We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

7,000
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Reverberation and its Binaural Reproduction: The Trade-off between Computational Efficiency and Perceived Quality

Isaac Engel and Lorenzo Picinali

Abstract

Accurately rendering reverberation is critical to produce realistic binaural audio, particularly in augmented reality applications where virtual objects must blend in seamlessly with real ones. However, rigorously simulating sound waves interacting with the auralised space can be computationally costly, sometimes to the point of being unfeasible in real time applications on resource-limited mobile platforms. Luckily, knowledge of auditory perception can be leveraged to make computational savings without compromising quality. This chapter reviews different approaches and methods for rendering binaural reverberation efficiently, focusing specifically on Ambisonics-based techniques aimed at reducing the spatial resolution of late reverberation components. Potential future research directions in this area are also discussed.

Keywords: binaural audio, reverberation, Auralisation, Ambisonics, perceptual evaluation

1. Introduction

Reverberation results from pairing a sound source with an acoustic space. After emanating from the source, a sound wave will interact with its environment, undergoing reflection, diffraction and absorption. Thus, a listener will receive filtered replicas of the original wavefront (echoes) arriving from various directions at different times, causing the impression that the original sound persists in time. According to the so-called precedence effect, the direct sound allows a listener to determine the position of the sound source, while early reflections are generally not perceived as distinct auditory events [1–3]. As stated by Wallach et al. [1], the maximum delay after which a reflection is no longer 'fused' with the direct sound depends on the signal, being around 5 ms for single clicks and as long as 40 ms for complex signals such as speech or music [4]. Nevertheless, early reflections can broaden the perceived width of the source and shift its apparent position, as shown experimentally by Olive and Toole [5]. Furthermore, they can modify the signal's spectrum due to phase cancellation and subsequent comb filtering, as shown by Bech in his study on small-room acoustics [6]. Such phenomena can alter the perception of the room on a higher level. For example, Barron and Marshall [7]

argued that the timing, direction of arrival, and spectra of early lateral reflections contribute to the sense of 'envelopment'—defined as the 'subjective impression of being surrounded by the sound'. The time delay between the direct sound and the first distinct echo has also been shown to be a relevant feature: in the case of small rooms, Kaplanis et al. [8] found that it was correlated with the perception of environment dimensions and 'presence'—or 'sense of being inside an enclosed space and feeling its boundaries'—while in the case of concert halls, Beranek [9] linked it to a sense of 'intimacy'.

As time passes and the sound waves that emanated from the source continue interacting with the environment, the temporal density of echoes increases, and the resulting sound field becomes more diffuse. At this point, temporal and spatial features of individual echoes become less relevant, and late reverberation can be characterised as a stochastic process. An important parameter used to define such process is the reverberation time (RT), or the 'duration required for the space-averaged sound energy density in an enclosure to decrease by 60 dB after the source emission has stopped' [10], which is generally proportional to the volume of the room. Yadav et al. [11] suggested that RT contributes to the perception of environment dimensions most significantly in large spaces, whereas early reflections have greater importance in small rooms. Although late reverberation is often modelled as diffuse and isotropic (i.e., with an even distribution of energy across directions from the listeners' point of view). Alary et al. [12] showed that this assumption may not always hold and directionality should be taken into account, especially for asymmetrical spaces, such as a corridor.

When reproduced binaurally (e.g., through headphones), it has been shown that reverberation increases the sense of externalisation, i.e., the illusion of virtual sound sources being outside the head, when compared to anechoic sounds [13, 14]. It has been suggested that this effect can be achieved even by just adding the early reflections [13], while the contribution of late reverberation (>80 ms) is smaller in comparison [15]. Previous studies have looked into the contribution of both monaural and binaural cues to the externalisation of reverberant binaural signals. Monaural cues have been shown to have limited importance by Hassager et al. [16] and Jiang et al. [17], who argued that spectral detail is not as critical in the reverberant sound as it is in the direct sound. Regardless, it has been reported that applying spectral correction (headphone equalisation) to binaural signals could increase externalisation and other subjective attributes when employing headphones with limited reproduction bandwidth [18, 19]. Binaural cues, on the other hand, have been shown to be critical: Leclere et al. [14] suggested that reverberation increases externalisation of a binaural signal as long as interaural differences are introduced. This is supported by Catic et al. [15], who reported a considerable decrease in externalisation when the reverberant part of auralised speech was presented diotically. Such effects have been linked to specific binaural cues, such as interaural level differences (ILDs) and interaural coherence (IC). Recent studies have reported correlations between the level of externalisation and the amount of temporal fluctuations of ILDs and IC in the binaural signals [14, 15, 20]. Moreover, Li et al. [21, 22] highlighted the importance of reverberation specifically in the contralateral ear signal, showing a stronger contribution to externalisation than its ipsilateral counterpart, which is explained by the fact that reverberation is proportionally louder on the contralateral side due to the head shadow effect. Finally, according to the 'room divergence effect', externalisation of simulated binaural signals increases when the rendered reverberation matches the listener's expectations given their prior knowledge of the room [23–25]. Head movements and vision also play an important role in spatial audio perception [26], but they are not covered here—for a thorough review on sound externalisation, the reader is referred to Best et al. [27].

In summary, reverberation greatly influences how a listener perceives an auditory scene by providing information on the room characteristics, the size and location of the sound sources and, in the case of binaural simulations, affecting the level of externalisation. Consequently, it should be modelled carefully when producing realistic acoustic simulations, although this can prove to be a challenging task in real-time systems with limited resources. The next sections of this chapter will address the issue of balancing computational efficiency and perceptual quality when simulating reverberation.

2. Simulating reverberation efficiently

Simulating reverberation can be useful in various applications. In some cases, such as music production, it has mainly an aesthetic value and may not require highly realistic simulations. In other cases, such as architectural acoustics, augmented reality (AR) and, to a lesser extent, virtual reality (VR), the goal is to recreate a real acoustic space, so reverberation needs to be modelled with sufficient accuracy. For instance, an AR system allows the users to perceive the real world integrated with a virtual layer, e.g., a videoconferencing application in which users, wearing a pair of AR glasses, see holograms of their interlocutors which look and sound as if they were in the same room. From the acoustic point of view, this is particularly challenging to implement because the listener is exposed to real sound sources as well as virtual ones, so the simulated acoustics should be realistic enough for the virtual and real sources to be appropriately blended. Even though highly realistic reverberation is often desired, it can easily become too expensive to simulate in real time for interactive applications, where the auditory scene is expected to vary over time—even more so if many virtual sources are simulated [28]. Therefore, it is relevant to explore simplified reverberation models that reduce computational costs without compromising quality.

In the most general case, reverberation is rendered by convolving a dry audio signal with a room impulse response (RIR), which is the time-domain acoustic transfer function between a sound source and a receiver in a given acoustic space (room), assuming that the system formed by these is linear and time-invariant. The RIR can be either measured acoustically [29] or in a simulated environment. Several simulation techniques have been proposed, which range from rigorous but computationally expensive physical models, such as the finite-difference time-domain method [30], to simpler but less accurate geometrical models, such as the image source method [31] or scattering delay networks [32]. Ray-tracing and cone-tracing are also popular techniques that allow for a variable degree of accuracy [28, 33–35], albeit the computational requirements can become rather intensive when sound sources move in space, and real-time implementations are often limited to very simplified models and/or renderings.

