applied sciences
MDPI

Review

Surround by Sound: A Review of Spatial Audio Recording and Reproduction

Wen Zhang [1,2,*], Parasanga N. Samarasinghe [1], Hanchi Chen [1] and Thushara D. Abhayapala [1]

[1] Research School of Engineering, College of Engineering and Computer Science, The Australian National University, Canberra 2601 ACT, Australia; prasanga.samarasinghe@anu.edu.au(P.N.S.); hanchi.chen@anu.edu.au (H.C.); thushara.abhayapala@anu.edu.au (T.D.A.)
[2] Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an 710072, Shaanxi, China
* Correspondence: wen.zhang@anu.edu.au; Tel.: +61-2-6125-1438

Academic Editors: Woon-Seng Gan and Jung-Woo Choi
Received: 14 March 2017; Accepted: 11 May 2017; Published: 20 May 2017

Abstract: In this article, a systematic overview of various recording and reproduction techniques for spatial audio is presented. While binaural recording and rendering is designed to resemble the human two-ear auditory system and reproduce sounds specifically for a listener's two ears, soundfield recording and reproduction using a large number of microphones and loudspeakers replicate an acoustic scene within a region. These two fundamentally different types of techniques are discussed in the paper. A recent popular area, multi-zone reproduction, is also briefly reviewed in the paper. The paper is concluded with a discussion of the current state of the field and open problems.

Keywords: spatial audio; binaural recording; binaural rendering; soundfield recording; soundfield reproduction; multi-zone reproduction

1. Introduction

Spatial audio aims to replicate a complete acoustic environment, or to synthesize realistic new ones. Sound recording and sound reproduction are two important aspects in spatial audio, where not only the audio content but also the spatial properties of the sound source/acoustic environment are preserved and reproduced to create an immersive experience.

Binaural recording and rendering refer specifically to recording and reproducing sounds in two ears [1,2]. It is designed to resemble the human two-ear auditory system and normally works with headphones [3,4] or a few loudspeakers [5,6], i.e., the stereo speakers. However, it is a complicated process, as a range of localization cues, including the static individual cues captured by the individualized HRTF (head-related transfer function), dynamic cues, due to the motion of the listener, and environmental scattering cues, should be produced in an accurate and effective way for creating realistic perception of the sound in 3D space [7,8]. In gaming and personal entertainment, binaural rendering has been widely applied in Augmented Reality (AR)/Virtual Reality (VR) products like the Oculus Rift [9] and Playstation VR [10].

Soundfield recording and reproduction adopts physically-based models to analyze and synthesize acoustic wave fields [11–14]. It is normally designed to work with many microphones and a multi-channel speaker setup. With conventional stereo techniques, creating an illusion of location for a sound source is limited to the space between the left and right speakers. However, with more channels included and advanced signal processing techniques adopted, current soundfield reproduction systems can create a 3D (full-sphere) sound field within an extended region of space. Dolby Atmos (Hollywood, CA, USA) [15] and Auro 3D (Mol, Belgium) [16] are two well-known examples in this area, mainly used in commercial cinema and home theater applications.

This article gives an overview of various recording and reproduction techniques in spatial audio. Two fundamentally different types of rendering techniques are covered, i.e., binaural recording/rendering and soundfield recording/reproduction. We review both early and recent methods in terms of apparatus design and signal processing for spatial audio. We conclude with a discussion of the current state of the field and open problems.

2. Binaural Recording and Rendering

Humans have only two ears to perceive sound in a 3D space. Hence, it is intuitive to use two locally separated microphones to record audio as it is heard; and when played back through headphones or a stereo dipole, a 3D sound sensation is created for the listener. This is known as binaural recording, a specific approach of the two-channel stereo recording, where two microphones are placed at two ears either on a human head (known as *listening subject recording*) or an artificial head (known as *dummy head recording*).

The first binaural audio system was demonstrated in 1881 on a device called a Theatrophone introduced by a French engineer Clément Ader. However, there was not too much interest in this technology. It was not until 1974 that the first clip-in binaural microphones for a human subject was offered by Sennheiser and the first completely in-ear binaural microphones were offered by Sound Professionals in 1999. Nowadays, with the widespread availability of headphones and cheaper methods of recording, there has been a renewed interest in binaural spatial sound.

The sounds received in two ears are scattered and shaped by the human head, torso and ear geometry, resulting in spatial cues for binaural hearing being made available. This is fully captured by the HRTF, a filter defined in the spatial frequency domain that describes sound propagation from a specific point to the listener's ear, and a pair of HRTFs for two ears include all of the localization cues, such as the inter-aural time difference (ITD), inter-aural level difference (ILD), and spectral cues [17].

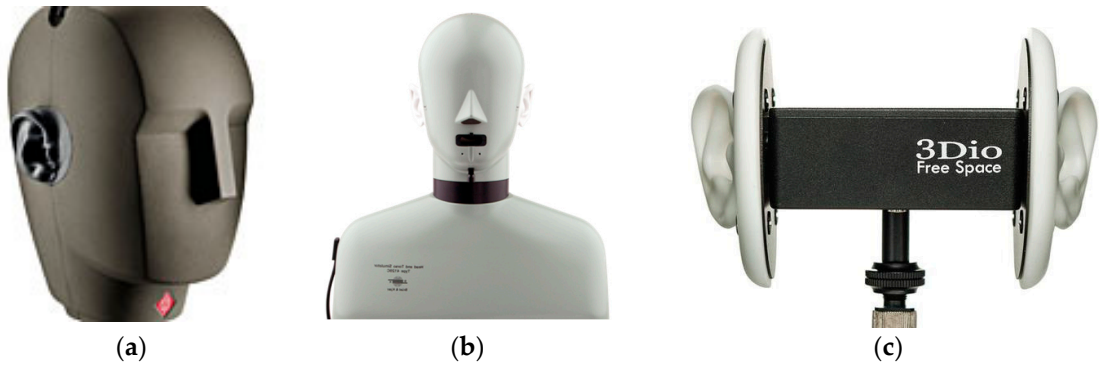
To generate spatial audio effects in binaural rendering, we can either use the recordings directly to simulate real objects or generate virtual objects by convolving a mono signal with the HRTFs corresponding to the virtual source positions. These are common practices used in Augmented Reality (AR) [3] and Virtual Reality (VR) [18].

2.1. Binaural Recording and HRTF Measurement

In both binaural recording and the HRTF measurement, a typical setup is to record acoustic properties of the dummy head or head and torso simulator (HATS), which are designed based on the average dimension of a human head/torso and have two high fidelity microphones inserted within each ear to record the two-ear signals. Some widely used binaural recording packages include (a) Dummy head: Neumann KU-100 (Berlin, Germany) [19], (b) head-and-torso simulator (HATS): Brüel & Kjær 4128D/C (Nærum, Denmark) [20], and (c) 3Dio Free Space Binaural Microphone (Vancouver, WA, USA) [21], as shown in Figure 1.

The dummy head recordings are supposed to achieve a good trade-off between individual variability and good spatial perception. However, as each human body has a unique size and shape features, it could cause confusion, especially in terms of elevation localization using the dummy head recordings directly for everyone's ears [22]. This problem is also known as one of the HRTF characteristics—that HRTFs differ greatly from person to person.

To understand the perceptual cues to spatial hearing, HRTFs are measured on both the HATS and listening subjects, and further analyzed for its characteristic variation as a function of source positions. The source positions are typically sampled at a predefined set of azimuths and elevations on a spherical surface with a radius of 1–2 m, which is when the HRTFs are assumed to be distance independent [23,24]. The whole measurement process, which involves the measurement system design/setup, is logistically complicated and takes at least 15–20 min to complete [25].



(**a**) (**b**) (**c**)

Figure 1. Binaural recording packages. (**a**) dummy head: Neumann KU-100; (**b**) head-and-torso simulator (HATS): Brüel & Kjær 4128D/C; (**c**) binaural microphones: 3Dio Free Space.

2.1.1. Measurement Resolution

One problem for the HRTF measurement is that there is a lack of recognized standards, especially in terms of the measurement resolution, resulting in the fact that databases with varying sample positions and the number of samples [26]. Even though interpolation is widely used to generate HRTFs at non-measured positions [27], the spatial resolution for HRTF measurement greatly influences the success and quality of binaural rendering.

For azimuth sampling, it is shown that an angular spacing of 5° or less is necessary to reconstruct the HRTF data up to 20 kHz in the horizontal plane (the plane perpendicular to a vertical direction and at the same height of a person's ear) [28]. Zhong and Xie [29] further studied the maximal azimuthal resolution, or equivalently the minimum number of azimuthal measurements (MNAM), at each elevation plane. The MNAM increases with increasing frequency, i.e., from five measurements at the lowest frequency to more than 60 measurements at 20 kHz in the horizontal plane, while, for a fixed frequency, it decreases as the elevation deviates from the horizontal plane. Minnaar et al. [30] investigated the directional resolution of HRTF measurement in the horizontal, frontal, and median plane. A resolution of 8° over the sphere was deemed to be enough without introducing audible interpolation errors in binaural rendering, but this resolution threshold was largely based on listening experiments.

