Investigating sound-field reproduction methods as perceived by bilateral hearing aid users and normal-hearing listeners

Fernandez, Janani; McCormack, Leo; Hyvarinen, Petteri; Anne Kressner, Abigail

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
10.1121/10.0024875

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

The Journal of the Acoustical Society of America

FEBRUARY 20 2024

Investigating sound-field reproduction methods as perceived by bilateral hearing aid users and normal-hearing listeners

Janani Fernandez ; Leo McCormack ; Petteri Hyvärinen ; Abigail Anne Kressner

J. Acoust. Soc. Am. 155, 1492–1502 (2024)
https://doi.org/10.1121/10.0024875

ASA ACOUSTICAL SOCIETY OF AMERICA

JASA ARTICLE

Investigating sound-field reproduction methods as perceived by bilateral hearing aid users and normal-hearing listeners

Janani Fernandez,[1,a)] Leo McCormack,[1] Petteri Hyvärinen,[1] and Abigail Anne Kressner[2,a)]
[1]*Department of Information and Communications Engineering, Aalto University, Espoo, Finland*
[2]*Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark*

ABSTRACT:
A perceptual study was conducted to investigate the perceived accuracy of two sound-field reproduction approaches when experienced by hearing-impaired (HI) and normal-hearing (NH) listeners. The methods under test were traditional signal-independent Ambisonics reproduction and a parametric signal-dependent alternative, which were both rendered at different Ambisonic orders. The experiment was repeated in two different rooms: (1) an anechoic chamber, where the audio was delivered over an array of 44 loudspeakers; (2) an acoustically-treated listening room with a comparable setup, which may be more easily constructed within clinical settings. Ten bilateral hearing aid users, with mild to moderate symmetric hearing loss, wearing their devices, and 15 NH listeners were asked to rate the methods based upon their perceived similarity to simulated reference conditions. In the majority of cases, the results indicate that the parametric reproduction method was rated as being more similar to the reference conditions than the signal-independent alternative. This trend is evident for both groups, although the variation in responses was notably wider for the HI group. Furthermore, generally similar trends were observed between the two listening environments for the parametric method. The signal-independent approach was instead rated as being more similar to the reference in the listening room.
 https://doi.org/10.1121/10.0024875
(Received 28 April 2023; revised 26 January 2024; accepted 26 January 2024; published online 20 February 2024)
[Editor: Pavel Zahorik] Pages: 1492–1502

I. INTRODUCTION

Hearing assistive devices (HADs), such as hearing aids and cochlear implants, are typically custom fitted and calibrated for each individual user. These personalised fittings are usually performed at a clinic, where the surrounding sound sources and the acoustical characteristics of the environment may deviate from the situations the users may later encounter in their day-to-day lives. Indeed, it is common for users of HADs to report dissatisfaction with their devices when experiencing different real world scenarios.[1,2] This may be because established laboratory and clinical tests consider only simplistic sound scenes and static listening conditions,[3–7] despite several studies suggesting that such scenes may be a poor indicator of real world HAD performance.[8–11] Therefore, the ability to faithfully reproduce a variety of recorded or ecologically valid simulated sound scenes within these clinical settings may be desirable since this may help facilitate more optimal fittings or adjustments of devices so that they may be better suited to real world scenarios. Such sound-field reproduction methods may also find application in perceptual studies and HAD research and development, or be utilized for training the hearing abilities of HAD users.[12]

There are several existing reproduction methods in the literature, and while there is evidence of the perceptual accuracy of these methods, the accuracy generally relates to normal-hearing (NH) listeners experiencing the reproductions deployed in anechoic chambers.[13–18] There is, on the other hand, relatively little evidence of how these methods compare when deployed in rooms that are acoustically non-ideal, especially in terms of how the methods perform when they are experienced by hearing-impaired (HI) listeners.[19–21] This article, therefore, focuses on the investigation of a subset of currently available sound-field reproduction methods that could be deployed within clinical settings, and the main objective is to characterize the perceptual differences between these methods as perceived by HI listeners. Moreover, there is an observation regarding how non-ideal listening conditions, as present in an example listening room located at a clinic, can potentially impact these differences.

One popular signal-independent processing framework for the sound-field reproduction task is Ambisonics.[22,23] Ambisonic pipelines are divided into two stages: (1) a so-called *encoding* stage, whereby the microphone array signals, or sound objects, are transformed into the spherical harmonic (SH) domain;[24] and (2) a *decoding* stage, whereby the SH signals (also referred to as Ambisonic signals) are mapped to the playback channels to reproduce the sound field over a valid listening area. In traditional Ambisonics-based rendering pipelines, both of these stages are linear and

[a)]Also at: Copenhagen Hearing and Balance Centre, Rigshospitalet, Copenhagen, Denmark.

time-invariant (LTI) operations. The encoding stage is typically realised based upon a frequency-dependent regularised least squares fitting[25] of the microphone array directivities to the SH patterns. These array directivities may be determined through either free-field measurements, simulations, or analytical descriptors.[26] The frequency-dependent and SH order-dependent performance of the encoding process is then largely determined by the number of microphones in the array, their relative placement, and the construction of the mounting hardware. Crucially, the incorporation of more microphones in the array allows higher SH orders to be obtained, which subsequently leads to a higher spatial resolution of the captured scene. The all-round Ambisonic decoder (AllRAD)[27] is largely considered to represent the current state-of-the-art Ambisonics decoding approach, owing to its inherent ability to accommodate for irregular (non-uniform) loudspeaker arrangements, which are often encountered in practice.