Reverberation may also be generated through computationally lighter 'convolution-less' methods, such as Schroeder reverberators [36] or feedback delay networks (FDN) [37–39]. Such techniques are generally less accurate than convolution-based methods but can be useful to efficiently model the less critical parts of the RIR such as the late-reverberation tail [40].

With the goal of finding a balance between computational cost and perceived quality, several parametric reverberation models have been proposed [40–47]. Most of them aim to alleviate computational costs by rendering early reflections with a higher temporal and spatial accuracy than late reverberation, based on the concept of mixing time, i.e., the instant after which the RIR does not perceivably change across different listeners' positions or orientations within the room (see **Figure 1**) [48].

An early example of this approach, known as 'hybrid' reverberation, was presented by Murphy and Stewart [40], who proposed to employ convolution-based rendering for early reflections and simpler methods (e.g., FDN) to produce late reverberation. A key aspect of the hybrid model is correctly establishing the mixing time, which depends on the room volume, being higher for larger rooms [48].

In spatial audio applications, it is important to accurately simulate the direction of arrival of early reflections (and of late reverberation, to a lesser extent) which adds yet another layer of difficulty to the process. This also means that the reproduction method should be able to replicate such spatial cues. An example of a playback system would be a loudspeaker array surrounding the listener that can simulate virtual sources and reflections through amplitude panning [49] or Ambisonics [50]. In the case of binaural audio, such systems may be mimicked through virtual loudspeakers, but other methods also exist, as discussed in Section 2.1.

Note that the scope of this chapter covers reverberation's spatial features from the listener's point of view, but not from the source's point of view. Therefore, sound source directivity is not discussed, even though it is an important topic on its own—e.g., it is essential to model it correctly in a six-degrees-of-freedom (6DoF) application where the listener is allowed to walk past a directional source [51].

2.1 The binaural case

When rendering reverberation binaurally, directional information of reflected sounds is encoded in the binaural room impulse response (BRIR), i.e., a pair of RIRs that are measured at the listener's ear canals, in the form of monaural and interaural cues. Therefore, the most effective and straightforward way to achieve an accurate binaural rendering is to convolve an anechoic audio signal with a BRIR. Static (non-head-tracked) BRIR-based renderings can produce highly authentic binaural signals, to the point of being indistinguishable from those emitted by real sound sources [52–55]. On the other hand, dynamic (head-tracked) renderings are more challenging to implement, as they require swapping between BRIRs as the listener or the source move. It is worth noting that, when dealing with binaural renderings of anechoic environments, an angular movement of a source relative to the listener is roughly equivalent to a head rotation of the listener, which is typically trivial to compute in the Ambisonics domain using rotation matrices ([56], Section 5.2.2). However, this does not generalise to reverberant environments, where the room
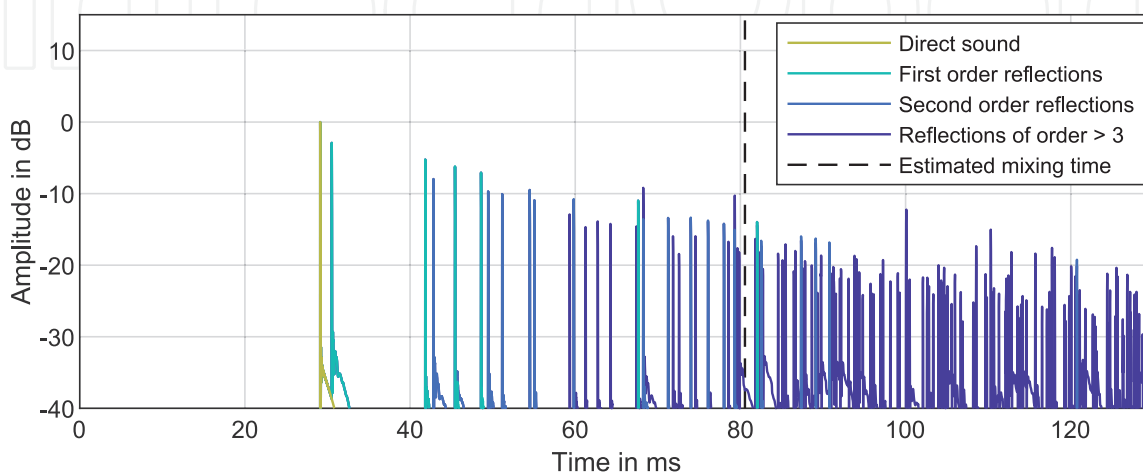
Figure 1.
First 130 ms of an RIR, expressed in decibels relative to the peak value. The RIR was simulated with the image source method [31] for an omnidirectional point source placed 10 m away from the receiver in a room with an approximate volume of 2342.7 m³. The mixing time, estimated according to Lindau et al. [48], is indicated.

provides a frame of reference, and the angular movement of a source is not equivalent to rotating the listener's head.

A recent study has suggested that BRIRs should be measured by varying the listener position in increments of 5 cm or less in a three-dimensional grid (which can be a costly process) to achieve a dynamic convolution-based rendering in which the swapping is seamless to the listener [57]. Alternatively, one may start from a coarser spatial grid and interpolate BRIRs at intermediate positions. Unfortunately, BRIR interpolation is not trivial because the time and direction of arrival of each reflection may vary depending on the receiver's position, changing the BRIR's temporal structure across the grid. Nevertheless, recent studies have shown promising progress by employing dual-band approaches and heuristics to match early reflections in the time domain [58, 59]. On a related note, another active research topic is the extrapolation of RIRs in the Ambisonics domain for 6DoF applications (e.g., [60–63]), which is further discussed in Section 4.

Although BRIRs are mainly obtained through binaural measurements made on a person's or a mannequin's head [55], they may also be generated from RIRs that were either measured with microphone arrays [64–68] or simulated [28, 35]. This approach typically involves identifying individual reflections and their direction of arrival, e.g., with the help of the spatial decomposition method (SDM) [65], and then convolving each reflection with a head-related impulse response (HRIR) for the corresponding direction [69]—which is equivalent to a multiplication with a head-related transfer function (HRTF) in the frequency domain. However, rendering the full length of the BRIR this way can easily become expensive, which is why simplified models such as the aforementioned 'hybrid' one become important: we can just render a few early reflections accurately while modelling late reverberation as a stochastic, non-directional process, and still produce binaural signals that are not perceptually different from properly rendered ones. This has been recently shown by Brinkmann et al. [47], who suggested that accurately rendering just six early reflections plus stochastic late reverberation may be enough to produce auralisations that are perceptually indistinguishable from a fully-rendered reference, for a simulation of a shoebox-type room.

It should be noted that modelling late reverberation as isotropic is computationally inexpensive but may lead to noticeable degradation when simulating asymmetrical rooms (e.g., a long and narrow corridor) where late reverberation is highly directional [12]. For such cases, Alary et al. have proposed to employ directional feedback delay networks (DFDN) [39], which extend the functionality of traditional FDNs to spatial audio and allow to inexpensively produce non-uniform reverberation, so that the RT is direction-dependent. A downside of DFDNs is their inability to correctly reproduce early reflections, which should be modelled separately for best results.