The Spherical Harmonics, an orthogonal basis function on the sphere, has been widely used for HRTF data representation [31,32], based on which the HRTF dimensionality [33], measurement resolution [34], and sampling scheme [35] are proposed. The Spherical Harmonics based analysis shows that, in order to capture the HRTF variations for the entire audible frequency range (i.e., 20 Hz to 20 kHz), the measurement grid should have a spatial resolution of 3–4° in elevation and azimuth, with fewer measurement points required towards polar directions.

2.1.2. Test Signal and Post-Processing

Various test signals with corresponding signal processing methods have been used for HRTF measurement, such as exponential swept-sine, maximum length sequences (MLS), Golay codes, etc. Post-processing is necessary to remove the responses of the measurement system as well as reflections from the measurement apparatus or reverberation for measurements performed in a semi-anechoic chamber. The swept-sine based method with appropriate time windowing has been shown effective to remove reflections [36]; in addition, a common practice is to divide the measured transfer function by the spatially averaged mean response, i.e., the Common Transfer Function, which is irrelevant to direction, in order to extract the so-called Directional Transfer Function [1]. Another major focus is to improve the measurement efficiency, in order to minimize the recording time but without reducing the spatial resolution. More specifically, the proposed approaches include the measurement via the reciprocity method [37], the multiple exponential sweep method [25], the continuous measurement method [38], and the HRTF acquisition with unconstrained movements of human subjects [39].

An ongoing project, the "Club Fitz", was initiated in 2004 with the aim to compare HRTF databases from laboratories across the world [27]. The database consists of physical measurements and numerical simulations performed on the same dummy head (Neumann KU-100), where HRTFs are converted to the widely-adopted open standard SOFA file format [40]. Investigations on 12 different HRTFs from 10 laboratories showed an observation of large ITD variations (up to 235 μs) and spectral magnitude variations (up to 12.5–23 dB) among these databases. This further demonstrates the profound impact of measurement systems and signal processing techniques on HRTF data acquisition.

2.2. Binaural Rendering

It requires rendering static, dynamic, and environmental cues within the audio stream to create an immersive spatial audio experience [41]. As stated earlier, the static cue is fully captured by the HRTF, which varies significantly between people [42]. Thus, it is normally claimed to use individualized HRTF for binaural rendering [4]. In addition, dynamic and reverberation cues must be added to generate a virtual audio scene with maximum fidelity—for example, with appropriate externalization and localization of the virtual source. Figure 2 shows a binaural rendering system for headphones with all of the signal processing techniques reviewed in this section.

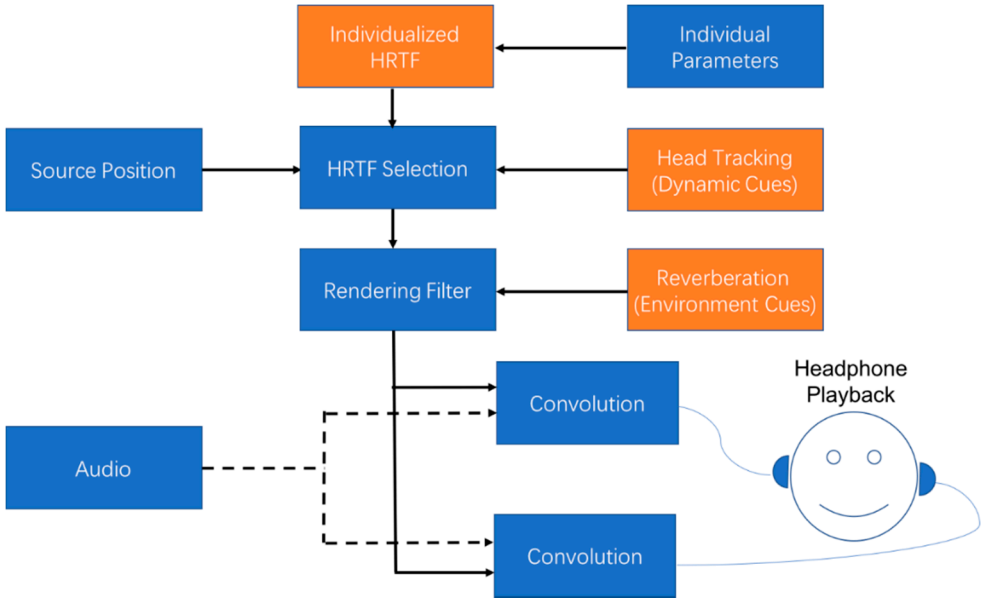


Figure 2. The signal flow of a binaural rendering system.

2.2.1. Individualized HRTF

The direct acoustic measurement on a listening subject is the most accurate and important way of obtaining the individualized HRTFs, which include all of the relevant static cues, such as ITD, ILD, and spectral cues, for sound localization in 3D, especially in terms of azimuth and elevation. However, these measurements are discrete over space and frequency, and thus should be used with interpolation for binaural rendering [43,44]. The methods can be classified as *local interpolation*, where HRTFs are computed from the measurements at its adjacent directions, i.e., bilinear interpolation [45] and inter-positional transfer function based interpolation [46], or *global interpolation*, where HRTFs are computed from all measurements based on a model using appropriate basis functions, i.e., the Spherical Harmonic based interpolation [26] and principle component analysis based interpolation [47]. Especially, the Spherical Harmonics, which are spatially continuous and orthogonal over sphere, are now widely used for HRTF representation [33,48]; based on this continuous representation, a novel spatial audio rendering technique for area and volumetric sources in VR was developed [18].

Due to the fact that HRTF individualization is strongly related to the anthropometry of a person, methods have been proposed for HRTF personalization by choosing a small set of anthropometry

features with a pre-trained model [49–54]. The training was established based on a direct linear or nonlinear relationship between the anthropometric data and the HRTFs, where the first step is to reduce the HRTF data dimensionality. The effectiveness of this kind of model is heavily dependent on the selection of the anthropometry features [55]. For example, the study in [53] investigated the contribution of the external ear to the HRTF and proposed a relation between the spectra notches and pinna contours.

Recently, researchers from Microsoft proposed an indirect method for HRTF synthesis using sparse representation [56,57]. The main idea is firstly to learn a sparse representation of a person's anthropometric features from the training set and then apply this sparse representation directly for HRTF synthesis. Further work shows that, in this method, the pre-processing and post-processing are crucial for the performance of HRTF individualization [58].

Some other techniques used for HRTF individualization include perceptual feedback (tunning-based HRTF manipulation) [59], binaural synthesis using frontal projection headphones [60], and by finite element modelling of the human head and torso to simulate the source-ear acoustic path [61]. Especially, a significant amount of work has been done in HRTF numerical simulations, such as using boundary element method [62,63] and finite difference time domain simulation [64], or estimating key localization cues of the HRTF data from anthropometric parameters [65]. Recently, a collection of 61-subject HRTFs and the high-resolution surface meshes of these subjects obtained from magnetic resonance imaging (MRI) data are released in the SYMARE database [66]. The predictions using the fast-multiple boundary element method show a high correlation with the measurement ones, especially at frequencies below 10 kHz.

Notice that, due to the logistical challenges of obtaining the individualized HRTF, the database of non-individualized HRTFs are often used in the acoustic research [67,68]. In addition, the perception using non-individualized HRTFs can be strengthened if dynamic cues are appropriately included as shown in the following section.

2.2.2. Dynamic Cues

The dynamic cue arising from the motion of the listener, which changes the relative position of the source, can reinforce localization. Studies on this cue show that the well-known front-back confusion problem in static binaural rendering disappears when the listener can turn their head to assist localization [69]. It also demonstrates that the perception using non-individualized HRTFs can be strengthened if dynamic cues are appropriately included [42]. Thus, it is necessary in binaural rendering to have the simulated scene change with the listener movement. This is normally achieved using a low-latency head tracking system, based on computer vision techniques (such as regular cameras) to estimate the orientation of the listener's head. In smartphones, the low-cost sensors, such as accelerometers and gyroscopes, can be used for head-tracking.

2.2.3. Environment Cues

Using the static cue, i.e., the HRTF, along to render the binaural signals will have one big problem that the recreated sounds are not well externalized. Dummy head recordings performed in reverberant rooms reveal that reverberation effects are essential for an immersive 3D sensation. For example, the direct-to-reverberant energy ratio enables source distance estimation in the room [70]. Thus, it is important to incorporate environmental scattering cues to achieve good externalization or distance perception [70,71].