The Ambisonics framework is also of particular interest as ecologically valid simulators for arbitrarily complex sound scenes have been made available,[28,29] which store the sound scene in this same Ambisonics format. In these simulators, room impulse responses are synthesised based on techniques such as ray-tracing or the image-source method for modelling specular reflections,[30] which is typically combined with separate handling of diffuse reverberation using shaped exponentially decaying noise sequences. The resultant spatial room impulse responses (one per source/receiver combination), may then be convolved with appropriate source stimuli, in order to obtain synthetic Ambisonic recordings. The room acoustics simulation procedures outlined in,[31,32] for example, are of particular note, since both reference loudspeaker array responses and Ambisonic responses (of arbitrary Ambisonic order) may be obtained via the simulator, which allows different reproduction methods to be easily compared against reference loudspeaker renders.

Previous studies exploring the use of Ambisonics reproduction within HAD or broader clinical contexts, however, have relied primarily on objective metrics to determine their feasibility,[33–36] or otherwise focused on speech intelligibility perceptual tests[37,38] or aspects related to motion-sickness within virtual reality environments.[39] Moreover, these perceptual studies investigating the performance of LTI Ambisonics reproduction pipelines have, for the most part, only been conducted with NH participants.[17,19,20,40–42] Nonetheless, the common conclusion is that using higher decoding orders leads to notable improvements in perceived spatial accuracy. However, this is problematic when considering that the most popular commercially-available Ambisonic array is comprised of only four microphones, which are often arranged in an open tetrahedral fashion. Such an array is only capable of first-order Ambisonics capture, and thus, due to this low resolution, directional sounds can become spatially blurred and lead to localisation ambiguities.[19,20] Furthermore, the spatial blurring of diffuse-sounds can lead to poor externalisation, timbral colourations, and reduced listener envelopment.[17,40,41] Naturally, these perceptual limitations may be alleviated by recording the sound scenes at higher orders. However, commercial Ambisonic arrays for higher-order capture are limited in availability, generally costly, and often offer these higher-order components only within narrow frequency bandwidths.

As an alternative to the decoding stage of the Ambisonics rendering framework, signal-dependent and parametric alternatives have been proposed for the task of adaptively mapping the Ambisonic signals to the playback channels.[13,15,43,44] These methods typically adopt a sound-field model, which formally describes the assumptions that are made regarding the composition of the sound field. The very first parametric method, intended for reproducing first-order Ambisonic sound scenes over loudspeakers, was directional audio coding (DirAC).[13] The method adopts a sound-field model that assumes the presence of a single source and/or diffuse isotropic reverberation, and, in practice, conducts direction-of-arrival (DoA) and diffuseness estimation in the time-frequency domain. While the first-order DirAC method is simplistic, formal perceptual studies have shown it to be comparable with LTI Ambisonic decoders operating at much higher-orders.[13] This is because sounds that are analysed as being directional (i.e., when diffuseness is low) are spatialised as a point source directly over the reproduction setup, which effectively represents a spatial sharpening operation. On the other hand, sounds that are analysed as being diffuse (i.e., when the diffuseness is high) are reproduced in a spatially-incoherent manner (using signal decorrelation), which is more in line with how such sounds would be experienced in nature. DirAC was also later extended to higher-orders[14,45] by subdividing the sound-field into directionally-constrained sectors, and conducting the DoA and diffuseness estimation independently for each. This allows the method to resolve more than one simultaneous sound source per time-frequency index, which has been shown to improve the perceived rendering accuracy.[14,45]

The more recent COMPASS method,[15] on the other hand, adopts an even more general sound-field model; which assumes the presence of a variable number of sound sources across time and frequency. Detection algorithms are employed to ascertain the number of active sources, followed by estimating their respective DoAs. Unlike DirAC, the diffuseness (or direct-to-reverberant ratio) parameter is not derived. Instead, COMPASS estimates and reproduces the diffuse ambience in the scene based on spatial filtering and decorrelation.[16] Here, after source beamformers have been steered towards the DoAs, and their signals subsequently spatialised over the target setup, the isolated source signals are re-encoded into the Ambisonics format and subtracted from the input recording. The resultant residual Ambisonic recording is then assumed to encapsulate the remaining diffuse reverberation and is reproduced via a plane wave decomposition (to a suitable spherical grid[46]), applying decorrelation operations, and then spatialising these decorrelated plane-waves over the same playback system.

In this article, a perceptual study involving ten HI listeners was conducted in an anechoic chamber to compare renders of a signal-independent Ambisonics decoder at first-, third-, and fifth-order and a parametric alternative at first- and third-order. The state-of-the-art all-round Ambisonic decoder (AllRAD)[27] and COMPASS[15] methods were selected as the candidate signal-independent and parametric decoders, respectively. The objective for this part of the study was to ascertain whether higher-order Ambisonics and/or parametric rendering would lead to measurable improvements in perceived similarity relative to simulated ground-truth recordings. The perceived accuracy by the HI listeners is compared with that of 15 NH subjects to assess potential differences. The second part of the study involved conducting the same tests, with the same 25 listeners, but instead using a comparable loudspeaker setup assembled in the Copenhagen Hearing and Balance Centre (CHBC), Rigshospitalet, Denmark, in an acoustically-treated (but not anechoic) clinical listening room. Importantly, the intention is not to directly compare the two listening environments, as the reference simulated scene will also be affected by the listening room acoustics, but rather to ascertain whether the same relative trends in the perceived accuracy between the different reproduction methods remain consistent between the two test environments.

II. METHODOLOGY

A. Participants

Ten bilateral hearing aid users with mild-to-moderate symmetrical hearing loss were recruited for the listening tests. A participant was considered to have symmetric hearing loss if the threshold differences between their left and right ears did not exceed 15 dB, i.e., 15 dB hearing level (HL), at any of the measured thresholds. The age range of this group was 24 to 76 years, with an average age of 54 years. Five participants were female. The participants wore their own hearing assistive devices during the tests. Device information for the HI group is presented in Table I. During the test, the HI participants were asked to keep their devices in the default, or most general purpose program, to best reflect their real world listening experience. The audiograms of all participants, averaged across both ears, are presented in Fig. 1. The participants for the NH group had auditory thresholds of 25 dB HL or less for the frequency range 125 Hz to 8 kHz. Fifteen NH listeners participated in testing. The age range of the NH group was 20 to 30 years, with an average of 26 years. Six of the participants in this group were female.