Another simplification consists in quantising the direction of arrival of reflections by 'snapping' them to the closest neighbour in a predefined grid. This method is explored by Amengual Garí et al. [69], who found that an RIR may be quantised to just 14 directions in a Lebedev grid [70] and still be used to render binaural signals through SDM without perceptual degradation when compared to the original. The scattering delay network method (SDN) is based on a similar premise, quantising the RIR to as many directions as first-order reflections, e.g., six for a cuboid room, while obtaining good results in perceptual evaluations [32]. The rationale of SDN is that early reflections are computed accurately, while later ones are approximated with higher error as time advances, which is a sensible approach from a perceptual point of view. However, it might lead to an inaccurate late reverberation tail, which is why combining SDN with an inexpensive method for late reverberation simulation (e.g., DFDN) might be a promising alternative.

On the other hand, rather than generating separate BRIRs for each rendered sound source, one may also 'encode' the sum of all of them into a single sound field, and then reproduce it binaurally, e.g., by means of a set of virtual loudspeakers. That way, only the virtual loudspeaker signals must be binaurally rendered, independently of the number of sources that form the sound field. This is a convenient simplification when many sources are rendered at once. As mentioned earlier, typical loudspeaker-based sound field reproduction methods include vector-based amplitude panning [49] and high-order Ambisonics [50, 56, 71]. The latter is by far the more popular method for binaural rendering, given its efficient simulation of head rotations ([56], Section 5.2.2) and manipulation of spatial resolution [72]. However, the Ambisonics processing may have perceivable effects on the binaural signals, which are still being investigated. Recent research on this topic is reviewed in Section 3.

3. Binaural Ambisonics-based reverberation and spatial resolution

The spherical harmonics framework (known as Ambisonics in the context of audio production) allows to express a sound field as a continuous function on a sphere around the listener. Ambisonics sound fields are typically generated from microphone array recordings [73] or plane-wave-based simulations. Alternatively, it is often convenient to measure or simulate an Ambisonics RIR that can be convolved with any anechoic audio signal to generate the sound field, e.g., as in [74]. Once encoded in the Ambisonics domain, a sound field can be mirrored, warped or rotated around the listener through inexpensive algebraic operations [56]. Additionally, it is possible to modify its spatial resolution, which allows to reduce computational costs in the rendering process in exchange for potential perceptual degradation [72, 75, 76].

When a sound field is encoded in the Ambisonics domain, its spatial resolution is defined by its inherent 'truncation order', which is an integer equal or greater than zero. Higher-order signals have a larger number of channels and allow to produce binaural renderings with finer spatial resolution and sound sources that are easier to localise, while lower-order signals are more lightweight (fewer channels) and produce renderings with lower resolution and 'blurry' sources (see **Figure 2**). This was shown by Avni et al. [77], who argued that truncating the order of an Ambisonics signal affected the perception of spaciousness and timbre in the resulting binaural signals. Later, Bernschütz [66] reported that, in perceptual evaluations, listeners could not generally detect differences in binaural signals rendered from Ambisonics sound fields of order 11 and above. Then, Ahrens and Andersson [74] showed that an order of 8 might be sufficient to simulate lateral sound sources that are indistinguishable from BRIR-based renderings, but slight differences were perceived up to order 29 for frontal sound sources.

It has also been shown that the relation between spatial order and perceived quality also depends on the 'decoding' method that is used to translate the Ambisonics sound field to a pair of binaural signals. For instance, the time-alignment method [78] and the magnitude least squares (MagLS) method [79] have both been shown to produce more accurate binaural signals at lower spatial orders than other approaches, such as the widely used virtual loudspeakers method [80]. In the case of MagLS, which focuses on minimising magnitude errors (disregarding phase) at high frequencies, Sun [81] showed that a conceptually similar method was able to produce binaural signals that were indistinguishable from a high-order reference at orders as low as 14.

Overall, previous studies have suggested that binaural signals can be accurately rendered from Ambisonics sound fields as long as the truncation order is high
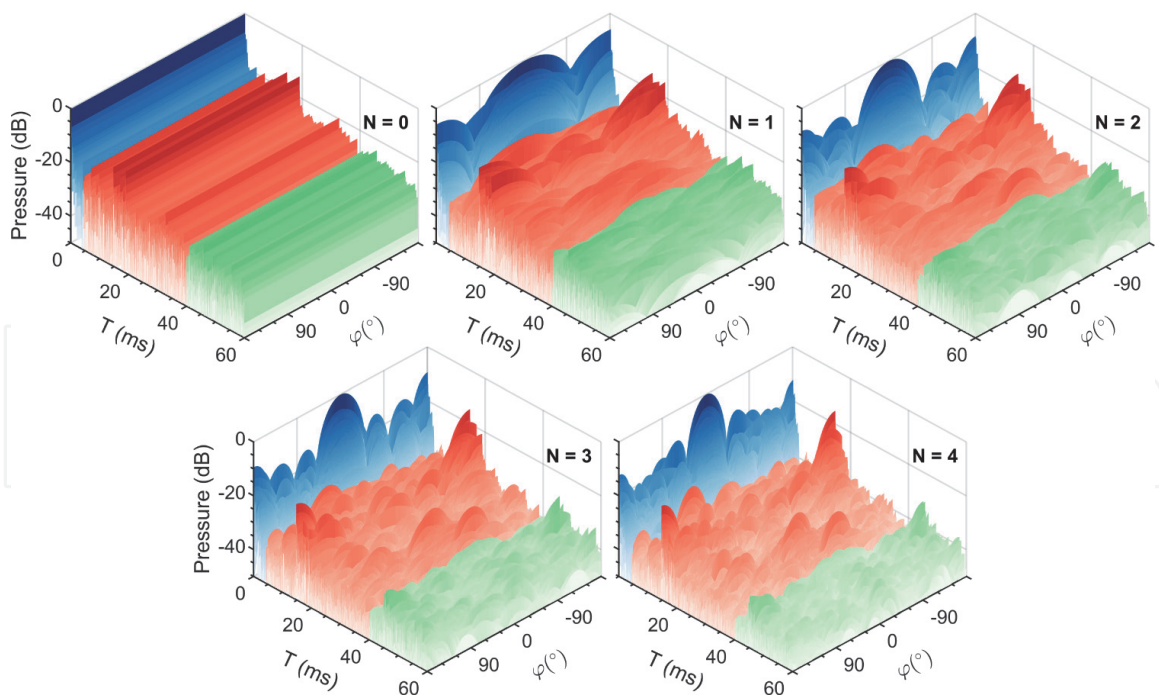
Figure 2.
Room impulse response encoded in the Ambisonics domain at different truncation orders (0 to 4), for a source placed in front of the listener. Data are plotted as sound pressure (in decibels relative to the peak value) along the time axis and over different azimuth angles on the horizontal plane. Source: Engel et al. [76] ('trapezoid' room).

enough, probably somewhere between 8 and 29. However, such orders may still be too high to be computationally efficient (the number of channels of an Ambisonics signal is proportional to the square of its truncation order) or just unfeasible in practice (commercially available microphone arrays operate at order 4 or lower). The remainder of this section discusses some recent perceptual studies that explored how the binaural rendering of reverberant sound fields is affected when simplifications are applied in the Ambisonics domain, e.g., reducing the truncation order of different parts of the RIR.