The environmental scattering is characterized by the room transfer function (RTF) or room impulse response (RIR), a function of the source and receiver positions that includes effects due to reflection at the boundaries, sound absorption, diffraction, and room resonance. The RIR can be separated into two parts, a small number of early strong reflections and very large numbers (hundreds or even thousands) of late weak reverberation. It is believed that the early reflections are helpful for source externalization, thus must be convolved with the appropriate HRTF for the corresponding image source direction [41]. The late reverberation, however, is directionless and thus can be approximated by a generic model for a given room geometry. It is clear that room

reverberation has a strong impact on RIR simulation. Physically based reverberation models are widely used in virtual reality to reproduce the acoustics of a given real or virtual space. One example is the image source model for computing the early reflections of RIR in a rectangular room [72]. For a review on different artificial reverberation simulation methods, see, e.g., [73].

The rendering filter is normally constructed as a Finite-Impulse-Response (FIR) filter, where only the direct path and the first few reflections are computed in real time and the rest of the filter is computed once given the room geometry and boundary. For playback synthesis, simulation of multiple virtual sources at different positons can be performed in parallel, where the audio stream of each source is convolved with the corresponding FIR filter. The convolution can be performed either in the time domain to have a zero-processing lag but with high computational complexity or in the frequency domain for low computational complexity but with unavoidable latency. The synthesized results are mixed together for playback. To further reduce computational complexity, binaural rendering on headphones [74] and for interactive simulations [75] were implemented using a Graphical Processing Unit (GPU).

2.2.4. Rendering by Headphones or Loudspeakers

Headphones are the natural and most effective way to reproduce binaural sounds. Equalization is commonly applied to make them meet an ideal target response. The original reference was that of a frontal free field (i.e., free-field equalization), but a preference for a diffuse field with random incidence was later proposed (i.e., diffuse-field equalization) [76] and now has been widely adopted by the headphone manufactures. In terms of binaural rendering itself, the recent results show that only presenting the variation of the sound spectrum due to the source position changes can provide the localization cues from the perception point of view [2]. However, from a practical application point of view, compensation is necessary for accurate binaural synthesis over headphones, either through individual measurements [77] or non-individual recordings of the headphone transfer function (HpTF) [78]. The metrics for evaluating the perception of equalized HpTF was recently proposed [79].

Binaural rendering through loudspeakers is intrinsically affected by the crosstalk and thus requires a pre-processing crosstalk cancellation system (CCS) [5]. This system, however, is sensitive to the listener's head movement or misalignment given limited sweet spot. The research has focused on developing robust CCS for a two-speaker or multiple-speaker setup [6,80]. The optimal loudspeaker positions for CCS are of interest, among which an effective dual-speaker system called "stereo dipole" [81] and the optimal loudspeaker array configuration [82,83] were developed. Other recent work includes investigating the sound source localization performance provided by CCS [84] and designing CCS with a head tracker based on an online monitoring of ITD [85].

3. Soundfield Recording and Reproduction

The system that uses many loudspeakers for soundfield reproduction is principally different from the above mentioned binaural rendering. The soundfield reproduction system creates spatial audio effects (i.e., exact positions of the sound sources) within an extended region of space while a binaural audio system produces natural 3D sound at the listener's ears without an expensive set of loudspeakers. The recording and reproduction techniques for soundfield reproduction are described in the following.

3.1. Soundfield Representation

For a soundfield within a source free region, the sound pressure at any point can be expressed in the 3D spherical coordinate as [86]

$$P(\boldsymbol{x}, w) \approx \sum_{n=0}^{N} \sum_{m=-N}^{N} \alpha_n^m(w) j_n(kr) Y_n^m(\hat{\boldsymbol{x}}), \tag{1}$$

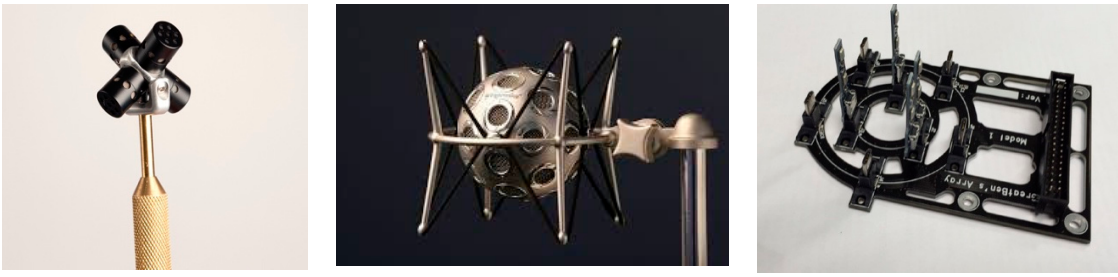
where $\alpha_n^m(w)$ are soundfield coefficients corresponding to the mode index (n, m), $j_n(kr)$ are Spherical Bessel functions of the first kind representing the mode amplitude at radius r, $Y_n^m(\hat{\boldsymbol{x}})$ are Spherical Harmonics that are functions of the angular variables. Spherical Harmonics and Spherical Bessel functions together represent the propagation modes. Especially, due to the low-pass

characteristics of the Spherical Bessel function, given the radius of a region of interest (ROI) r_0 and the wavenumber k, the truncation number $N \approx \lceil kr_0 \rceil$ [87]. This means that the soundfield within the ROI can be represented by a finite number of, i.e., $D = (N+1)^2$, coefficients.

In soundfield recording and analysis, normally, the soundfield coefficients $\alpha_n^m(w)$ and the Spherical Bessel functions $j_n(kr)$ are integrated as one component. Equation (1) becomes an expansion with respect to Spherical Harmonics solely. The expansion order N is also the order of the system. For example, when $N = 1$, the system is called the first-order system (or the widely known Ambisonics); and when $N \geq 2$, the system is called the higher-order system (or higher-order Ambisonics).

3.2. 3D Soundfield Recording

The soundfield microphone, arranged as a microphone unit composed of multiple microphone capsules and a signal processor, with 3D pick up capability is commonly used for soundfield recording. This microphone normally has a 3D geometry, such as the B-Format Ambisonic microphone [88] and EigenMike (a spherical microphone array) [89], as shown in Figure 3. The design of the microphone for soundfield recording is based on the decomposition of the 3D soundfield using Spherical Harmonics, i.e., Equation (1).



(**a**) (**b**) (**c**)

Figure 3. 3D soundfield recording microphones. (**a**) TetraMic; (**b**) EigenMike; (**c**) Planar Microphone Array.

3.2.1. 3D Microphone Array

The Ambisonic microphone was firstly designed by Dr. Jonathan Halliday at Nimbus Records (Monmouth, England) for recording the first-order Spherical Harmonics decomposition of a soundfield, i.e., B-format. The original design, known as a native or Nimbus/Halliday microphone array, has three coincident microphones, i.e., an omnidirectional microphone, one forward-facing, and one left-facing figure of eight microphone, to record the W, X, and Y components separately. Since it is impossible to build a perfectly coincident microphone array, Michael A. Gearzon developed an improved version, the tetrahedral microphone, which has four cardioid or sub-cardioid microphone capsules arranged in a tetrahedron and equalized for uniform diffuse-field response [90]. The recorded signals later are converted to B-format through a matrix operation. On the market, TetraMic developed in 2007 [88] (as shown in Figure 3a) the first portable single point, stereo and surround sound Ambisonic microphone.

Above the first-order, multiple microphones with very sophisticated digital signal processing are required to obtain the higher-order expansion components directly. An ideal higher order microphone would be comprised of a continuous spherical microphone array, which can decompose the measured 3D soundfield into its spherical harmonic coefficients. In practice, spherical microphone arrays are implemented with a discrete array that uniformly samples a spherical surface [11,12]. To record an Nth order soundfield, the minimum sampling requirement along the spherical surface is $(N+1)^2$. Note that, since the soundfield order $N \approx \lceil kr_0 \rceil$ is proportional to the frequency and the ROI size (or radius), the number of microphones required in the array is also proportional to those two quantities. The Eigenmike [89], as shown in Figure 3b, is an example for a commercially available 4th order microphone array, which consists of 32 microphones uniformly spaced on a spherical baffle of radius 4 cm. Instead of using a spherical geometry, a novel array structure consisting of a set of

parallel circular microphone arrays to decompose a wavefield into spherical harmonic components was proposed recently [91].

3.2.2. Plannar Microphone Array

Planar microphone arrays are widely used for applications such as beamforming and circular harmonic analysis of 2D spatial sound. However, generally speaking, due to the limitation of the 2D geometry, a single planar array cannot capture full 3D spatial sound.

A special planar microphone array configuration that is capable of recording 3D spatial sound was proposed by Chen et al. [92], as shown in Figure 3c. In this configuration, the combined use of omni-directional and vertically placed first order (cardioid or differential) microphones enables detection of the acoustic particle velocity in the vertical direction, which can be used to solve for the spatial soundfield components that are normally "invisible" to planar microphone arrays. It is shown that this planar microphone array offers the same capability as a spherical microphone array of the same radius, in terms of spatial sound recording.

