scenario and by the quality requirements desired by a user. Among such constraints, a particular attention in the algorithmic design and development will be given to the available computational resources. Indeed, the available hardware is subject to the scenario considered. For instance, in indoor scenarios distributed hardware architectures may be used, whereas more limited computational resources will be available in outdoor scenarios (e.g., portable devices). In that sense, the methodology aimed at will be compatible and adaptable to the different application contexts that will be considered and the proposed algorithms will be designed to be efficient also from a computational point of view. The proposed methodology will show a significant degree of novelty, mainly due to the novelty of the proposed application scenarios, to the design of algorithms that must be able to satisfy the imposed constraints (e.g., quality, hardware), to the development of algorithms deriving from the joint use of machine learning and signal processing methods. Taking into account all the aspects mentioned above, the envisioned methodology includes diverse sets of algorithms, as described below.

A first set of methods is devoted to the quality enhancement of the audio captured by microphones. To this end, the geometric arrangement of microphones with

respect to the sound sources requires an accurate study that depends on the specific application. In stationary and controlled environments, arrays can be easily placed on fixed and even bulky supports and their dimensions are not an issue. In contrast, the array size may be required to be spatially limited e.g., when microphones are integrated in a portable device. However, regardless of the specific application, the arrangement and correct calibration of the microphones significantly affect the qualitative improvement of the audio signal. Methodologies to accomplish this task include: spatial filtering design for defining the optimal geometries of microphone arrays in well-defined environments; optimization algorithms for the selection of a subset of microphones for a fixed geometry; deep learning techniques for relaxing the geometrical constraints and the necessity of microphones calibration, and super-resolution techniques that use a priori information to increase the signal resolution.

Different natural and artificial artifacts usually affect the quality of the captured audio, including background noise, reverberation and interfering sounds. A first significant enhancement of the audio quality is provided by beamforming techniques, which aim at reducing the noise sounds coming from outside the sound field of interest where a desired sound source is located. Advanced space-time processing methods can be developed according to some general characteristics of the microphones employed (e.g., single microphone, distributed sensors, coincident microphone arrays). In a second stage, the residual noise can be further reduced by implementing signal enhancement algorithms to improve the audio intelligibility. Enhancement methods also include the plenacoustic representation, which provides a ray space image of the directional components of the sound field by using small sub-arrays among the employed microphones, machine learning techniques for dereverberation, as well as the separation of mixtures of sound signals.

The identification of the source position is fundamental to correctly recognize the nature of a sound and provide an accurate rendering of the audio signal that can be pleasant to the human listening. To this end, several methodologies will be involved for the sound localization, including binaural techniques based on HRTF, plenacoustic framework for distributed microphones, deep learning techniques for 3D sound localization.

Once signals have been acquired, enhanced and spatially localized, it is important to perform a description of the whole scenario with a high-level detail. An acoustic scene is characterized by particular sounds that can be associated with a specific event. Advanced machine learning algorithms can be developed to detect a well-known sound or even to identify an "anomalous" event that might represent a potential risk for human safety. However, the detection of an anomalous sound event should be followed by an adequately trained classification architecture to correctly evaluate the possible risks. To this end, novel deep learning solutions involving semi-supervised learning, data augmentation, and transfer learning strategies will be taken into account. Sound events can be further analyzed to describe a particular environment by "listening" to the sound rebounding in it. Such analysis, known as "acoustic scene understanding", can be performed by involving data-driven models, providing a scene classification based on the analysis of frames with different temporal depth.

The high-quality audio signal rendering for an enhanced experience is also included within the proposed methodology. In order to provide the user with a spatial perception of the augmented audio, several methods based on the HRTF will be developed. In particular, research efforts will be focused on the spatial audio rendering engine and its personalization, as well as on methods for personalizing HRTF according to the user preferences, by using morphing of data both from publicly available datasets and from real measurements. In order to improve directional cues of warning messages, low-latency solutions will be investigated involving time difference and interaural level difference. The audio rendering for augmented reality also involves adaptive equalization for acoustic transparency in headset reproduction. Indeed, novel equalization algorithms will be developed to make headsets perceptually transparent to specific external sounds, while providing at the same time an optimal reproduction of the infotainment audio message. Moreover, some procedures to automatically select desired sound events will be developed based on spatial filtering and advanced active noise cancellation, besides integrating the information provided by the acoustic scene understanding subsystem.