3.1 Hybrid Ambisonics

A recent listening experiment by Lübeck et al. [75] showed that early reflections and late reverberation may be encoded in Ambisonics at a significantly lower order than the direct sound and still produce binaural signals that are indistinguishable from a BRIR-based rendering. The reason why this may happen is illustrated in **Figure 2**, which shows an RIR encoded in Ambisonics at different truncation orders. It can be seen how the lowest order (0) produces an isotropic signal which does not vary across directions in the horizontal plane, while higher orders achieve a more faithful representation of the sound field by allowing for spatially 'sharper' patterns— e.g., note how the direct sound becomes narrower as order increases, converging towards a spatial Dirac delta. Looking at this figure, it becomes apparent that earlier parts of the RIR (blue) are more sensible to spatial resolution changes due to order truncation, compared to late reverberation (green) which is less directional.

According to these observations, it is reasonable to propose an Ambisonics-based binaural rendering method that employs a high truncation order for the direct sound (and, possibly, some early reflections) and lower orders for the rest of the RIR. Such a method could be highly efficient given that late reverberation usually accounts for the majority of the duration of the RIR. This approach, reminiscent to

the hybrid models discussed earlier, has been tentatively coined as 'hybrid Ambisonics'.

A perceptual study by Engel et al. [76] evaluated binaural signals generated with hybrid Ambisonics and the virtual loudspeaker method, and found that an order between 2 or 3 (dependent on the room) may be enough to render reverberation, assuming that the direct sound path is accurately reproduced through convolution with HRIRs (see **Figure 3**). This is a promising precedent for future efficient binaural rendering methods, although further investigations would be needed to generalise these results to a wider selection of rooms and stimuli types. In the future, a more general model could estimate the needed truncation order adaptively based on the Ambisonics signal (e.g., measuring its directivity over time), which could be used in efficient binaural renderers or as a way to compress spatial audio data.

3.2 Reverberant virtual loudspeaker (RVL)

In real-time interactive binaural simulations, RIRs are typically recomputed when there is a change in the scene such as movements of the listener or sources. When working in the Ambisonics domain, this recomputation is not needed in order to simulate a head rotation from the listener, as the signal can be efficiently rotated via a rotation matrix ([56], Section 5.2.2). However, translational movements of either the listener or a source still require to recompute the RIRs. As a result, the number of sources that can be rendered simultaneously in a low-cost scenario might be limited.

In such cases, it may be beneficial to employ a rendering method that scales well with the number of sources. One such example is the reverberant virtual loudspeaker method (RVL), an Ambisonics-based approach that has the advantage of requiring a fixed amount of real-time convolutions regardless of the number of sources [72, 76, 83]. This method takes inspiration from the virtual loudspeakers approach [71, 80], which decodes an Ambisonics sound field to a virtual loudspeaker grid around the listener and convolves the resulting signals with HRIRs to generate the binaural output. RVL performs this same process but, instead of HRIRs, the virtual loudspeaker signals are convolved with BRIRs, so the acoustics of the room are effectively integrated with the binaural rendering without the need for additional steps. Therefore, the number of real-time convolutions depends only on the truncation order of the sound field, independently of the number of rendered
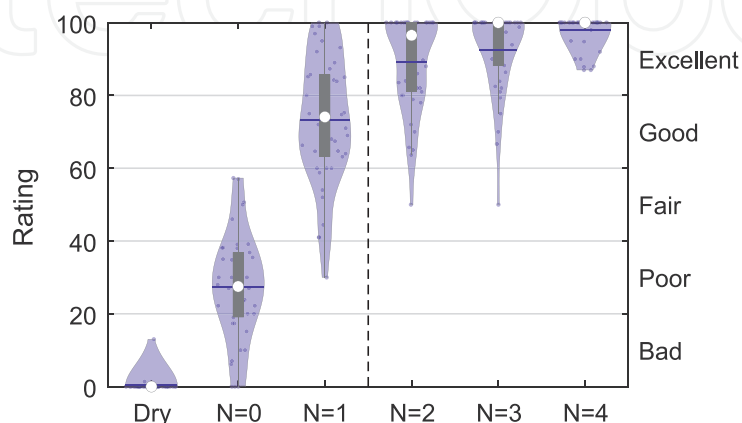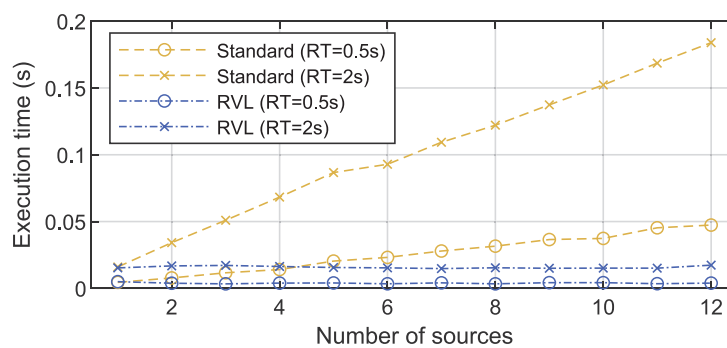

Figure 3.
*Perceptual ratings of binaural renderings generated from the hybrid-Ambisonics RIRs of orders 0 to 4 are shown in **Figure 2**, where the direct sound was reproduced via convolution with a single HRIR. A dry rendering was used as the anchor signal and the 4th order signal, as the reference. The vertical dotted lines indicate that the groups on the left are significantly different ($p < 0.05$) from the groups on the right. Source: Engel et al. [76].*

sources. For this reason, RVL is highly efficient at rendering a large number of sources in real time (see **Figure 4**). Its main limitation is that the room is head-locked due to the set of BRIRs being fixed, so head rotations may lead to inaccurate reflections, as shown in **Figure 5**.

RVL was perceptually evaluated in [76], paying particular attention to its effect on head rotations. For the assessment, the method was applied only to the reverberant sound (direct sound was generated through convolution with HRIRs) and the implementation was done with the 3D Tune-In Toolkit spatial audio library [84]. Listeners were asked to compare RVL to first-order hybrid Ambisonics renderings (both head-tracked) of speech and music, by being asked 'Considering the given room [shown in a picture], which example is more appropriate?'. Results suggested that the inaccurate head rotations could indeed be detected by listeners but were not necessarily perceived as a degradation in quality with respect to the more accurate rendering—note the bimodal distribution shown in **Figure 6**, which indicates that there was not a unanimous preference towards either rendering.