3.2.3. Array of Higher Order Microphones

While higher order soundfield recording can be conveniently achieved via the microphone designs discussed above, with increasing size of the desired spatial region and increasing frequency, the array's minimum microphone requirement increases to impractical numbers. A recently proposed method to overcome this limitation is via utilizing an array of higher order microphones [14,93]. For example, a distributed array of fourth order Eigenmike/planar higher order microphones can replace a spherical array of 121 omnidirectional microphones. Such a design is highly suitable for spatial soundfield recording over large regions because it significantly reduces the implementation complexity of the array (reduced amount of cabling, reduced spatial samples, etc.) at the expense of increased complexity at each microphone unit.

3.3. Soundfield Reproduction

Spatial soundfield reproduction aims to create an immersive soundfield over a predefined spatial region so that the listener inside the region can experience a realistic but virtual replication of the original soundfield. This is achieved by controlling the placement of a set of loudspeakers usually put on the boundary that encloses the spatial region of interest and deriving the signals emitted from the loudspeakers. Loudspeaker array design and audio processing are two key aspects to control sound radiation and to deal with the complexity and uncertainty associated with soundfield reproduction.

3.3.1. Reproduction Methods

Given multi-channel (or multi-microphone) recordings of an acoustic scene, the *channel-based reproduction* plays these recordings with certain down/up mixing to replicate the scene. The *object-based reproduction*, on the other hand, is based on the location information of loudspeakers and the objects (or virtual source) in order to determine which loudspeakers are used and their driving signals for playing back the object's audio. The object-based reproduction formed on the panning law can render the audio objects in real time. Two well-known object-based pieces of audio equipment are stereo systems that consist of two channels, left and right, and surround sound systems that consist of multiple channels surrounding the listener. The stereo can provide good imaging (the perceived spatial location of the sound source) in the front quadrant while the surround sound can offer imaging around the listener.

Based on the amplitude panning principle [94,95], the audio object can be positioned in an arbitrary 2D or 3D setup using the vector base amplitude panning (VBAP) or distance-based amplitude panning (DBAP). The VBAP is based on a triplet-wise panning law, where three loudspeakers are arranged in a triangle layout to generate sound imaging [96]. The DBAP only takes the positions of the virtual source and loudspeakers into account for reproduction and thus is well

suited to irregular loudspeaker layouts [97]. In object-based reproduction, the listener normally is required to be at the centre of the speaker array (i.e., the sweep spot position) where the generated audio effect is best; DBAP, however, is a noticeable exception to this restriction.

Table 1. Summary of soundfield reproduction methods and commercial systems.

Method	Typical Systems	Characteristics
Stereo/Surround	Dolby Stereo, Dolby 5.1/7.1 Dolby Atmos, NHK 22.2 Auro 3D	■ Channel-based & Object-based reproduction ■ Amplitude/phase encoding
VBAP, DBAP	Software Demo [98]	■ Object-based reproduction ■ Amplitude encoding
Ambisonics (B-format) HOA	Youtube 360° VR Audio Kit [99]	■ Model-based reproduction ■ Amplitude/phase encoding
WFS	IOSONO	■ Model-based reproduction ■ Amplitude/phase encoding
Inverse Filtering	--	■ Channel-based reproduction ■ Amplitude/phase encoding

VBAP: vector base amplitude panning; DBAP: distance-based amplitude panning; HOA: higher order Ambisonics; WFS: wave-field synthesis.

The first known demonstration of reproducing a soundfield within a given region of space was conducted by Camras at the Illinois Institute of Technology in 1967, where loudspeakers were distributed on the surface enclosing the selected region and the listeners can move freely within the region [100]. Later, the well-known Ambisonics was designed based on Huygen's principle for more advanced spatial soundfield reproduction over a large region of space [101,102]. The system is based on the zero and first order spherical harmonic decomposition of the original soundfield into four channels, i.e., Equation (1), and from a linear combination of these four channels to derive the loudspeaker driving signals. This low-order system is optimum at low frequencies but less accurate at high frequencies and when the listener is away from the center point. Higher Order Ambisonics (HOA) based on the higher order decomposition of a soundfield, such as cylindrical two-dimensional (2D or horizontal plane) harmonic [87,103,104] or spherical three-dimensional (3D or full sphere) harmonic [13,105,106] decomposition, was developed especially for high reproduction frequencies and large reproduction regions. Based on the same principle, soundfield reproduction using the plane wave decomposition approach was recently proposed [107–109].

Wave-Field Synthesis (WFS) is another well-known sound reproduction technique initially conceived by Berkhout [110,111]. The fundamental principle is based on the Kirchhoff–Helmholtz integral to represent a soundfield in the interior of a bounded region of the space by a continuous distribution of monopole and normally oriented dipole secondary sources, arranged on the boundary of that region [112]. An array of equally spaced loudspeakers is used to approximate the continuous distribution of secondary sources. Reproduction artifacts due to the finite size of the array and the spatial discretization of the ideally continuous distribution of secondary sources were investigated [113]. The WFS technique has been mainly implemented in 2D sound reproduction using linear and planar arrays [114,115], for which a 2.5D operator was proposed to replace the secondary line sources by point sources (known as 2.5D WFS) [112,116,117]. In WFS, a large number of closely spaced loudspeakers is necessary.

For arbitrarily placed loudspeakers, a simple approach in sound reproduction known as inverse filtering is to use multiple microphones as matching points based on the least-square match to derive the loudspeaker weights [118], with the knowledge of the acoustic channel between the loudspeakers and matching points, i.e., the RTF. Tikhonov regularization is the common method for obtaining loudspeaker weights with limited energy and also for improving the system robustness. In addition, the fluctuation of the RTF requires an accurate online estimation of the inverse filter coefficients in this method [119].

Table 1 summarises the above mentioned soundfield reproduction techniques. A comparison of different soundfield reproduction techniques especially from the perception point of view was

presented in the review paper [120]. One of the current research interests is to further improve these techniques with a thorough perceptual assessment [121].

3.3.2. Listening Room Compensation

In an acoustic reverberant environment, the multi-path propagation effect introduces echoes and spectral distortions into the generated soundfield. Room equalization has been studied in theory and applied in practice to cancel this effect for sound reproduction in cinema halls, home theaters and teleconferencing applications [122]. An efficient method for correcting room reverberation is by using active room compensation, that is, by applying compensation signals to the loudspeaker input to make a reverberant room problem look like an anechoic room problem. This method, however, requires the knowledge of the underlying acoustic system, i.e., the RIR (room impulse response) or the RTF (room transfer function).

Compensation schemes based on RIR/RTF modelling are theoretically capable of good performance; however, imperfections in the modelling process generally lead to a reduction of dereverberation performance in real environments [123]. This is further exacerbated by non-static room conditions, e.g., time-variant room responses because of a temperature variation or source/receiver position variations [124,125]. For soundfield reproduction in time-varying environments, a multiple-loudspeaker-multiple-microphone setup is employed and the RIR/RTF of this acoustic system must be determined online in an adaptive manner. In addition, as the purpose here is to reproduce sound over a region of interest, room compensation should be achieved essentially within the entire region as well.

In massive multichannel soundfield reproduction systems, for which the number of loudspeakers and microphones are large, active room compensation can be solved computationally and efficiently by using a wave-domain approach [103,126]. The principle of wave-domain signal representation is to use fundamental solutions of the Helmholtz wave-equation as basis functions to express a wave field over a spatial region as shown in Label (1). Processing directly on the decomposition coefficients therefore controls sound within the region.

The wave-domain adaptive filtering (WDAF) approach transforms the signals at the microphones and the loudspeaker signals into the wave domain, and then adaptively calculates the loudspeaker compensations signals (Figure 4). Especially, in the wave domain, the compensation filter is forced to be diagonal, and each diagonal entry can be determined from the decoupled adaptive filters [127,128]. This technique results in parallel implementation and significantly reduces the complexity of the adaption process.

More complex approaches for room compensation are by using fixed or variable directivity higher-order loudspeakers to minimize the acoustic energy directed towards the walls of a room [129,130], or by exploiting room reflections to reproduce a desired soundfield [131–133].