4 Conclusions

Digital signal processing brought exciting achievements and innovations in the audio domain, during the last fifty years. Among them, this chapter focused on advanced audio augmented reality solutions, involving both virtual audio sensors and transducers, to design enhanced spatial hearing experiences in diverse application contexts, spanning from entertainment to safety. According to the authors' knowledge gained in the field and future perspectives, all these innovative techniques will lead to new applications able to substantially improve quality of experience and comfort in people's daily life.

References

1. Acampora G, Cook DJ, Rashidi P, Vasilakos AV (2013) A survey on ambient intelligence in healthcare. Proc IEEE 101(12):2470–2494
2. Alsina-Pagès R, Navarro J, Alías F, Hervás M (2017) Homesound: real-time audio event detection based on high performance computing for behaviour and surveillance remote monitoring. Sensors 17(4):854
3. Azpicueta-Ruiz LA, Zeller M, Figueiras-Vidal AR, Arenas-Garcia J, Kellermann W (2011) Adaptive combination of volterra kernels and its application to nonlinear acoustic echo cancellation. IEEE Trans Audio Speech Lang Process 19(11):97–110
4. Bharitkar S, Kyriakakis C (2006) Immersive audio signal processing. Springer Science & Business Media
5. Bonfigli R, Ferroni G, Principi E, Squartini S, Piazza F (2014) A real-time implementation of an acoustic novelty detector on the beagleboard-xm. In: 2014 6th European embedded design in education and research conference (EDERC). IEEE, pp 307–311

6. Burton TG, Goubran RA (2011) A generalized proportional subband adaptive second order volterra filter for acoustic echo cancellation in changing environments. IEEE Trans Audio Speech Lang Process 19(8):2364–2373
7. Carini A, Cecchi S, Piazza F, Omiciuolo I, Sicuranza GL (2012) Multiple position room response equalization in frequency domain. IEEE Trans Audio Speech Lang Process 20(1):122–135
8. Carini A, Cecchi S, Orcioni S (2018) Orthogonal lip nonlinear filters. In: Comminello D, Príncipe JC (eds) Adaptive learning methods for nonlinear system modeling, chapter 2. Elsevier
9. Carini A, Cecchi S, Terenzi A, Orcioni S (2018) On room impulse response measurement using perfect sequences for wiener nonlinear filters. In 2018 26th European signal processing conference (EUSIPCO). IEEE, pp 982–986
10. Carini A, Romoli L, Cecchi S, Orcioni S (2016) Perfect periodic sequences for nonlinear wiener filters. In 2016 24th European signal processing conference (EUSIPCO), pp 1788–1792
11. Cecchi S, Palestini L, Peretti P, Romoli L, Piazza F, Carini A (2011) Evaluation of a multi-point equalization system based on impulse response prototype extraction. J Audio Eng Soc 59(3):110–123
12. Cecchi S, Romoli L, Carini A, Piazza F (2014) A multichannel and multiple position adaptive room response equalizer in warped domain: real-time implementation and performance evaluation. Appl Acoust 82:28–37
13. Cecchi S, Carini A, Spors S (2018) Room response equalization—a review. Appl Sci 8(1):16
14. Chin-Feng L, Sung-Yen C, Han-Chieh C, Yueh-Min H (2011) Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling. IEEE Sens J 11(3):763–770
15. Droghini D, Ferretti D, Principi E, Squartini S, Francesco F, (2017) A combined one-class svm and template-matching approach for user-aided human fall detection by means of floor acoustic features. Comput Intell Neurosci
16. Gamper H et al (2014) Enabling technologies for audio augmented reality systems. PhD thesis, Aalto University
17. García-Hernández A, Galván-Tejada C, Galván-Tejada J, Celaya-Padilla J, Gamboa-Rosales H, Velasco-Elizondo P, Cárdenas-Vargas R (017) A similarity analysis of audio signal to develop a human activity recognition using similarity networks. Sensors 17(11):2688
18. George NV, Panda G (2013) Advances in active noise control: a survey, with emphasis on recent nonlinear techniques. Signal Process 93(2):363–377
19. Goussard Y, Krenz W, Stark L (1985) An improvement of the lee and schetzen cross-correlation method. IEEE Trans Autom Control AC-30(9):895–898
20. Hai ND, Chaudhary NK, Peksi S, Ranjan R, He J, Gan WS (2017) Fast HRFT measurement system with unconstrained head movements for 3d audio in virtual and augmented reality applications. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 6576–6577
21. Härmä A, Jakka J, Tikander M, Karjalainen M, Lokki T, Hiipakka J, Lorho G (2004) Augmented reality audio for mobile and wearable appliances. J Audio Eng Soc 52(6):618–639
22. Hatziantoniou PD, Mourjopoulos JN (2004) Errors in real-time room acoustics dereverberation. J Audio Eng Soc 52(9):883–899
23. Huggins-Daines D, Kumar M, Chan M, Black AW, Ravishankar M, Rudnicky AI (2006) Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices. In 2006 IEEE international conference on acoustics, speech and signal processing ICASSP 2006 proceedings, vol 1. IEEE, p I
24. Hu S, Rajamani R, Yu X (2011) Active noise control for selective cancellation of external disturbances. In American control conference (ACC). IEEE, pp 4737–4742
25. Kim H-G, Moreau N, Sikora T (2006) MPEG-7 audio and beyond: audio content indexing and retrieval. Wiley
26. Kim C, Stern RM (2012) Power-normalized cepstral coefficients (PNCC) for robust speech recognition. In 2012 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 4101–4104