One could speculate that the RVL method was preferred by some listeners due to the BRIR-based rendering leading to highly uncorrelated binaural signals, which are typically associated with higher perceived quality when evaluating late reverberation (see the binaural quality index by Beranek [9]). An additional investigation to explore the matter further would be to compare the RVL method to other approaches that specifically aim to optimise interaural coherence, such as the
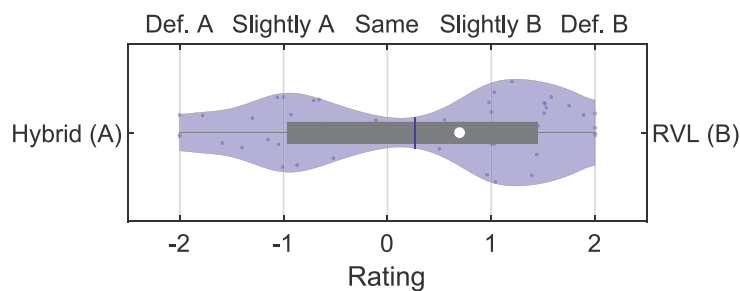

Figure 4.
Comparison between the average execution time of the convolution stage in Ambisonics binaural rendering ('standard') and RVL binaural rendering, as a function of the number of rendered sources, for two different reverberation times (RT). A random input signal with a length of 1024 samples was used as input. Simulations were done in MATLAB (MathWorks) using the overlap-add method [82], running on a quad-core processor at 2.8 GHz. Source: Engel et al. [76].
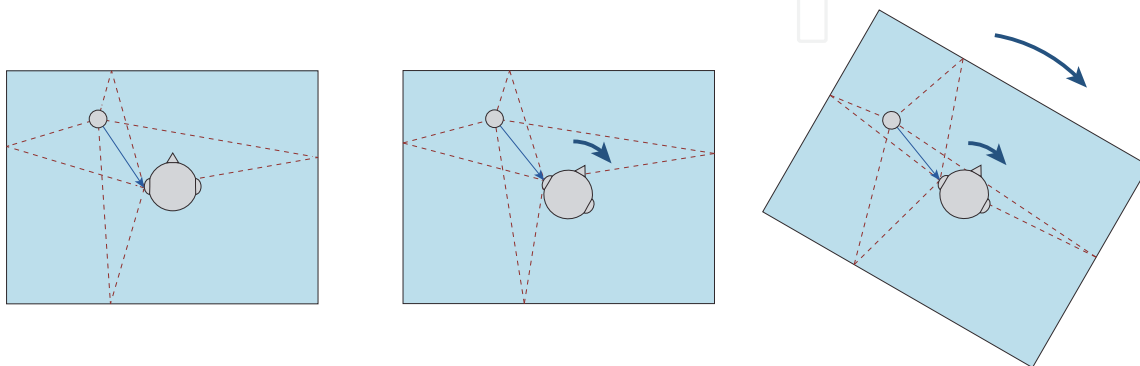
Figure 5.
Direct sound path and first-order early reflections as they reach the left ear of a listener in three scenarios: (left) before any head rotation; (middle) canonical rendering after a head rotation of 30 degrees clockwise; and (right) RVL rendering after the same head rotation. Note how, in the third scenario, the direct sound path is accurate, whereas the room is head-locked, affecting the incoming direction of reflections. Source: Engel et al. [76].

Figure 6.
Violin plot showing perceptual ratings from paired comparisons between first-order hybrid Ambisonics (A) and RVL (B) binaural renderings. Negative values represent preference towards a, while positive values represent preference towards B. Source: Engel et al. [76].

covariance constraint method proposed by Zaunschirm et al. [78] and described by Zotter and Frank ([56], Section 4.11.3).

Regardless, further perceptual evaluations (e.g., in more rooms) would be needed to generalise these results. Overall, RVL could be a viable option to render binaural reverberation of a large number of sources in real time in a low-resource scenario.

4. Future directions

The trade-off between complexity and perceived quality when rendering binaural reverberation is still an area of major interest that has to be further explored. Recent studies have looked at the perceptual impact of varying spatial resolution of Ambisonics-based reverberation, but there are yet aspects of it that warrant further research. For instance, it would be interesting to explore an approach to compress Ambisonics RIRs by truncating their order depending on their directional and temporal information, as a way to compute and store them more efficiently.

Another set of very relevant challenges will come from using artificial binaural reverberation in different contexts and tasks. For example, binaural audio has been used in the past for assisting blind individuals in learning about the spatial configuration of a closed environment before being physically introduced to it [85, 86]. Within that context, the creation of geometrically and spatially accurate real-time reverberation was extremely important and could be achieved only by performing a series of case-specific optimisations in the processing chain, for example, limiting navigation paths to a series of lines rather than a 2-dimensional space, and pre-calculating a set of Ambisonics RIRs computing in real-time only rotations and interpolations. Such optimisations can be allowed only within a research environment, therefore real-life applications of such techniques are currently very limited. A better understanding of both the computational and perceptual sides of reverberation, possibly specifically for blind and visually impaired individuals, could lead to major advancements in the development and use of auditory displays and assistive technologies, tools and devices.

Looking ahead, AR applications could offer an interesting testbed for further research on binaural reverberation perception and rendering. One of the key research areas in AR/VR is 6DoF (or position-dynamic) audio rendering, where the listener is allowed to move around the scene, as opposed to traditional Ambisonics rendering where only head rotations are allowed (three degrees of freedom). Several methods have been recently proposed to efficiently extrapolate spatial audio signals from one listener position to another, either via simple parametric methods

[87] or more complex Ambisonics-based approaches that often rely on parametrising the sound field in 'direct' and 'ambient' components [60, 61], or according to the source distance [62, 63]. Significant advancements have also been made in terms of recording complex auditory scenes and to make them navigable in 6DoF—in this case, specialised hardware and software has been released and is already available commercially [88]. Future improvements in 6DoF recording and rendering techniques will in turn allow for an increased level of interactivity within the simulation, as well as more effective evaluations of different audio rendering technologies using AR/VR systems.

Focusing on the AR case, in order to blend real with virtual audio, it is essential to develop techniques for the automatic estimation of the reverberant characteristics of the real environment. New methods will need to be developed and evaluated for blending virtual audio sources within real scenes and to evaluate the impact of blending accuracy through metrics related to perceived realism and scene acceptability. This can be achieved, for example, by characterising the acoustical environment surrounding the AR user, using this in-situ data to synthesise virtual sounds with matching acoustic properties. Machine learning (ML) techniques could be employed to address the issue of blind acoustical environment characterisation by focusing first on overall room fingerprint evaluation (late reverberation), then on the finer details of the room response that vary depending on specific source positions (early reflections). The scene analysis could also be used to extract the direction-of-arrival for multiple sound sources and direct-to-reverberant energy ratio by separating source information from room and user acoustic properties. The data extracted by the model could then be employed to generate realistic virtual reverberation, which will be matched with the real-world reverberation. Of course for each step of this scenario several open challenges still exist, both from the computational point of view (e.g., how to generate geometrically and directionally accurate reverberation in real-time) and from the perceptual point of view (e.g., what is perceptually relevant and should therefore be computationally modelled and rendered, and what can be approximated).

Better understanding the extent and origin of sensory thresholds in terms of reverberation perception, therefore, presents still a very open set of challenges, which will need to be addressed in the future through extensive listening experiments and, why not, also by means of binaural auditory models and ML-trained 'artificial listeners'.

5. Conclusions

Within this chapter, an overview was presented on perception and efficient simulation of reverberation. A special focus has been put on the case of binaural audio and, in particular, on Ambisonics-based and convolution-based rendering methods. The issue of the trade-off between computational cost and perceived quality has been discussed at length, mainly looking at the case of varying spatial resolution and implementation choices of Ambisonics-based renderings, highlighting the results of some recent studies on this matter. Considering the very rapid development and uptake of VR and AR technologies, it is particularly evident the importance of further research focusing on better understanding how computational optimisations and simplifications can have an impact on the perceived quality and realism of the rendering. Some of the most relevant challenges in this area have been outlined at the end of the chapter, and will hopefully serve as a guideline for future research in the area.