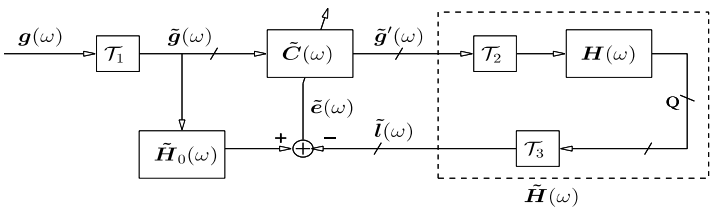
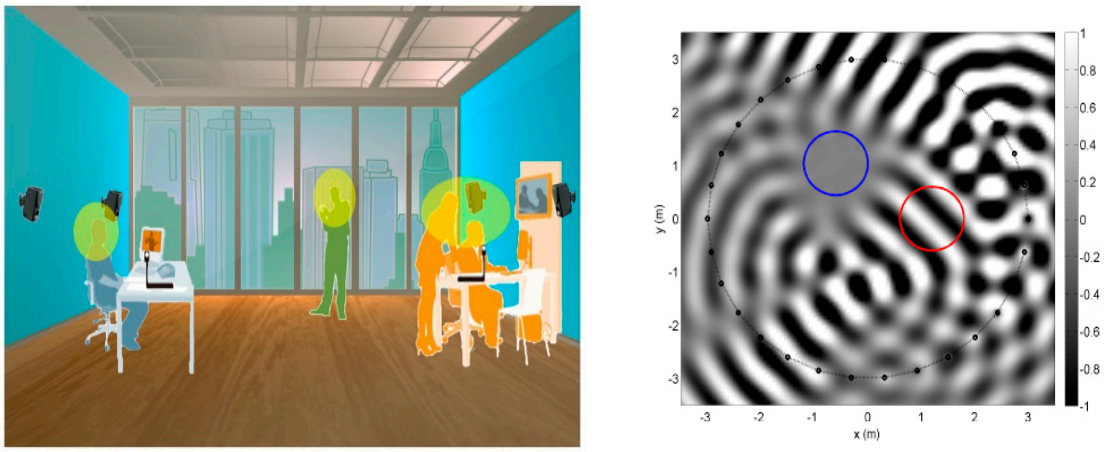


Figure 4. The listening room compensation using WDAF (wave-domain adaptive filtering). The free-field transformed loudspeaker signals $\tilde{\boldsymbol{g}}$ are used in a reverberant room with the filter matrix $\tilde{\boldsymbol{C}}$ to compensate the RTFs in matrix $\tilde{\boldsymbol{H}}$, $\boldsymbol{\mathcal{T}_1}, \boldsymbol{\mathcal{T}_3}$ and $\boldsymbol{\mathcal{T}_2}$ represent the forward and backward WDAF, respectively.

4. Multi-Zone Sound Reproduction

Multi-zone reproduction aims to extend spatial sound reproduction over multiple regions so that different listeners can enjoy their audio material simultaneously and independently of each other but without physical isolation or using headphones (Figure 5). The concept of multi-zone soundfield control has recently drawn attention due to a whole range of audio applications, such as controlling sound radiation from a personal audio device, creating independent sound zones in different kinds of enclosures (such as shared offices, private transportation vehicles, exhibition centres, etc.), and generating quiet zones in a noisy environment. A single array of loudspeakers is used, where sound zones can be placed at any desired location and the listener can freely move between zones; thus, the whole system provides significant freedom and flexibility.



(**a**) (**b**)

Figure 5. (**a**) an illustration of multi-zone sound reproduction in an office environment; (**b**) a plane wave of 500 Hz from 45° is reproduced in the bright zone (red circle) with a dark or quiet zone (blue circle) generated using a circular array of 30 loudspeakers [134].

The multi-zone reproduction was firstly formulated as creating two kinds of sound zones, the bright zone within which certain sounds with high acoustic energy are reproduced and the dark zone (or the quiet zone) within which the acoustic energy is kept at a low level [135]. The proposed method is to maximise the ratio of the average acoustic energy density in the bright zone to that in the dark zone, which is known as the acoustic contrast control (ACC) method. Since then, different forms of contrast control based on the same principle have been proposed, including an acoustic energy difference formulation [136], direct and indirect acoustic contrast formulations using the Lagrangian [137]. The technique has been implemented in different personal audio systems in an anechoic chamber [138,139] or in a car cabin [140]; over 19 dB contrast was achieved under the ideal condition, while, for real-time systems in the car cabin, the acoustic contrast was limited to a maximum value of 15 dB. This contrast control method, however, does not impose a constraint on the phase of the soundfield and thus cannot control the spatial aspects of the reproduced soundfield in the bright zone. A recent work by Coleman et al. proposed refining the cost function of the ACC with the aim of optimizing the extent to which the reproduced soundfield resembles a plane wave, thus optimising the spatial aspects of the soundfield [141]. Another issue in ACC is the self-cancellation problem, which results in a standing wave produced within the bright zone [142].

The pressure matching (PM) approach aims to reproduce a desired soundfield in the bright zone while producing silence in the dark zone [143]. The approach uses a sufficiently dense distribution of microphones within all the zones as the matching points and adopts the least-squares method to control the pressure at each point. A constraint on the loudspeaker weight energy (or the array effort) is added to control the sound leakage outside the sound zones and to ensure the implementation is robust against speaker positioning errors and changes in the acoustic environment [144]. When the desired soundfeld in the bright zone is due to a few virtual source directions,

the multi-zone sound control problem can be solved using a compressive sensing idea dolwhere the loudspeaker weights are regularised with the L_1 norm. This results in only a few loudspeakers placed closely to the virtual source directions being activated for reproduction [145,146]. More recent works have been focusing on the combination of the ACC and PM formulations using the Lagrangian with a weighting factor to tune the trade-off between the two performance measures, i.e., the acoustic contrast and bright zone error (i.e., the reproduction error within the bright zone) [147–149]. The idea of performing time-domain filters for personal audio was recently investigated [150].

The multi-zone reproduction is formulated in the modal domain based on representing the soundfield within each zone through a spatial harmonic expansion, i.e., Equation (1). The local sound field coefficients are then transformed to an equivalent global soundfield coefficient using the harmonic translation theroem, from which the loudspeaker signals are obtained throung the mode matching [151]. The modal-domain approach can provide theoretical insights into the multi-zone problem. For example, through the modal domain analysis, a theoretical basis is established for creating two sound zones with no interference [152]. Modal-domain sparsity analysis shows that a significantly reduced number of microphone points could be used quite effectively for multi-zone reproduction over a wide frequency range [146]. The synthesis of soundfields with distributed modal constraints and quiet zones having an arbitrary predefined shape have also been investigated [153,154]. Based on modal-domain analysis, a parameter, the coefficient of realisability, is developed to indicate the achievable reproduction performance given the sound zone geometry and the desired soundfield in the bright zone [155].

5. Conclusions

In this article, we presented the recording and reproduction techniques for spatial audio. The techniques that have been explored include binaural recording and rendering, soundfield recording and reproduction, and multi-zone reproduction. Binaural audio that works with a pair of headphones or a few loudspeakers has the advantage of being easily incorporated into the personal audio products, while soundfield reproduction techniques that rely on many loudspeakers to control sounds within a region are mainly used in commercial and professional audio applications. Both techniques strive for the best immersive experience; however, in soundfield reproduction, the spatial audio effect is not restricted to a single user or some spatial points. Therefore, a wide range of emerging research directions has appeared in this area, such as multi-zone reproduction and active noise control over space.

In binaural techniques, generating individualized dynamic auditory scenes are still challenging problems. Future research directions include adaptation to an individualized HRTF based on the available datasets, real-time room acoustic simulations for a natural AR/VR listening experience, and sound source separation for augmented audio reality. In soundfield reproduction, interference mitigation and room compensation robust to acoustic environment changes remain as the major challenges. Further opportunities exist in higher-order surround sound using an array of directional sources and wave-domain active room compensation to perform sound field reproduction in reverberant rooms.

Acknowledgments: The authors acknowledge National Natural Science Foundation of China (NSFC) No. 61671380 and Australian Research Council Discovery Scheme DE 150100363.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hammershoi, D.; Møller, H. Methods for binaural recording and reproduction. *Acta Acust. United Acust.* **2002**, *88*, 303–311.
2. Møller, H. Fundamentals of binaural technology. *Appl. Acoust.* **1992**, *36*, 171–218.
3. Ranjan, R.; Gan, W.-S. Natural listening over headphones in augmented reality using adaptive filtering techniques. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1988–2002.