TABLE I. Hearing device brand and model for the HI group participants.

Participant ID	Brand	Model
1	Oticon	Selectic Napoli Pro2
2	Oticon	Viron 9 miniRite
3	Siemens	Pure 701
4	Interon	Centro 2
5	Oticon	OPN 3 Minirite
6	Oticon	More 1 miniRite
7	Widex	Beyond B-F2 440
8	Rexton	Emerald S
9	Widex	Beyond B-F2 440
10	Oticon	Ruby 2 miniRITE

B. Test cases

The reference loudspeaker audio files utilised in the listening test were rendered via a combination of ODEON (ODEON A/S, Lyngby, Denmark) and the Loudspeaker-based Room Auralization (LoRA) toolbox.[31,32] This procedure for creating the reference test case was validated in previous studies.[28,29] Two rooms were simulated (see Fig. 2). One room was a moderately sized seminar room set in a "group work" configuration, and the other was a small meeting room based on an existing room at the Technical University of Denmark (DTU). The RT_{60} of the two rooms was calculated according to the ISO 3382-1:2009 standard[47] and found to be 1.46 and 0.51 s, respectively. Additional details of the two simulated rooms are described in a previous publication.[48] Three sound sources were placed in these rooms to the left, right, and directly in front of the listener, as shown in Fig. 2.

The echograms and directional metadata for the simulated rooms were exported as text files from ODEON. These files were then fed to the LoRA toolbox to create multichannel reference room impulse responses for each sound source position within each room. The toolbox uses the metadata to directly render the early parts of the impulse response to the appropriate loudspeaker room impulse response, i.e., the room impulse responses corresponding to each loudspeaker position, via the use of nearest loudspeaker mapping (conducted independently for the two different playback loudspeaker setups), while the late part of the impulse response is modelled by deriving the energy and intensity envelopes of the late reflections in octave bands, then convolving these envelopes with uncorrelated noise sequences.[32] These multichannel impulse responses were then convolved with monophonic sound source signals to produce the reference sound

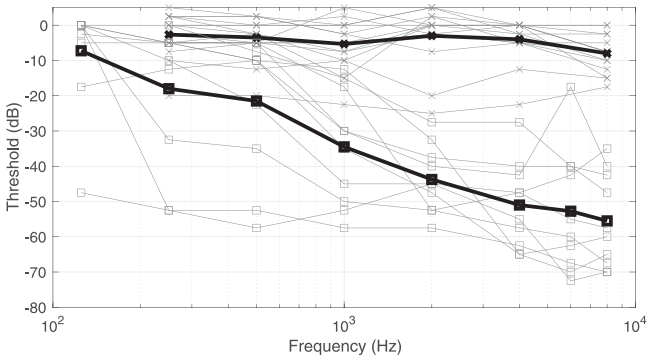

FIG. 1. The audiograms of the test participants. Each gray line represents the audiogram of a single participant, averaged over both ears. The audiograms of HI group participants are marked with the square symbol, while the NH group is marked with crosses. The average audiogram of each group is represented by the black lines.

https://doi.org/10.1121/10.0024875

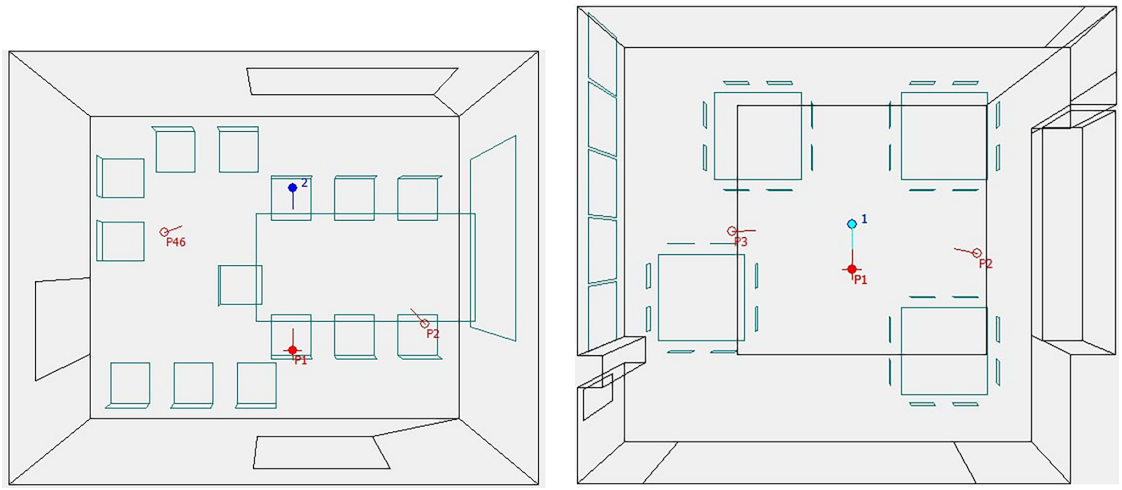

FIG. 2. (Color online) Top down view of the meeting room (left) and seminar room (right). Sound source positions are indicated by the red circles labelled P1, P2, and P3. The listener position is noted by the blue and cyan circles labelled 2 and 1, respectively.

scenes. The sound scenes chosen for testing consisted of three categories: *speech*, in which three competing talkers were present; *band*, which contained three different musical instruments (bass guitar, a shaker, and strings); and *mix*, which comprised a speaker and two noise sources (a pair of hands clapping and a water fountain). There were, therefore, six reference sound scenes in total (i.e., three per simulated room).