Acknowledgements

Nomenclature

6DoF	Six degrees of freedom
AR	Augmented reality
BRIR	Binaural room impulse response
DFDN	Directional feedback delay network
FDN	Feedback delay network
HRIR	Head-related impulse response
HRTF	Head-related transfer function
IC	Interaural coherence
ILD	Interaural level difference
ML	Machine learning
RIR	Room impulse response
RT	Reverberation time
RVL	Reverberant virtual loudspeaker
SDM	Spatial decomposition method
SDN	Scattering delay network
VR	Virtual reality

Author details

Isaac Engel* and Lorenzo Picinali
Dyson School of Design Engineering, Imperial College London, London,
United Kingdom

*Address all correspondence to: isaac.engel@imperial.ac.uk

IntechOpen

References

[1] Wallach H, Newman EB, Rosenzweig MR. A precedence effect in sound localization. The Journal of the Acoustical Society of America. 1949; **21**(4):468-468. DOI: 10.1121/1.1917119

[2] Litovsky RY, Steven Colburn H, Yost WA, Guzman SJ. The precedence effect. The Journal of the Acoustical Society of America. 1999;**106**(4): 1633-1654. DOI: 10.1121/1.427914

[3] Brown AD, Christopher Stecker G, Tollin DJ. The precedence effect in sound localization. JARO: Journal of the Association for Research in Otolaryngology. 2015;**16**(1):1-28. DOI: 10.1007/s10162-014-0496-2

[4] Moore BCJ. An Introduction to the Psychology of Hearing. Leiden, Netherlands: Brill; 2012

[5] Olive SE, Toole FE. The detection of reflections in typical rooms. Journal of the Audio Engineering Society. 1989;**37** (7/8):539-553. https://aes.org/e-lib/browse.cfm?elib=6079

[6] Bech S. Timbral aspects of reproduced sound in small rooms. II. The Journal of the Acoustical Society of America. 1996;**99**(6):3539-3549. DOI: 10.1121/1.414952

[7] Barron M, Marshall AH. Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. Journal of Sound and Vibration. 1981; **77**(2):211-232. DOI: 10.1016/ S0022-460X(81)80020-X

[8] Kaplanis N, Bech S, Jensen SH, van Waterschoot T. Perception of reverberation in small rooms: A literature study. In: Audio Engineering Society Conference: 55th International Conference: Spatial Audio. Helsinki, Finland: Audio Engineering Society; 2014.

https://aes.org/e-lib/browse.cfm?elib= 17348

[9] Beranek LL. Concert hall acoustics— 2008. Journal of the Audio Engineering Society. 2008;**56**(7/8):532-544. https:// aes.org/e-lib/browse.cfm?elib=14398

[10] International Organization for Standardization. Measurement of Room Acoustic Parameters — Part 1: Performance Spaces (ISO Standard No. 3382-1:2009). 2009. https://www.iso. org/standard/40979.html

[11] Yadav M, Cabrera DA, Miranda L, Martens WL, Lee D, Collins R. Investigating auditory room size perception with autophonic stimuli. In: Audio Engineering Society Convention 135. New York, USA: Audio Engineering Society; 2013. https://aes. org/e-lib/browse.cfm?elib=16984

[12] Alary B, Massé P, Välimäki V, Noisternig M. Assessing the anisotropic features of spatial impulse responses. In: EAA Spatial Audio Signal Processing Symposium. Paris, France: Sorbonne Université; 2019. pp. 43-48. DOI: 10.25836/sasp.2019.32

[13] Begault DR, Wenzel EM, Anderson MR. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. Journal of the Audio Engineering Society. 2001; **49**(10):904-916. http://aes.org/e-lib/b rowse.cfm?elib=10175

[14] Leclère T, Lavandier M, Perrin F. On the externalization of sound sources with headphones without reference to a real source. The Journal of the Acoustical Society of America. 2019;**146**(4): 2309-2320. DOI: 10.1121/1.5128325

[15] Catic J, Santurette S, Dau T. The role of reverberation-related binaural cues in

the externalization of speech. The Journal of the Acoustical Society of America. 2015;**138**(2):1154-1167. DOI: 10.1121/1.4928132

[16] Hassager HG, Gran F, Dau T. The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment. The Journal of the Acoustical Society of America. 2016;**139**(5):2992-3000. DOI: 10.1121/1.4950847

[17] Jiang Z, Sang J, Zheng C, Li X. The effect of pinna filtering in binaural transfer functions on externalization in a reverberant environment. Applied Acoustics. 2020;**164**:107257. DOI: 10.1016/j.apacoust.2020.107257

[18] Engel I, Alon DL, Robinson PW, Mehra R. The effect of generic headphone compensation on binaural renderings. In: Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio. York, UK: Audio Engineering Society; 2019. https://www.aes.org/e-lib/browse.cfm?elib=20387

[19] Engel I, Alon DL, Scheumann K, Mehra R. Listener-preferred headphone frequency response for stereo and spatial audio content. In: Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality. Virtual Reality: Audio Engineering Society; 2020. https://aes.org/e-lib/browse.cfm?elib=20868

[20] Catic J, Santurette S, Buchholz JM, Gran F, Dau T. The effect of interaural-level-difference fluctuations on the externalization of sound. The Journal of the Acoustical Society of America. 2013;**134**(2):1232-1241. DOI: 10.1121/1.4812264

[21] Li S, Schlieper R, Peissig J. The effect of variation of reverberation parameters in contralateral versus ipsilateral ear signals on perceived externalization of a lateral sound source in a listening room. The Journal of the Acoustical Society of America. 2018;**144**(2):966-980. DOI: 10.1121/1.5051632

[22] Li S, Schlieper R, Peissig J. The role of reverberation and magnitude spectra of direct parts in contralateral and ipsilateral ear signals on perceived externalization. Applied Sciences. 2019;**9**(3):460. DOI: 10.3390/app9030460

[23] Werner S, Klein F, Mayenfels T, Brandenburg K. A summary on acoustic room divergence and its effect on externalization of auditory events. In: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). Lisbon, Portugal: IEEE; 2016. pp. 1-6. DOI: 10.1109/QoMEX.2016.7498973

[24] Werner S, Götz G, Klein F. Influence of head tracking on the externalization of auditory events at divergence between synthesized and listening room using a binaural headphone system. In: Audio Engineering Society Convention 142. Berlin, Germany: Audio Engineering Society; 2017. https://www.aes.org/e-lib/browse.cfm?elib=18568

[25] Klein F, Werner S, Mayenfels T. Influences of training on externalization of binaural synthesis in situations of room divergence. Journal of the Audio Engineering Society. 2017;**65**(3):178-187. https://www.aes.org/e-lib/browse.cfm?elib=18553

[26] Engel I, Goodman DFM, Picinali L. The effect of auditory anchors on sound localization: A preliminary study. In: Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio. York, UK: Audio Engineering Society; 2019. https://aes.org/e-lib/browse.cfm?elib=20388

[27] Best V, Baumgartner R, Lavandier M, Majdak P, Kopčo N. Sound externalization: A review of recent research. Trends in Hearing. 2020;**24**:2331216520948390. DOI: 10.1177/2331216520948390

[28] Schissler C, Stirling P, Mehra R. Efficient construction of the spatial room impulse response. In: 2017 IEEE Virtual Reality (VR). Los Angeles, USA: IEEE; 2017. pp. 122-130. DOI: 10.1109/VR.2017.7892239

[29] Farina A. Advancements in impulse response measurements by sine sweeps. In: Audio Engineering Society Convention 122. Vienna, Austria: Audio Engineering Society; 2007. https://aes.org/e-lib/browse.cfm?elib=14106