4. Sunder, K.; He, J.; Tan, E.-L.; Gan, W.-S. Natural sound rending for headphones: Integration of signal processing techniques. *IEEE Signal Process. Mag.* **2015**, *23*, 100–114.
5. Bauer, B.B. Stereophonic earphones and binaural loudspeakers. *J. Acoust. Soc. Am.* **1961**, *9*, 148–151.
6. Huang, Y.; Benesty, J.; Chen, J. On crosstalk cancellation and equalization with multiple loudspeakers for 3-D sound reproduction. *IEEE Signal Process. Lett.* **2007**, *14*, 649–652.
7. Ahveninen, J.; Kopčo, N.K.; Jääskeläinen, I.P. Psychophysics and neuronal bases of sound localization in humans. *Hear. Res.* **2014**, *307*, 86–97.
8. Kolarik, A.J.; Moore, B.C.J.; Zahorik, P.; Cirstea, S.; Pardhan, S. Auditory distance perception in humans: A review of cues, development, neuronal bases and effects of sensory loss. *Atten. Percept. Pyschophys.* **2016**, *78*, 373–395.
9. Oculus Rift | Oculus. Available online: https://www.oculus.com/rift/ (accessed on 26 April 2017).
10. PlayStation VR—Virtual Reality Headset for PS4. Available online: https://www.playstation.com/en-us/explore/playstation-vr/ (accessed on 26 April 2017).
11. Abhayapala, T.D.; Ward, D.B. Theory and design of high order sound field microphones using spherical microphone array. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 2002; pp. 1949–1952.
12. Meyer, J.; Elko, G. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL, USA, 2002; pp. 1781–1784.
13. Poletti, M.A. Three-dimensional surround sound systems based on spherical harmonics. *J. Audio Eng. Soc.* **2005**, *53*, 1004–1025.
14. Samarasinghe, P.N.; Abhayapala, T.D.; Poletti, M.A. Spatial soundfield recording over a large area using distributed higher order microphones. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2011; pp. 221–224.
15. Dolby Atmos Audio Technology. Available online: https://www.dolby.com/us/en/brands/dolby-atmos.html (accessed on 26 April 2017).
16. Auro-3D/Auro Technologies: Three-dimensional sound. Available online: http://www.auro-3d.com/ (accessed on 26 April 2017).
17. Cheng, C.I.; Wakefield, G.H. Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in time, frequency and space. *J. Audio Eng. Soc.* **2001**, *49*, 231–249.
18. Schissler, C.; Nicholls, A.; Mehra, R. Efficient HRTF-based spatial audio for area and volumetric sources. *IEEE Trans. Vis. Comput. Gr.* **2016**, *22*, 1356–1366.
19. Neumann—Current Microphones, Dummy Head KU-100 Description. Available online: http://www.neumann.com/?lang=en&id=current_microphones&cid=ku100_description (accessed on 10 March 2017).
20. Brüel & Kjær—4128C, Head and Torso Simulator HATS. Available online: http://www.bksv.com/Products/transducers/ear-simulators/head-and-torso/hats-type-4128c?tab=overview (accessed on 10 March 2017).
21. 3Dio—The Free Space Binaural Microphone. Available online: http://3diosound.com/index.php?main_page=product_info&cPath=33&products_id=45 (accessed on 10 March 2017).
22. Wenzel, E.M.; Arruda, M.; Kistler, D.J.; Wightman, F.L. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* **1993**, *94*, 111–123.
23. Brungart, D.S. Near-field virtual audio displays. *Presence Teleoper. Virtual Environ.* **2002**, *11*, 93–106.
24. Otani, M.; Hirahara, T.; Ise, S. Numerical study on source distance dependency of head-related transfer functions. *J. Acoust. Soc. Am.* **2009**, *125*, 3253–3261.
25. Majdak, P.; Balazs, P.; Laback, B. Multiple exponential sweep method for fast measurment of head related transfer functions. *J. Audio Eng. Soc.* **2007**, *55*, 623–630.
26. Andreopoulou, A.; Begault, D.R.; Katz, B.F.G. Inter-laboratory round robin HRTF measurement comparison. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 895–906.
27. Duraiswami, R.; Zotkin, D.N.; Gumerov, N.A. Interpolation and range extrapolation of HRTFs. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, QC, Canada, 2004; pp. 45–48.
28. Ajdler, T.; Faller, C.; Sbaiz, L.; Vetterli, M. Sound field analysis along a circle and its applications to HRTF interpolation. *J. Audio Eng. Soc.* **2008**, *56*, 156–175.
29. Zhong, X.L.; B.S., X. Maximal azimuthal resolution needed in measurements of head-related transfer functions. *J. Acoust. Soc. Am.* **2009**, *125*, 2209–2220.

30. Minnaar, P.; Plogsties, J.; Christensen, F. Directional resolution of head-related transfer functions required in binaural synthesis. *J. Audio Eng. Soc.* **2005**, *53*, 919–929.
31. Zhang, W.; Abhayapala, T.D.; Kennedy, R.A.; Duraiswami, R. Modal expansion of HRTFs: Continuous representation in frequency-range-angle. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, 19–24 April 2009; pp 285–288.
32. Zhang, M.; Kennedy, R.A.; Abhayapala, T.D. Empirical determination of frequency representation in spherical harmonics-based HRTF functional modeling. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 351–360.
33. Zhang, W.; Abhayapala, T.D.; Kennedy, R.A.; Duraiswami, R. Insights into head-related transfer function: Spatial dimensionality and continuous representation. *J. Acoust. Soc. Am.* **2010**, *127*, 2347–2357.
34. Zhang, W.; Zhang, M.; Kennedy, R.A.; Abhayapala, T.D. On high-resolution head-related transfer function measurements: An efficient sampling scheme. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 575–584.
35. Bates, A.P.; Z.;, K.; Kennedy, R.A. Novel sampling scheme on the sphere for head-related transfer function measurements. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1068–1081.
36. Muller, A.; Massarani, P. Transfer-function measurement with sweeps. *J. Audio Eng. Soc.* **2001**, *49*, 443–471.
37. Zotkin, D.N.; Duraiswami, R.; Grassi, E.; Gumerov, N.A. Fast head-related transfer function measurement via reciprocity. *J. Acoust. Soc. Am.* **2006**, *120*, 2202–2215.
38. Fukudome, K.; Suetsugu, T.; Ueshin, T.; Idegami, R.; Takeya, K. The fast measurment of head related impulse responses for all azimuthal directions using the continuous measurement method with a servoswiveled chair. *Appl. Acoust.* **2007**, *68*, 864–884.
39. He, J.; Ranjan, R.; Gan, W.-S. Fast continuous HRTF acquisition with unconstrained movements of human subjects. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 321–325.
40. Majdak, P.; Iwaya, Y.; Carpentier, T. Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions. In Proceedings of the 134th Audio Engineering Society Convention, Rome, Italy, 4–7 May 2013; pp. 1–11.
41. Zotkin, D.N.; Duraiswami, R.; Davis, L.S. Rendering localized spatial audio in a virtual auditory scene. *IEEE Trans. Multimedia* **2004**, *6*, 553–563.
42. Xie, B. *Head-Related Transfer Function and Virtual Auditory Display*; J Ross Publishing: Plantation, FL, USA, 2013.
43. Gamper, H. Head-related transfer function interpolation in azimuth, elevation, and distance. *J. Acoust. Soc. Am.* **2013**, *134*, EL547–554.
44. Queiroz, M.; de Sousa, G.H.M.A. Efficient binaural rendering of moving sound sources using HRTF interpolation. *J. New Music Res.* **2011**, *40*, 239–252.
45. Savioja, L.; Huopaniemi, J.; Lokki, T.; Väänänen, R. Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.* **1999**, *47*, 675–705.
46. Freeland, F.P.; Biscinho, L.W.P.; Diniz, P.S.R. Efficient HRTF interpolation in 3D moving sound. In Proceedings of the 22nd AES International Conference: Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 15–17 June 2002; pp. 1–9.
47. Kistler, D.J.; Wightman, F.L.L. A model of HeadRelated Transfer Functions based on Principal Components Analysis and Minimum-Phase reconstruction. *J. Acoust. Soc. Am.* **1992**, *91*, 1637–1647.
48. Romigh, G.D.; Brungart, D.S.; Stern, R.M.; Simpson, B.D. Efficient real spherical harmonic representation of head-related transfer function. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 921–930.
49. Zotkin, D.N.; Hwang, J.; Duraiswami, R.; Davis, L.S. HRTF personalization using anthropometric measurements. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 19–22 October 2003; pp. 157–160.
50. Hu, H.; Zhou, L.; Ma, H.; Wu, Z. HRTF personalization based on airtificial neural network in individual virtual auditory space. *Appl. Acoust.* **2008**, *69*, 163–172.
51. Li, L.; Huang, Q. HRTF personalization modeling based on RBF neural network. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013; pp. 3707–3710.
52. Grindlay, G.; Vasilescu, M.A.O. A multilinear (tensor) framework for HRTF analysis and synthesis. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, HI, USA, 16–20 April 2007; pp. 161–164.