Each simulated reference sound scene was then also encoded into the Ambisonics format via the appropriate transforms,[24] which were applied to these reference loudspeaker scenes. These encoded sound scenes were then rendered for the same respective loudspeaker array setups using the reproduction methods under test, which were AllRAD[27] and COMPASS. The open-source AllRAD implementation found in the SPARTA audio plugin suite (v1.6.2)[49] was selected for this task, whereas the COMPASS renderings were obtained using the COMPASS decoder audio plugin, which may also be obtained via the SPARTA plugin suite installer.[50] First-, third-, and fifth-order AllRAD renderings and first- and third-order COMPASS renderings of the scenes were thus obtained. Note that it was not possible to obtain fifth-order renderings using COMPASS, as the Virtual Studio Technology (VST) implementation is limited to a maximum of third-order input.

The same pipeline and parameter settings were used to obtain the stimuli for both test environments, with the major difference being the respective loudspeaker configurations for the two rooms. Additionally, as one of the loudspeakers configurations was not spherical in nature, the distance compensator plugin from the IEM suite[51] was used to mitigate the effects this difference would cause.

C. Test environments

The tests were conducted in both a free-field and a non-free-field (but acoustically treated) environment. The Audio Visual Immersion Lab (AVIL) at DTU served as the free field environment. This room is an anechoic chamber of dimensions $7.0 \times 8.0 \times 6.0\,\mathrm{m}^3$, fitted with 64 loudspeakers (KEF, Maidstone, UK) arranged in a three-dimensional (3D) spherical configuration, of which the 44 loudspeakers that comprise the upper hemisphere were utilised for testing. The loudspeakers utilized for this study were arranged in the concentric circles, with 24, 12, 6, and 2 loudspeakers at elevation angles 0°, 28°, 56°, and 80°, respectively. Impulse response measurements taken at the central listening position were used to apply the appropriate time, level, and magnitude corrections to the stimuli signals in this environment.

The second (non-free-field) listening room environment was the Spatial Hearing Lab at the Copenhagen Hearing and Balance Centre (CHBC) at Rigshospitalet, Copenhagen, Denmark. The room consisted of an acoustically treated room of dimensions $3.4 \times 4.4 \times 2.8\,\mathrm{m}$ in which 41 loudspeakers (KEF, Maidstone, UK) were fitted. These loudspeakers were placed flush with the walls in a rectangular arrangement. All of the loudspeaker directions, with respect to the central listener position, corresponded to the loudspeaker directions of the 3D spherical grid arrangement in AVIL, with the exception of the 56° ring having four uniformly-spaced loudspeakers (instead of six), and the two uppermost loudspeakers are instead represented by a single loudspeaker at 90° elevation in CHBC. The room had a broadband RT_{60} of 0.13 s (mean over all loudspeaker directions, as measured in the listening position). The magnitude corrections for this setup were performed by placing a microphone at the listening position, playing white noise through each individual speaker, and then calculating the required corrections. Time and level differences for the individual loudspeakers were calculated based on impulse response measurements.

D. Test design and procedure

The study involved a multiple-stimulus listening test in which participants were presented with a known simulated

reference sound scene and the output of five different reproduction methods: first-, third-, and fifth-order AllRAD and first- and third-order COMPASS. The simulated reference sound scene was also included as a hidden reference. Therefore, the total number of test cases for each sound scene was six. The participants were able to listen to the reference scene and the six reproduced outputs as many times as they wished before making their judgements. They were asked to answer the question "To what extent are the sound samples different from the reference?" Their answers were recorded as ratings of each test case on a scale of 0 to 100 based on the perceived similarity when compared to the reference, with 100 being perceptually identical and 0 being perceptually very different to the reference. Participants gave their ratings using virtual sliders on a graphical user interface running on an iPad (Apple Inc., Cupertino, CA). As not all participants were familiar with such perceptual tests, the participants were also given the following questions to help guide their ratings: "Are the sounds coming from the same direction as the reference? Do the sounds seem like they are the same size as in the reference? Do you perceive a variation in pitch?" The participants were encouraged to use the full range of the scale for each trial. The test was run twice in each of the two test environments. The first run was considered as training and used as a way to familiarize the participants with the test interface. It was, therefore, excluded from the results. Thereafter, the order of testing was randomized, with some participants performing the listening test in the free-field environment first, while others performed the test in the listening room environment first.

E. Statistical analysis

The listening test results data were analyzed using Matlab version 2022a. Violin plots were created with the aid of the GitHub repository maintained by Bastian Bechtold.[52] Friedman tests and *post hoc* pairwise multiple comparison tests were performed using the MATLAB `multicompare` function in order to ascertain statistically significant differences between the ratings for different rendering methods. Correction for multiple comparisons using Tukey's HSD procedure was applied during *post hoc* analysis.

III. RESULTS AND DISCUSSION

Figure 3 shows the violin plots of the NH control group ratings for the free-field room, while Fig. 4 displays the violin plots of the HI group ratings for the same test environment. Similarly, Figs. 5 and 6 are violin plots of the control and HI group results in the listening room test environment. To further analyse the results, a Friedman test was performed to determine if there were significant differences between the ratings within each of the two groups for each of the six test cases, i.e., sound scenes. These tests revealed that there were indeed statistically significant differences in the ratings across all six sound scenes for all the data cases, with the exceptions of the HI group results in the listening room for condition Meeting Room/Speech. See Table II for Friedman test statistics of the free-field environment and for the listening room. Subsequently, the *post hoc* analyses revealed several statistically significant differences in the ratings between the linearly decoded Ambisonic (AllRAD) renderings, COMPASS renderings, and the respective