27. Kleiner M, Dalenbäck BI, Svensson P (1993) Auralization-an overview. J Audio Eng Soc 41(11):861–875
28. Kuo SM, Mitra S, Gan WS (2006) Active noise control system for headphone applications. IEEE Trans Control Syst Tech 14(2):331–335
29. Kyriakakis C (1998) Fundamental and technological limitations of immersive audio systems. Proc IEEE 86(5):941–951
30. Lee YW, Schetzen M (1965) Measurement of the wiener kernels of a nonlinear system by crosscorrelation. 2(3):237–254
31. Li Y, Ho KC, Popescu M (2012) A microphone array system for automatic fall detection. IEEE Trans Biomed Eng 59(5):1291–1301
32. Li Y, Ho KC, Popescu M (2014) Efficient source separation algorithms for acoustic fall detection using a microsoft kinect. IEEE Trans Biomed Eng 61(3):745–755
33. Markos M, Sameer S (2003) Novelty detection: a review-part 1: statistical approaches. Signal process 83(12):2481–2497
34. Markos M, Sameer S (2003) Novelty detection: a review-part 2: neural network based approaches. Signal proces 83(12):2499–2521
35. Mourjopoulos J (1985) On the variation and invertibility of room impulse response functions. J Sound Vib 102(2):217–228
36. Mourjopoulos J (2003) Comments on'analysis of traditional and reverberation-reducing methods of room equalization'. J Audio Eng Soc 51(12):1186–1188
37. Muhammad M, Ling S, Luke S (2013) A survey on fall detection: principles and approaches. Neurocomputing 100:144–152
38. Neely ST, Allen JB (1979) Invertibility of a room impulse response. J Acousti Soc Am 66(1):165–169
39. Ntalampiras S, Potamitis I, Fakotakis N (2011) Probabilistic novelty detection for acoustic surveillance under real-world conditions. IEEE Trans Multimed 13(4):713–719
40. Orcioni S (2014) Improving the approximation ability of volterra series identified with a cross-correlation method. Nonlinear Dyn 78(4):2861–2869
41. Orcioni S, Carini A, Cecchi S, Terenzi A, Piazza F (2018) Identification of nonlinear audio devices exploiting multiple-variance method and perfect sequences. In Audio engineering society AES 144th convention paper
42. Orcioni S, Cecchi S, Carini A (2017) Multivariance nonlinear system identification using wiener basis functions and perfect sequences. In 2017 25th European signal processing conference (EUSIPCO), pp 2748–2752
43. Orcioni S, Pirani M, Turchetti C (2005) Advances in Lee-Schetzen method for volterra filter identification. Multidimens Sys Sig Process 16(3):265–284
44. Orcioni S, Pirani M, Turchetti C, Conti M (2002) Practical notes on two volterra filter identification direct methods. In Proceedings of IEEE international symposium on circuits and systems ISCAS'02, vol 3. Scottsdale, Arizona, pp 587–590
45. Orcioni S, Terenzi A, Cecchi S, Piazza F, Carini A (2018) Identification of Volterra models of tube audio devices using multiple-variance method. J Audio Eng Soc 66(10):823–838
46. Paoli R, Fernández-Luque FJ, Doménech G, Martínez F, Zapata J, Ruiz R (2012) A system for ubiquitous fall monitoring at home via a wireless sensor network and a wearable mote. Expert Syst Appl 39(5):5566–5575
47. Pimentel MA, Clifton DA, Clifton L, Tarassenko L (2014) A review of novelty detection. Signal Process 99:215–249
48. Pirani M, Orcioni S, Turchetti C (2004) Diagonal kernel point estimation of n-th order discrete Volterra-wiener systems. EURASIP J Appl Signal Process 12:1807–1816
49. Pires IM, Santos R, Pombo N, Garcia NM, Florez-Revuelta F, Spinsante S, Goleva R, Zdravevski E (2018) Recognition of activities of daily living based on environmental analyses using audio fingerprinting techniques: a systematic review. Sensors 18(160):23
50. Principi E, Droghini D, Squartini S, Olivetti O, Piazza F (2016) Acoustic cues from the floor: a new approach for fall classification. Expert Syst Appl 60:51–61