53. Spagnol, S.; Geronazzo, M.; Avanzini, F. On the relation between pinna reflection patterns and head-related transfer functon features. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 508–519.
54. Geronazzo, M.; Spagnol, S.; Bedin, A.; Avanzini, F. Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4496–4500.
55. Zhang, M.; Kennedy, R.A.; Abhayapala, T.D.; Zhang, W. Statistical method to identify key anthropometric parameters in HRTF individualization. In Proceedings of the Hands-free Speech Communication and Microphone Arrays (HSCMA), Edinburgh, UK, 30 May–1 June 2011; pp. 213–218.
56. Bilinski, P.; Ahrens, J.; Thomas, M.R.P.; Tasheve, I.J.; Platt, J.C. HRTF magnitude synthesis via sparse representation of anthropometric features. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 4468–4472.
57. Tasheve, I.J. HRTF phase synthesis via sparse representation of anthropometric features. In Proceedings of the Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 9–14 February 2014; pp. 1–5.
58. He, J.; Gan, W.-S.; Tan, E.-L. On the preprocessing and postprocessing of HRTF individualizaion based on sparse representation of anthropometric features. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 639–643.
59. Fink, K.J.; Ray, L. Individualization of head related transfer functions using principal component analysis. *Appl. Acoust.* **2015**, *87*, 162–173.
60. Sunder, K.; Tan, E.-L.; Gan, W.-S. Individualization of binaural synthesis using frontal projection headphones. *J. Audio Eng. Soc.* **2013**, *61*, 989–1000.
61. Cai, T.; Rakerd, B.; Hartmann, W.M. Computing interaural differences through finite element modeling of idealized human heads. *J. Acoust. Soc. Am.* **2015**, *138*, 1549–1560.
62. Katz, B.F.G. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.* **2001**, *110*, 2440–2448.
63. Otani, M.; Ise, S. Fast calculation system specialized for head-related transfer function based on boundary element method. *J. Acoust. Soc. Am.* **2006**, *119*, 2589–2598.
64. Prepeliță, S.; Geronazzo, M.; Avanzini, F.; Savioja, L. Influence of voxelization on finite difference time domain simulations of head-related transfer functions. *J. Acoust. Soc. Am.* **2016**, *139*, 2489–2504.
65. Mokhtari, P.; Takemoto, H.; Nishimura, R.; Kato, H. Preliminary estimation of the first peak of HRTFs from pinna anthropometry for personalized 3D audio. In Proceedings of the 5th International Conference on Three Dimensional Systems and Applications, Osaka, Japan, 26–28 June 2013; p. 3.
66. Jin, C.T.; Guillon, P.; Epain, N.; Zolfaghari, R.; van Schaik, A.; Tew, A.I.; Hetherington, C.; Thorpe, J. Creating the sydney york morphological and acoustic recordings of ears database. *IEEE Trans. Multimedia* **2014**, *16*, 37–46.
67. Voss, P.; Lepore, F.; Gougoux, F.; Zatorre, R.J. Relevance of spectral cues for auditory spatial processing in the occipital cortex of the blind. *Front. Psychol.* **2011**, *2*, 48.
68. Kolarik, A.J.; Cirstea, S.; Pardhan, S. Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues. *J. Acoust. Soc. Am.* **2013**, *134*, 3395–3398.
69. Wightman, F.L.; Kistler, D.J. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *J. Acoust. Soc. Am.* **1999**, *102*, 2325–2332.
70. Kolarik, A.J.; Cirstea, S.; Pardhan, S. Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. *Exp. Brain Res.* **2013**, *224*, 623–633.
71. Shinn-Cunningham, B.G. Distance cues for virtual auditory space. In Proceedings of the IEEE Pacific Rim Conference (PRC) on Multimedia, Sydney, Australia, 26-28 August 2001; pp. 227–230.
72. Allen, J.B.; Berkeley, D.A. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* **1979**, *75*, 943–950.
73. Valimaki, V.; Parker, J.D.; Savioja, L.; Smith, J.O.; Abel, J.S. Fifty years of artificial reverberation. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 1421–1448.
74. Belloch, J.A.; Ferrer, M.; Gonzalez, A.; Martinez-Zaldivar, F.J.; Vidal, A.M. Headphone-based virtual spatialization of sound with a GPU accelerator. *J. Audio Eng. Soc.* **2013**, *61*, 546–561.
75. Taylor, M.; Chandak, A.; Mo, Q.; Lauterbach, C.; Schissler, C.; Manocha, D. Guided multiview ray tracing for fast auralization. *IEEE Trans. Vis. Comput. Gr.* **2012**, *18*, 1797–1810.

76. Theile, G. On the standardization of the frequency response of high-quality studio headphones. *J. Audio Eng. Soc.* **1986**, *34*, 959–969.
77. Hiipakka, M.; Kinnari, T.; Pulkki, V. Estimating head-related transfer functions of human subjects from pressure-velocity measurements. *J. Acoust. Soc. Am.* **2012**, *13*, 4051–4061.
78. Lindau, A.; Brinkmann, F. Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings. *J. Audio Eng. Soc.* **2012**, *60*, 54–62.
79. Boren, B.; Geronazzo, M.; Brinkmann, F.; Choueiri, E. Coloration metrics for headphone equalization. In Proceedings of the 21st International Conference on Auditory Display, Graz, Austria, 6–10 July 2015; pp. 29–34.
80. Takeuchi, T.; Nelson, P.A.; Hamada, H. Robustness to head misalignment of virtual sound imaging system. *J. Acoust. Soc. Am.* **2001**, *109*, 958–971.
81. Kirkeby, O.; Nelson, P.A.; Hamada, H. Local sound field reproduction using two closely spaced loudspeakers. *J. Acoust. Soc. Am.* **1998**, *104*, 1973–1981.
82. Takeuchi, T.; Nelson, P.A. Optimal source distribution for binaural synthesis over loudspeakers. *J. Acoust. Soc. Am.* **2002**, *112*, 2786–2797.
83. Bai, M.R.; Tung, W.W.; Lee, C.C. Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the generic algorithm. *J. Acoust. Soc. Am.* **2005**, *117*, 2802–2813.
84. Majdak, P.; Masiero, B.; Fels, J. Sound localization in individualized and non-individualized crosstalk cancellation systems. *J. Acoust. Soc. Am.* **2013**, *133*, 2055–2068.
85. Lacouture-Parodi, Y.; Habets, E.A. Crosstalk cancellation system using a head tracker based on interaural time differences. In Proceedings of the International Workshop on Acoustic Signal Enahcancement, Aachen, Germany, 4–6 September 2012; pp. 1–4.
86. Williams, E.G. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*; Academic Press: San Diego, CA, USA, 1999.
87. Ward, D.B.; Abhayapala, T.D. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 697–707.
88. Core Sound TetraMic. Available online: http://www.core-sound.com/TetraMic/1.php (accessed on 10 March 2017).
89. Eigenmike Microphone. Available online: https://www.mhacoustics.com/products#eigenmike1 (accessed on 10 March 2017).
90. Gerzon, M.A. The design of precisely conincident microphone arrays for stereo and surround sound. In Proceedings of the 50th Audio Engineering Society Covention, London, UK, 4–7 March 1975; pp. 1–5.
91. Abhayapala, T.D.; Gupta, A. Spherical harmonic analysis of wavefields using multiple circular sensor arrays. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *18*, 1655–1666.
92. Chen, H.; Abhayapala, T.D.; Zhang, W. Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis. *J. Acoust. Soc. Am.* **2015**, *138*, 3081–3092.
93. Samarasinghe, P.N.; Abhayapala, T.D.; Poletti, M.A. Wavefield analysis over large areas using distributed higher order microphones. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 647–658.
94. Pulkki, V.; Karjalainen, M. Localization of amplitude-panned virtual sources, Part 1: Stereophonic panning. *J. Audio Eng. Soc.* **2001**, *49*, 739–752.
95. Pulkki, V. Localization of amplitude-panned virtual sources, Part 2: Two and three dimensional panning. *J. Audio Eng. Soc.* **2001**, *49*, 753–767.
96. Pulkki, V. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* **1997**, *45*, 456–466.
97. Lossius, T.; Baltazar, P.; de la Hogue, T. DBAP—distance-based amplitude panning. In Proceedings of the 2009 International Computer Music Conference, Montreal, QC, Canada, 16–21 August 2009; pp. 1–4.
98. VBAP Demo. Available online: http://legacy.spa.aalto.fi/software/vbap/VBAP_demo/ (accessed on 26 April 2017).
99. Developers—3D Sound Labs. Availabe online: http://www.3dsoundlabs.com/category/developers/ (accessed on 26 April 2017).
100. Cameras, M. Approach to recreating a sound field. *J. Acoust. Soc. Am.* **1967**, *43*, 1425–1431.
101. Gerzon, M.A. Periphony: With-height sound reproduction. *J. Audio Eng. Soc.* **1973**, *21*, 2–10.
102. Gerzon, M.A. Ambisonics in multichannel broadcasting video. *J. Audio Eng. Soc.* **1985**, *33*, 859–871.