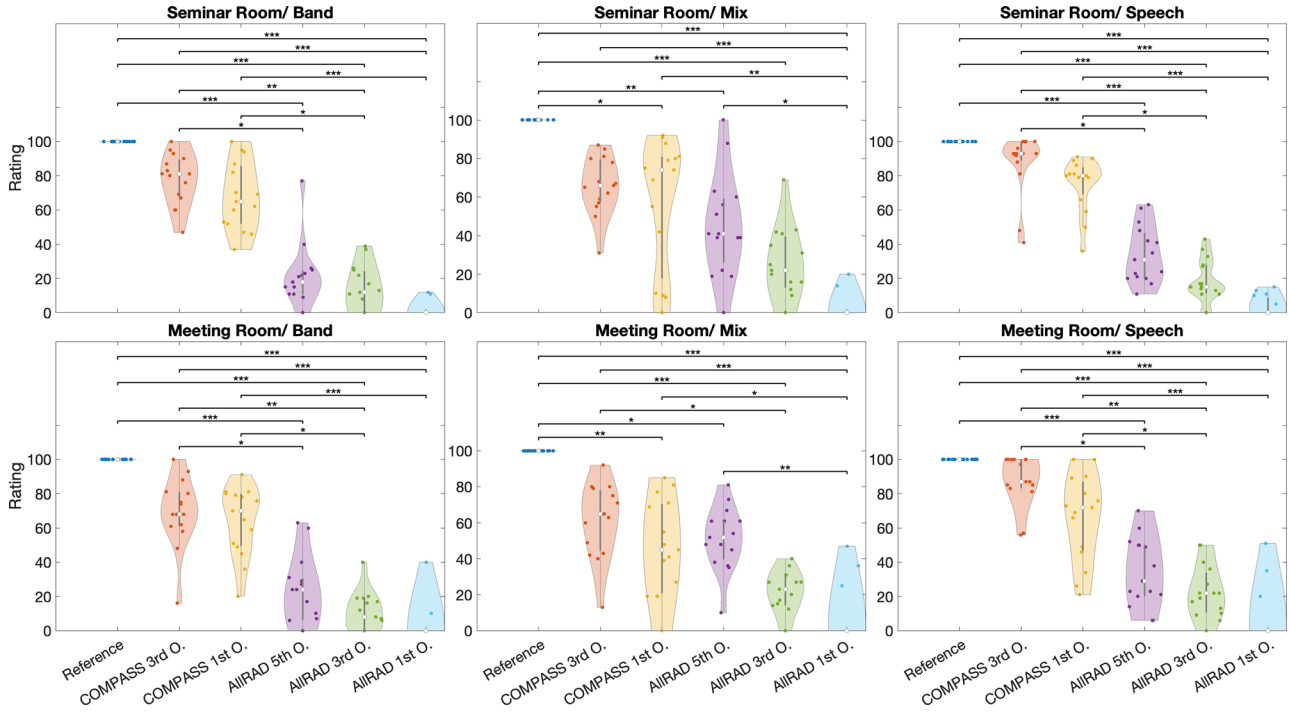

FIG. 3. (Color online) Violin plots of the free-field environment results for the NH group, where * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

https://doi.org/10.1121/10.0024875

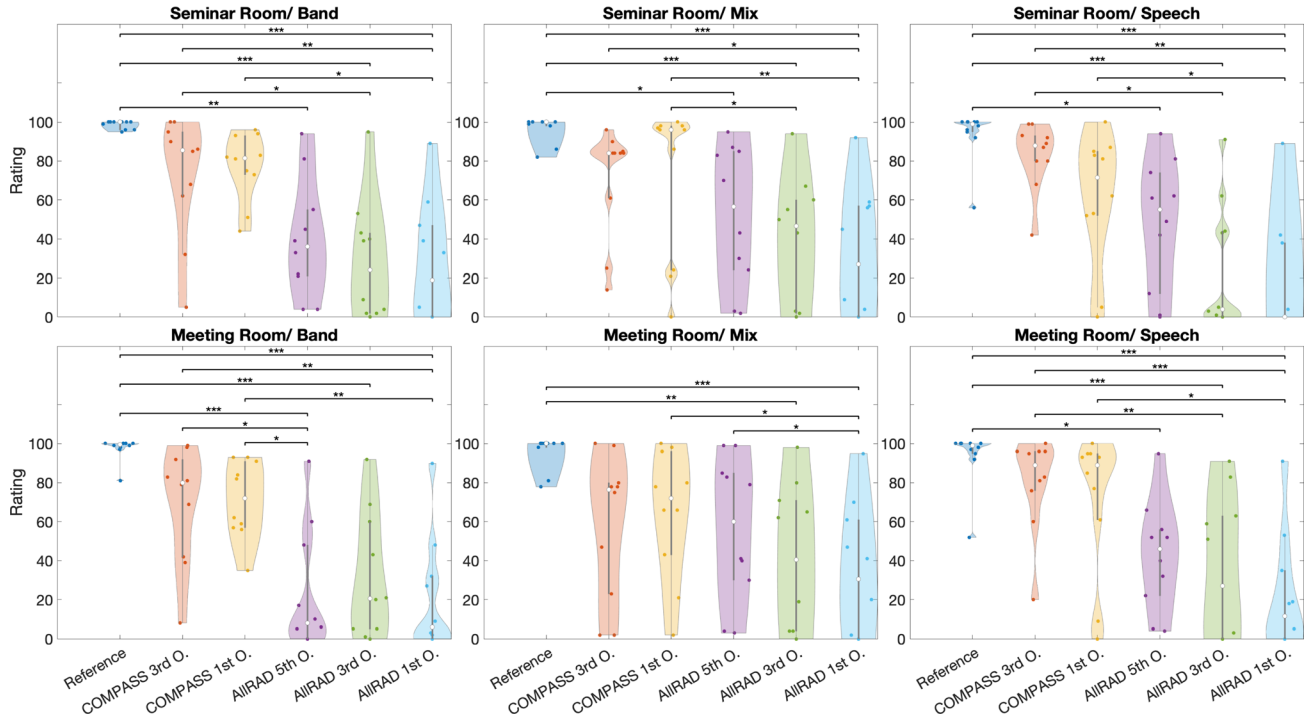

FIG. 4. (Color online) Violin plots of the free-field environment results for the HI group, where * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

reference of each sound scene. These significant differences are indicated in the respective results figures as horizontal black lines linking the two groups between which the statistical difference was discovered, with the number of asterisks above the horizontal lines indicating the level of significance.