51. Principi E, Squartini S, Bonfigli R, Ferroni G, Piazza F (2015) An integrated system for voice command recognition and emergency detection based on audio signals. Expert Syst Appl 42(13):5668–5683
52. Principi E, Squartini S, Piazza F, Fuselli D, Bonifazi M (2013) A distributed system for recognizing home automation commands and distress calls in the italian language. In Interspeech, pp 2049–2053
53. Rämö J, Välimäki V (2012) Digital augmented reality audio headset. J Electr Comput Eng
54. Ranjan R, Gan WS (2015) Natural listening over headphones in augmented reality using adaptive filtering techniques. IEEE/ACM Trans Audio Speech Lang Process (TASLP) 23(11):1988–2002
55. Rougier C, Meunier J, St-Arnaud A, Rousseau J (2011) Robust video surveillance for fall detection based on human shape deformation. IEEE Trans circuits syst video Technol 21(5):611–622
56. Schetzen M (1974) A theory of non-linear system identification. Int J Control 20(4):577–592
57. Squartini S, Principi E, Rotili R, Piazza F (2012) Environmental robust speech and speaker recognition through multi-channel histogram equalization. Neurocomputing 78(1):111–120
58. Tan L, Jiang J (1997) Filtered-X second-order Volterra adaptive algorithms. Electron Lett 33(8):671–672
59. Tronchin L (2012) The emulation of nonlinear time-invariant audio systems with memory by means of Volterra series. J Audio Eng Soc 60(12):984–996
60. Tronchin L, Coli VL (2015) Further investigations in the emulation of nonlinear systems with Volterra series. J Audio Eng Soc 63(9):671–683
61. Valimaki V, Franck A, Ramo J, Gamper H, Savioja L (2015) Assisted listening using a headset: enhancing audio perception in real, augmented, and virtual environments. IEEE Signal Process Mag 32(2):92–99
62. Wiener N (1966) Nonlinear problems in random theory. The MIT Press, Cambridge, MA
63. Yazar A, Keskin F, Töreyin BU, Çetin AE (2013) Fall detection using single-tree complex wavelet transform. Pattern Recognit Lett 34(15):1945–1952
64. Zhou G, Hansen JH, Kaiser JF (2001) Nonlinear feature based classification of speech under stress. IEEE Trans Speech Audio Process 9(3):201–216
65. Zhuang X, Huang J, Potamianos G, Hasegawa-Johnson M (2009) Acoustic fall detection using gaussian mixture models and gmm supervectors
66. Zigel Y, Litvak D, Gannot I (2009) A method for automatic fall detection of elderly people using floor vibrations and sound-proof of concept on human mimicking doll falls. IEEE Trans Biomed Eng 56(12):2858–2867