103. Betlehem, T.; Abhayapala, T.D. Theory and design of sound field reproduction in reverberant rooms. *J. Acoust. Soc. Am.* **2005**, *117*, 2100–2111.
104. Wu, Y.; Abhayapala, T.D. Theory and design of soundfield reproducion using continuous loudspeakers concept. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *17*, 107–116.
105. Daniel, J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In Proceedings of the 23rd AES International Conference: Signal Processing in Audio Recording and Reproduction, Copenhagen, Denmark, 23–25 May 2003.
106. Ahrens, J.; Spors, S. Applying the ambisonics approach to planar and linear distributions of secondary sources and combinations thereof. *Acta Acust. United Acust.* **2012**, *98*, 28–36.
107. Ahrens, J.; Spors, S. Wave field synthesis of a sound field described by spherical harmonics expansion coefficients. *J. Acoust. Soc. Am.* **2012**, *131*, 2190–2199.
108. Bianchi, L.; Antonacci, F.; Sarti, A.; Turbaro, S. Model-based acoustic rendering based on plane wave decomposition. *Appl. Acoust.* **2016**, *104*, 127–134.
109. Okamoto, T. 2.5D higher-order Ambisonics for a sound field described by angular spectrum coefficients. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016; pp. 326–330.
110. Berkhout, A.J. A holographic approach to acoustic control. *J. Audio Eng. Soc.* **1988**, *36*, 977–995.
111. Berkhout, A.J.; de Vries, D.; Vogel, P. Acoustic control by wave field synthesis. *J. Acoust. Soc. Am.* **1993**, *93*, 2764–2778.
112. Spors, S.; Rabenstein, R.; Ahrens, J. The theory of wave field synthesis revisited. In Proceedings of the 124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 17–20 May 2008.
113. Spors, S.; Rabenstein, R. Spatial aliasing aritifacts produced by linear and circular loudspeaker arrays used for wave field synthesis. In Proceedings of the 120th Audio Engineering Society Convention, Paris, France, 20–23 May 2006.
114. Boone, M.M.; Verheijen, E.N.G.; Tol, P.F.V. Spatial sound-field reproduction by wave-field synthesis. *J. Audio Eng. Soc.* **1995**, *43*, 1003–1012.
115. Boone, M.M. Multi-actuator panels (MAPs) as loudspeaker arrays for wave field synthesis. *J. Audio Eng. Soc.* **2004**, *52*, 712–723.
116. Spors, S.; Ahrens, J. Analysis and improvement of pre-equalization in 2.5-dimensional wave field synthesis. In Proceedings of the 128 Audio Engineering Society Convention, London, UK, 23–25 May 2010.
117. Firtha, G.; Fiala, P.; Schultz, F.; Spors, S. Improved referencing schemes for 2.5D wave field synthesis driving functions. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2017**, *25*, 1117–1127.
118. Kirkeby, O.; Nelson, P.A. Reproduction of plane wave sound fields. *J. Acoust. Soc. Am.* **1993**, *94*, 2992–3000.
119. Tatekura, Y.; Urata, S.; Saruwatari, H.; Shikano, K. On-line relaxation algorithm applicable to acoustic fluctuation for inverse filter in multichannel sound reproduction system. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2005**, *E88-A*, 1747–1756.
120. Spors, S.; Wierstorf, H.; Raake, A.; Melchior, F.; Frank, M.; Zotter, F. Spatial sound with loudspeakers and its perception: A review of the current state. *Proc. IEEE* **2013**, *101*, 1920–1938.
121. Wierstorf, H. *Perceptual Assessment of Sound Field Synthesis*; Technical University of Berlin: Berlin, Germany, 2014.
122. Bharitkar, S.; Kyriakakis, C. *Immersive Audio Signal Processing*; Springer: New York, NY, USA, 2006.
123. Corteel, E.; Nicol, R. Listening room compensation for wave field sysnthesis. What can be done? In Proceedings of the 23rd Audio Engineering Society Convention, Copenhagen, Denmark, 23–25 May 2003.
124. Mourjopoulos, J.N. On the variation and invertibility of room impulse response functions. *J. Sound Vib.* **1985**, *102*, 217–228.
125. Hatziantoniou, P.D.; Mourjopoulos, J.N. Erros in real-time room acoustics dereverberation. *J. Audio Eng. Soc.* **2004**, *52*, 883–889.
126. Spors, S.; Buchner, H.; Rabenstein, R.; Herbordt, W. Active listening room compensation for massive multichannel sound reproduction systems. *J. Acoust. Soc. Am.* **2007**, *122*, 354–369.
127. Talagala, D.; Zhang, W.; Abhayapala, T.D. Efficient multichannel adaptive room compensation for spatial soundfield reproduction using a modal decomposition. *IEEE Trans. Audio Speech Lang. Process.* **2014**, *22*, 1522–1532.
128. Schneider, M.; Kellermann, W. Multichannel acoustic echo cancellation in the wave domain with increased robustness to nonuniqueness. *IEEE Trans. Audio Speech Lang. Process.* **2016**, *24*, 518–529.

129. Poletti, M.A.; Fazi, F.M.; Nelson, P.A. Sound-field reproduction systems using fixed-directivity loudspeakers. *J. Acoust. Soc. Am.* **2010**, *127*, 3590–3601.
130. Poletti, M.A.; Abhayapala, T.D.; Samarasinghe, P.N. Interior and exterior sound field control using two dimensional higher-order variable-directivity sources. *J. Acoust. Soc. Am.* **2012**, *131*, 3814–3823.
131. Betlehem, T.; Poletti, M.A. Two dimensional sound field reproduction using higher-order sources to exploit room reflections. *J. Acoust. Soc. Am.* **2014**, *135*, 1820–1833.
132. Canclini, A.; Markovic, D.; Antonacci, F.; Sarti, A.; Tubaro, S. A room-compensated virtual surround sound system exploiting early reflections in a reverberant room. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 1029–1033.
133. Samarasinghe, P.N.; Abhayapala, T.D.; Poletti, M.A. Room reflections assisted spatial sound field reproduction. In Proceedings of the European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014; pp. 1352–1356.
134. Betlehem, T.; Zhang, W.; Poletti, M.A.; Abhayapala, T.D. Personal sound zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Process. Mag.* **2015**, *32*, 81–91.
135. Choi, J.-W.; Kim, Y.-H. Generation of an acoustically bright zone with an illuminated region using multiple sources. *J. Acoust. Soc. Am.* **2002**, *111*, 1695–1700.
136. Shin, M.; Lee, S.Q.; Fazi, F.M.; Nelson, P.A.; Kim, D.; Wang, S.; Park, K.H.; Seo, J. Maximization of acoustic energy difference between two spaces. *J. Acoust. Soc. Am.* **2010**, *128*, 121–131.
137. Elliott, S.J.; Cheer, J.; Choi, J.-W.; Kim, Y.-H. Robustness and regularization of personal audio systems. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 2123–2133.
138. Chang, J.-H.; Lee, C.-H.; Park, J.-Y.; Kim, Y.-H. A realization of sound focused personal audio system using acoustic contrast control. *J. Acoust. Soc. Am.* **2009**, *125*, 2091–2097.
139. Okamoto, T.; Sakaguchi, A. Experimental validation of spatial Fourier transform-based multiple sound zone generation with a linear loudspeaker array. *J. Acoust. Soc. Am.* **2017**, *141*, 1769–1780.
140. Cheer, J.; Elliott, S.J.; Gálvez, M.F.S. Design and implementation of a car cabin personal audio system. *J. Audio Eng. Soc.* **2013**, *61*, 414–424.
141. Coleman, P.; Jackson, P.; Olik, M.; Pederson, J.A. Personal audio with a planar bright zone. *J. Acoust. Soc. Am.* **2014**, *136*, 1725–1735.
142. Coleman, P.; Jackson, P.; Olik, M.; M'øller, M.; Olsen, M.; Pederson, J.A. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array. *J. Acoust. Soc. Am.* **2014**, *135*, 1029–1940.
143. Poletti, M.A. An investigation of 2D multizone surround sound systems. In Proceedings of the 125th Audio Engineering Society Convention, San Francisco, CA, USA, 2–5 October 2008; 9p.
144. Betlehem, T.; Withers, C. Sound field reproduction with energy constraint on loudspeaker weights. *IEEE Trans. Audio Speech Lang. Process.* **2012**, *20*, 2388–2392.
145. Radmanesh, N.; Burnett, I.S. Generation of isolated wideband soundfield using a combined two-stage Lasso-LS algorithm. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 378–387.
146. Jin, W.; Kleijn, W.B. Theory and design of multizone soundfield reproduction using sparse methods. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 2343–2355.
147. Chang, J.-H.; Jacobsen, F. Sound field control with a circular double-layer array of loudspeakers. *J. Acoust. Soc. Am.* **2012**, *131*, 4518–4525.
148. Chang, J.-H.; Jacobsen, F. Experimental validation of sound field control with a circular double-layer array of loudspeakers. *J. Acoust. Soc. Am.* **2013**, *133*, 2046–2054.
149. Cai, Y.; Wu, M.; Yang, J. Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint. *J. Acoust. Soc. Am.* **2014**, *135*, 734–741.
150. Gálvez, S.; Marcos, F.; Elliott, S.J.; Jordan, C. Time domain optimisation of filters used in a loudspeaker array for personal audio. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1869–1878.
151. Wu, Y.J.; Abhayapala, T.D. Spatial multizone soundfield reproduction: Theory and design. *IEEE Trans. Audio Speech Lang. Process.* **2011**, *19*, 1711–1720.
152. Poletti, M.A.; Fazi, F.M. An approach to generating two zones of silence with application to personal sound systems. *J. Acoust. Soc. Am.* **2015**, *137*, 1711–1720.
153. Menzies, D. Sound field synthesis with distributed modal constraints. *Acta Acust. United Acust.* **2012**, *98*, 15–27.

154. Helwani, K.; Spors, S.; Buchner, H. The synthesis of sound figures. *Multidimens. Syst. Signal Process.* **2014**, *25*, 379–403.
155. Zhang, W.; Abhayapala, T.D.; Betlehem, T.; Fazi, F.M. Analysis and control of multi-zone sound field reproduction using modal-domain approach. *J. Acoust. Soc. Am.* **2016**, *140*, 2134–2144.