[30] Botteldooren D. Finite-difference time-domain simulation of low-frequency room acoustic problems. The Journal of the Acoustical Society of America. 1995;**98**(6):3302-3308. DOI: 10.1121/1.413817

[31] Allen JB, Berkley DA. Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America. 1979;**65**(4):943-950. DOI: 10.1121/1.382599

[32] De Sena E, Hacıhabiboğlu H, Cvetković Z, Smith JO. Efficient synthesis of room acoustics via scattering delay networks. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015;**23**(9): 1478-1492. DOI: 10.1109/TASLP.2015.2438547

[33] Vorländer M. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. The Journal of the Acoustical Society of America. 1989;**86**(1):172-178. DOI: 10.1121/1.398336

[34] Lentz T, Schröder D, Vorländer M, Assenmacher I. Virtual reality system with integrated sound field simulation and reproduction. EURASIP Journal on Advances in Signal Processing. 2007; **2007**(1):187. DOI: 10.1155/2007/70540

[35] Schissler C, Mehra R, Manocha D. High-order diffraction and diffuse reflections for interactive sound propagation in large environments. ACM Transactions on Graphics. 2014;**33** (4):1-12. DOI: 10.1145/2601097.2601216

[36] Schroeder MR, Logan BF. "Colorless" artificial reverberation. IRE Transactions on Audio, AU. 1961;**9**(6):209-214. DOI: 10.1109/TAU.1961.1166351

[37] Jot J-M, Chaigne A. Digital delay networks for designing artificial reverberators. In: Audio Engineering Society Convention 90. Paris, France: Audio Engineering Society; 1991. https://aes.org/e-lib/browse.cfm?elib=5663

[38] Jot J-M. Efficient models for reverberation and distance rendering in computer music and virtual audio reality. In: ICMC: International Computer Music Conference. Ann Arbor, Michigan, USA: Thessaloniki, Greece; 1997. pp. 236-243. https://hal.archives-ouvertes.fr/hal-01106168

[39] Alary B, Politis A, Schlecht S, Välimäki V. Directional feedback delay network. Journal of the Audio Engineering Society. 2019;**67**(10): 752-762. http://www.aes.org/e-lib/browse.cfm?elib=20693

[40] Murphy DT, Stewart R. A hybrid artificial reverberation algorithm. In: Audio Engineering Society Convention 122. Vienna, Austria: Audio Engineering Society; 2007. https://aes.org/e-lib/browse.cfm?elib=14006

[41] Carpentier T, Noisternig M, Warusfel O. Hybrid reverberation processor with perceptual control. In: 17th International Conference on Digital Audio Effects - DAFx-14. Erlangen,

Germany: International Audio Laboratories Erlangen; 2014. pp. 93-100. https://hal.archives-ouvertes.fr/hal-01107075

[42] Coleman P, Franck A, Jackson PJB, Hughes RJ, Remaggi L, Melchior F. Object-based reverberation for spatial audio. Journal of the Audio Engineering Society. 2017;**65**(1/2):66-77. https://www.aes.org/e-lib/browse.cfm?elib=18544

[43] Coleman P, Franck A, Menzies D, Jackson PJB. Object-based reverberation encoding from first-order ambisonic RIRs. In: Audio Engineering Society Convention 142. Boston, Massachusetts, USA: Audio Engineering Society; 2017. https://www.aes.org/e-lib/browse.cfm?elib=18608

[44] Stade P, Arend JM, Pörschmann C. A parametric model for the synthesis of binaural room impulse responses. In: Proceedings of Meetings on Acoustics. Vol. 30. Boston, Massachusetts, USA: Acoustical Society of America; 2017. p. 015006. DOI: 10.1121/2.0000573

[45] Raghuvanshi N, Snyder J. Parametric directional coding for precomputed sound propagation. ACM Transactions on Graphics. 2018;**37**(4): 1-14. DOI: 10.1145/3197517.3201339

[46] Godin K, Gamper H, Raghuvanshi N. Aesthetic modification of room impulse responses for interactive auralization. In: Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio. York, UK: Audio Engineering Society; 2019. https://www.aes.org/e-lib/browse.cfm?elib=20444

[47] Brinkmann F, Gamper H, Raghuvanshi N, Tashev I. Towards encoding perceptually salient early reflections for parametric spatial audio rendering. In: Audio Engineering Society Convention 148. Audio

Engineering Society; 2020. https://aes.org/e-lib/browse.cfm?elib=20797

[48] Lindau A, Kosanke L, Weinzierl S. Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses. Journal of the Audio Engineering Society. 2012;**60**(11): 887-898. https://aes.org/e-lib/browse.cfm?elib=16633

[49] Pulkki V. Virtual sound source positioning using vector base amplitude panning. Journal of the Audio Engineering Society. 1997;**45**(6): 456-466. https://www.aes.org/e-lib/browse.cfm?elib=7853

[50] Gerzon MA. Periphony: With-height sound reproduction. Journal of the Audio Engineering Society. 1973; **21**(1):2-10. https://www.aes.org/e-lib/browse.cfm?elib=2012

[51] Werner S, Klein F, Neidhardt A, Sloma U, Schneiderwind C, Brandenburg K. Creation of auditory augmented reality using a position-dynamic binaural synthesis system— Technical components, psychoacoustic needs, and perceptual evaluation. Applied Sciences. 2021;**11**(3):1150. DOI: 10.3390/app11031150

[52] Langendijk EHA, Bronkhorst AW. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. The Journal of the Acoustical Society of America. 1999;**107**(1): 528-537. DOI: 10.1121/1.428321

[53] Moore AH, Tew AI, Nicol R. An initial validation of individualized crosstalk cancellation filters for binaural perceptual experiments. AES: Journal of the Audio Engineering Society. 2010;**58**(1–2):36-45

[54] Masiero B, Fels J. Perceptually robust headphone equalization for binaural reproduction. In: Audio Engineering Society Convention 130.

London, UK: Audio Engineering Society; 2011. pp. 1-7. DOI: 10.13140/2.1.1598.6882

[55] Brinkmann F, Lindau A, Weinzierl S. On the authenticity of individual dynamic binaural synthesis. The Journal of the Acoustical Society of America. 2017;**142**(4):1784-1795. DOI: 10.1121/1.5005606

[56] Zotter F, Frank M. Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality. Basingstoke, UK: Springer Nature; 2019. DOI: 10.1007/978-3-030-17207-7

[57] Neidhardt A, Reif B. Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis. In: Audio Engineering Society Convention 148. Audio Engineering Society; 2020. https://www.aes.org/e-lib/browse.cfm?elib=20788

[58] Zaunschirm M, Zotter F, Frank M. Perceptual evaluation of variable-orientation binaural room impulse response rendering. In: Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio. York, UK: Audio Engineering Society; 2019. https://www.aes.org/e-lib/browse.cfm?elib=20395

[59] Bruschi V, Nobili S, Cecchi S, Piazza F. An innovative method for binaural room impulse responses interpolation. In: Audio Engineering Society Convention 148. Audio Engineering Society; 2020. https://www.aes.org/e-lib/browse.cfm?elib=20802

[60] Kentgens M, Behler A, Jax P. Translation of a higher order ambisonics sound scene based on parametric decomposition. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE; 2020.