A. Free-field environment

In the free-field environment (Fig. 3), the NH group correctly identified the reference in all the test cases. The first-order linearly decoded AllRAD renderings are consistently rated the lowest of all the test cases. The COMPASS

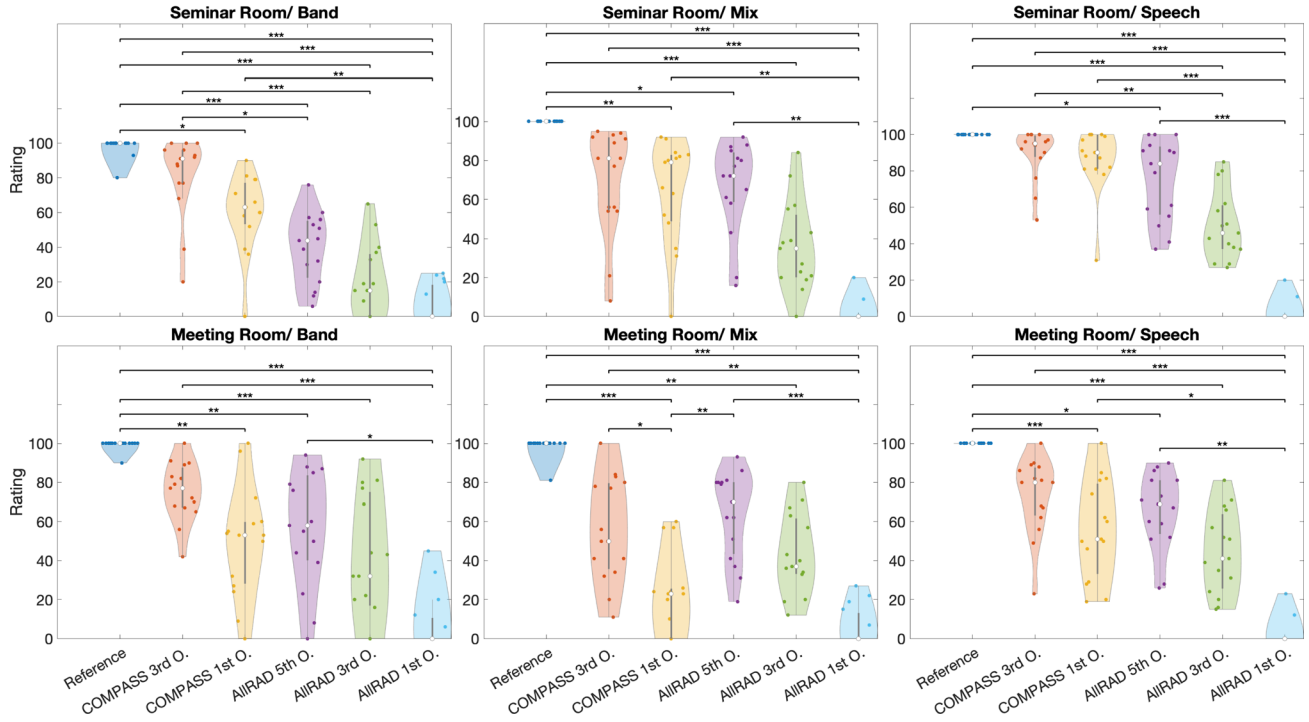

FIG. 5. (Color online) Violin plots of the listening room environment results for the NH group, where * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