pp. 151-155. DOI: 10.1109/ICASSP40776.2020.9054414

[61] Müller K, Zotter F. Auralization based on multi-perspective ambisonic room impulse responses. Acta Acustica. 2020;**4**(6):25. DOI: 10.1051/aacus/2020024

[62] Kentgens M, Jax P. Translation of a higher-order ambisonics sound scene by space warping. In: Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality. Virtual Reality: Audio Engineering Society; 2020. https://www.aes.org/e-lib/browse.cfm?elib=20873

[63] Birnie L, Abhayapala T, Tourbabin V, Samarasinghe P. Mixed source sound field translation for virtual binaural application with perceptual validation. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2021;**29**:1188-1203. DOI: 10.1109/TASLP.2021.3061939

[64] Merimaa J, Pulkki V. Spatial impulse response rendering I: Analysis and synthesis. Journal of the Audio Engineering Society. 2005;**53**(12):1115-1127. https://aes.org/e-lib/browse.cfm?elib=13401

[65] Tervo S, Pätynen J, Kuusinen A, Lokki T. Spatial decomposition method for room impulse responses. Journal of the Audio Engineering Society. 2013;**61**(1/2):17-28. http://www.aes.org/e-lib/browse.cfm?elib=16664

[66] Bernschütz B. Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording. [Doctoral thesis]. Berlin, Germany: Technische Universität Berlin; 2016. DOI: 10.14279/depositonce-5082

[67] Zaunschirm M, Frank M, Zotter F. BRIR synthesis using first-order

microphone arrays. In: Audio Engineering Society Convention 144. Milan, Italy: Audio Engineering Society; 2018. https://www.aes.org/e-lib/browse.cfm?elib=19461

[68] Amengual Garí SV, Brimijoin WO, Hassager HG, Robinson PW. Flexible binaural resynthesis of room impulse responses for augmented reality research. In: EAA Spatial Audio Signal Processing Symposium. Paris, France: Sorbonne Université; 2019. pp. 161-166. DOI: 10.25836/sasp.2019.31

[69] Amengual Garí SV, Arend JM, Calamia PT, Robinson PW. Optimizations of the spatial decomposition method for binaural reproduction. Journal of the Audio Engineering Society. 2021;**68**(12): 959-976. https://www.aes.org/e-lib/browse.cfm?elib=21010

[70] Vyacheslav Ivanovich Lebedev. Spherical quadrature formulas exact to orders 25–29. Siberian Mathematical Journal, 18(1):99–107, 1977. https://https://link.springer.com/content/pdf/10.1007/BF00966954.pdf.

[71] Noisternig M, Musil T, Sontacchi A, Holdrich R. 3D binaural sound reproduction using a virtual ambisonic approach. In: IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems. Lugano, Switzerland: IEEE; 2003. VECIMS '03, 2003. pp. 174-178. DOI: 10.1109/VECIMS.2003.1227050

[72] Engel I, Henry C, Amengual Garí SV, Robinson PW, Poirier-Quinot D, Picinali L. Perceptual comparison of ambisonics-based reverberation methods in binaural listening. In: EAA Spatial Audio Signal Processing Symposium. Paris, France: Sorbonne Université; 2019. pp. 121-126. DOI: 10.25836/sasp.2019.11

[73] Rafaely B. Analysis and design of spherical microphone arrays. IEEE Transactions on Speech and Audio Processing. 2005;**13**(1):135-143. DOI: 10.1109/TSA.2004.839244

[74] Ahrens J, Andersson C. Perceptual evaluation of headphone auralization of rooms captured with spherical microphone arrays with respect to spaciousness and timbre. The Journal of the Acoustical Society of America. 2019; **145**(4):2783-2794. DOI: 10.1121/1.5096164

[75] Lübeck T, Pörschmann C, Arend JM. Perception of direct sound, early reflections, and reverberation in auralizations of sparsely measured binaural room impulse responses. In: Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality. Virtual Reality: Audio Engineering Society; 2020. https://aes.org/e-lib/browse.cfm?elib=20865

[76] Engel I, Henry C, Amengual Garí SV, Robinson PW, Picinali L. Perceptual implications of different Ambisonics-based methods for binaural reverberation. The Journal of the Acoustical Society of America. 2021;**149**(2):895-910. DOI: 10.1121/10.0003437

[77] Avni A, Ahrens J, Geier M, Spors S, Wierstorf H, Rafaely B. Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution. The Journal of the Acoustical Society of America. 2013;**133**(5):2711-2721. DOI: 10.1121/1.4795780

[78] Zaunschirm M, Schörkhuber C, Höldrich R. Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. The Journal of the Acoustical Society of America. 2018; **143**(6):3616-3627. DOI: 10.1121/1.5040489

[79] Schörkhuber C, Zaunschirm M, Höldrich R. Binaural rendering of ambisonic signals via magnitude least squares. In: Fortschritte Der Akustik–DAGA. Munich, Germany: Deutsche Gesellschaft für Akustik; 2018. pp. 339-342. https://researchgate.net/publication/325080691_Binaural_Rendering_of_Ambisonic_Signals_via_Magnitude_Least_Squares

[80] McKeag A, McGrath DS. Sound Field Format to Binaural Decoder with Head Tracking. In: Audio Engineering Society Convention 6r. Melbourne, Australia: Audio Engineering Society; 1996. https://aes.org/e-lib/browse.cfm?elib=7477

[81] Sun D. Generation and Perception of Three-Dimensional Sound Fields Using Higher Order Ambisonics. [Doctoral thesis]. Sydney, Australia: University of Sydney; 2012

[82] Oppenheim AV, Buck JR, Schafer RW. Discrete-Time Signal Processing. Vol. 2. Upper Saddle River, N.J: Prentice Hall; 2001

[83] Picinali L, Wallin A, Levtov Y, Poirier-Quinot D. Comparative perceptual evaluation between different methods for implementing reverberation in a binaural context. In: Audio Engineering Society Convention 142. Berlin, Germany: Audio Engineering Society; 2017. https://hal.archives-ouvertes.fr/hal-01790217

[84] Cuevas-Rodríguez M, Picinali L, González-Toledo D, Garre C, de la Rubia-Cuestas E, Molina-Tanco L, et al. 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation. PLoS One. 2019;**14**(3): e0211899. DOI: 10.1371/journal.pone.0211899

[85] Katz BFG, Picinali L. Spatial audio applied to research with the blind. In: Advances in Sound Localization. London, UK: InTech; 2011. pp. 225-250

[86] Picinali L, Afonso A, Denis M, Katz BFG. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. International Journal of Human-Computer Studies. 2014;**72**(4): 393-407. DOI: 10.1016/j.ijhcs.2013.12.008

[87] Pörschmann C, Stade P, Arend JM. Binauralization of Omnidirectional Room Impulse Responses – Algorithm and Technical Evaluation. In: 20th International Conference on Digital Audio Effects. Edinburgh, UK: University of Edinburgh; 2017. http://dafx.de/paper-archive/2017/papers/DAFx17 paper 25.pdf

[88] Ciotucha T, Ruminski A, Zernicki T, Mróz B. Evaluation of six degrees of freedom 3d audio orchestra recording and playback using multi-point ambisonics interpolation. In: Audio Engineering Society Convention 150. Audio Engineering Society; 2021. https://www.aes.org/e-lib/browse.cfm?elib=21052