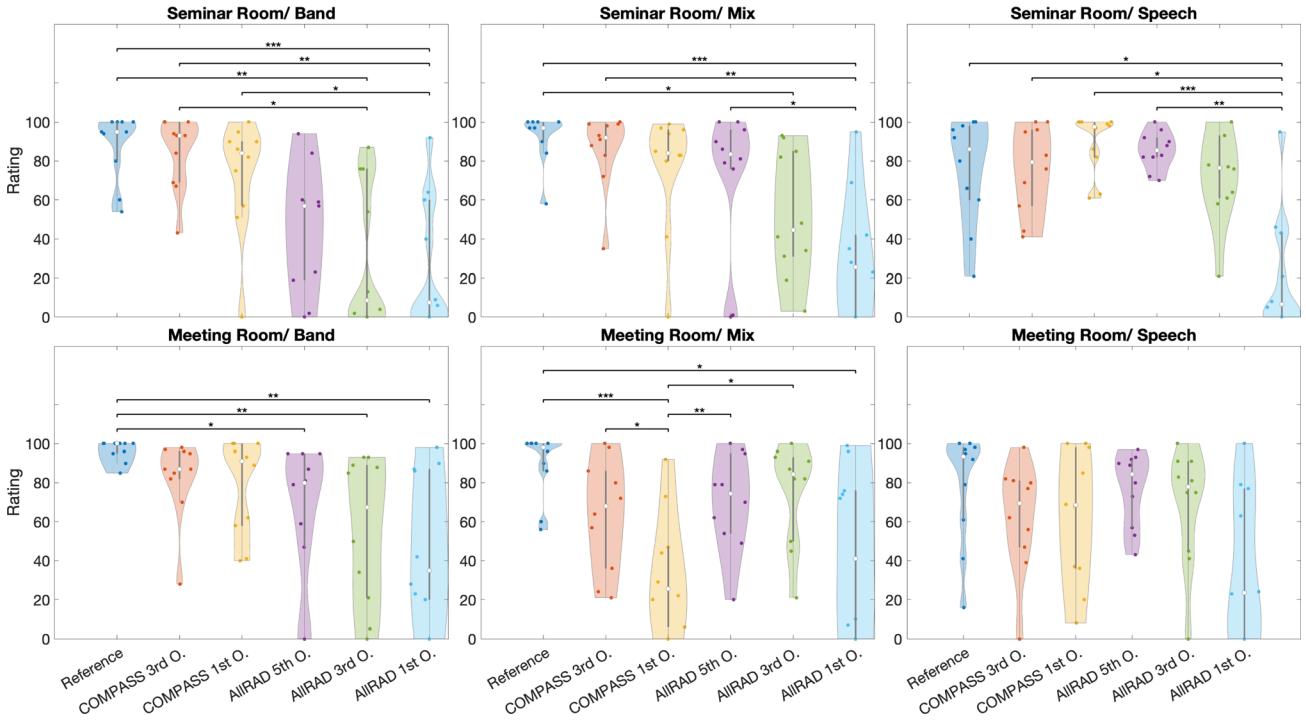

FIG. 6. (Color online) Violin plots of the listening room environment results for the HI group, where * indicates $p < 0.05$, ** indicates $p < 0.01$, and *** indicates $p < 0.001$.

renderings were rated relatively high on the scale in comparison to the AllRAD renderings and there were no statistically significant differences found between the ratings for these renderings and the reference ratings in the majority of the test cases. In four of six test cases, first-order COMPASS ratings were rated higher than and found to be significantly different to the ratings for first- and third-order AllRAD renderings. For the same four test cases, third-order COMPASS ratings were rated higher than first-, third-, and fifth-order AllRAD renderings, findings that did reach significance in *post hoc* analyses.

TABLE II. Results of the Friedman rank sum test for each condition for the free-field environment, where *** indicates $p < 0.001$ and ns. indicates $p > 0.05$.

Group	Condition	χ^2_5	
		Free-field	Listening Room
NH	Seminar room / Band	68.14***	65.04***
NH	Seminar room / Mix	58.07***	59.98***
NH	Seminar room / Speech	71.06***	25.35***
NH	Meeting room / Band	67.88***	51.03***
NH	Meeting room / Mix	26.73***	61.60***
NH	Meeting room / Speech	64.31***	55.66***
HI	Seminar room / Band	40.58***	30.47***
HI	Seminar room / Mix	37.19***	25.20***
HI	Seminar room / Speech	36.48***	61.70***
HI	Meeting room / Band	45.68***	21.20***
HI	Meeting room / Mix	30.22***	26.73***
HI	Meeting room / Speech	38.62***	9.80$^{ns.}$

The results of the HI group (Fig. 4) in the free-field environment also display, in general, the same trend in the median scores as was seen in the control group, but with fewer significant findings. These scores were also more variable than the scores for the NH group, particularly with the AllRAD rendered test cases. However, similarly to the control NH group, the median scores of the reference test case are consistently the highest for each test scene, indicating that the hidden references were correctly identified in the majority of cases. The median scores for the COMPASS rendered test cases are, in general, also in the top half of the scale. *Post hoc* analyses for the four test cases, which did not use the "mix" stimuli, found these differences to be statistically significant in pairwise comparisons between COMPASS and AllRAD renderings of the same order in all test cases except the Meeting room "band" stimuli test case, in which the pair-wise comparison between third-order COMPASS and the third-order AllRAD renderings were not found to be statistically significant. The analyses also found the difference between third-order COMPASS ratings and first-order AllRAD ratings to be statistically significant for these four test cases.

Two of the six test cases, in which the "mix" stimuli were used, deviate from the previously noted trends. For the other test cases, the majority of the AllRAD scores were compressed towards the bottom half of the violin plot figures, whereas for the "mix" stimuli test cases fifth-order AllRAD renderings show more variance and the medians have now moved towards the top half of the scale for both NH and HI groups. In contrast, the scores for the COMPASS renderings show more variance and there are

https://doi.org/10.1121/10.0024875

now fewer significant findings between the ratings for these rendered scenes and the AllRAD rendered scenes. For such "mix" stimuli, it is well-known that such a mixture of speech stimuli alongside impulsive stimuli is a difficult scenario for parametric audio rendering methods, especially at lower-orders, as these can violate their assumed sound-field model.[14,45] Notably, with COMPASS renderings being less perceptually similar to the reference in these test cases, all AllRAD renderings were rated higher, indicating more similarity to the reference than in other test cases, yet still received lower median scores than the COMPASS renderings, indicating COMPASS renderings are still perceived as being more similar to the reference than the AllRAD renderings.

The greater perceived inaccuracies of AllRAD in comparison to COMPASS renderings may be due to the coherent spreading caused by the former method, in which the same signal arrives at a listener's ears from multiple different angles of incidence simultaneously. This effect is inherent to linearly decoded Ambisonics and is especially prevalent at lower orders. This may be perceived as: pin-point sources being spatially blurred and more difficult to localise,[19,20] comb-filtering artefacts,[41] and diffuse sounds being rendered as coherent sounds; with the latter potentially sounding unnatural, and leading to a reduced sense of listener envelopment.[40] However, since COMPASS collapses the energy of directional sounds into one specific direction (reproduced using amplitude panning[53]) and applies decorrelation operations to sounds deemed to be more diffuse, this aforementioned coherent spreading problem of AllRAD (and the resulting perceptual issues that this incurs) is largely circumvented. This may be a reason for test participants perceiving AllRAD renderings, particularly at lower orders, to be perceptually very different from the reference, as they may have heard these localisation shifts and timbral issues, and this unnatural coherent rendering of diffuse sounds.

While none of the rendering methods were truly indistinguishable from the reference in all listening conditions, these findings imply that some methods were perceived as being closer to the reference than others, i.e., they were perceived to be more accurate relative to the "ground truth". First- and third-order COMPASS renderings appear to outperform the higher orders of AllRAD renderings in the majority of cases. In the test cases in which they performed poorly, i.e., the "mix" stimuli, COMPASS still seemed to render more perceptually accurate scenes than AllRAD, as indicated by these renderings receiving scores with higher medians than the equivalent order of AllRAD in all test cases; with most of these differences in scores found to be significantly different. These findings lend support to the use of COMPASS, as opposed to AllRAD, for rendering in free-field environments for both NH and HI participants; at least for the types of sound scenes considered in the present study.

B. Listening room environment

In Fig. 5, it can be seen that the same general trends that were reported for the free-field environment persist in the listening room environment for the NH group. However, the median scores of the AllRAD rendered test cases are higher in this environment than in the free field environment, with the average difference between median scores being 33.67 for fifth-order AllRAD renderings and 17.33 for third-order AllRAD renderings. This may be due to the effect of the room reverberation time, which imposes some degree of signal decorrelation onto the loudspeaker playback signals due to reflection and scattering effects in this environment. This may mitigate the perceptual problems arising due to the coherent spreading of directional and diffuse sounds (as described in the previous subsection), which in turn may explain why fewer significant differences were found between the ratings of COMPASS and AllRAD renderings. Third- and first-order COMPASS do appear to be less perceptually distinguishable with the reference than first-order AllRAD, as indicated by the higher median ratings and the significant differences in the pair-wise comparisons. There are, however, fewer significant differences found in comparisons between third-order COMPASS ratings and the equivalent or higher order AllRAD ratings, and no significant differences in comparisons with the reference ratings. This indicates that depending on the sound scene, COMPASS produces renderings that are as perceptually distinguishable, if not less distinguishable, with the reference than the equivalent AllRAD rendering.

In Fig. 6, it can be seen that most pair-wise comparisons were not found to be significant in this test environment. Nevertheless, the HI group ratings display a similar trend in the median scores for the listening room environment as the trends observed in the free-field environment. Notably, there are more instances of the reference not being rated the highest amongst the test cases, implying that the participants may have struggled to perceive differences in this setting. There is also a high variance in these results. The ratings for the first-order AllRAD renderings were, in nearly all test scenes, statistically different from the ratings for the reference, while scores for third-order AllRAD renderings were found to be statistically different in three of the six test scenes. For test scenes involving the seminar room simulation, the COMPASS renderings were found to have significant differences when compared with first-order AllRAD renderings. The ratings for fifth-order AllRAD renderings were found to be statistically distinguishable from the ratings for the reference test case in only one sound scene, as were the findings for the first-order COMPASS rendered scenes. Notably, in the sound scenes involving just speech stimuli, fewer statistically significant differences were found between the ratings for the various methods and the reference test cases, with one test case having no significant findings. Additionally, consistent with the previous findings, the third-order COMPASS renderings were not statistically different from the reference for any of the sound scenes, however, fewer significant differences were found between these ratings and the ratings for third- and fifth-order AllRAD renderings.

While it is difficult to form conclusions based on these findings due to the effect of the room on the reference as

well as the rendered sound scenes, they do indicate that COMPASS may be more suitable than AllRAD when used for the purpose of rendering sound scenes in a listening room setup. The evidence for this is stronger in the case of the NH listeners, as the picture is in general less clear for the HI group due to a large amount of variability in the test results. Interestingly, the variance of the ratings for AllRAD renderings tended to be larger than for the COMPASS renderings and reference in both environments, which in of itself may be a disadvantage of the AllRAD method.

IV. FUTURE WORK

This study compared COMPASS rendered sound scenes and AllRAD renderings of the same scene to a simulated ideal reference. While using this particular simulation method for the purposes of creating an ideal reference has been validated by other studies, more research is of course needed to fully understand the perceptual implications of rendering real recorded sound scenes. Moreover, one possible limitation of this study design is that the reference cases between the two rooms used for testing differed and participants attended the test at each location on two separate days, making comparisons across the listening environments more challenging. An alternative method of testing would be to conduct the study only in the free-field environment and instead simulate the room acoustics of listening room environments using a room simulator, and then impose these characteristics onto the test stimuli; similar to the approach described in a previous study.[17] In this case, it would be possible to retain the same reference scene across different simulated listening room environments and, therefore, to facilitate more direct comparisons between different simulated listening rooms.

Another possible limitation of the current study is that the HI listeners had different levels of hearing loss and also wore a variety of hearing devices that were programmed by different hearing aid dispensers. This latter limitation means that the fitting procedures themselves could also have varied, especially in regards to whether and to what extent the devices were optimized binaurally. This variability, alongside the varying levels of hearing loss and the difference in ages between the two groups, likely contributed to the high level of variance in the perceptual ratings across the HI group. A future study could introduce controls for the fitting procedure and the hearing aid devices themselves in order to clarify whether the variability is a characteristic of hearing aid users generally or simply confounded with device differences.

Only one specific parametric sound field rendering method was explored in this study, while other methods, such as DirAC, are also popular and may be explored in a similar context. Additionally, the sound playback in this study was via loudspeakers. It would, however, be interesting to explore the feasibility of parametric spatial audio reproduction methods as clinical tools when playback is over headphones instead, as the outcomes may differ. In particular, it would be worthwhile to compare parametric binaural rendering methods to signal-independent binaural rendering methods, as the latter have recently received proposals for perceptually-motivated optimisations.[54,55] While such comparisons have been conducted involving NH listeners,[16] the current study has highlighted the need for investigations to specifically include HI users if the intended use of such methods is indeed within the context of HI users.

V. CONCLUSION

This article details the findings of a study that explored the feasibility of recreating different sound scenes using two different sound field rendering methods. A signal-independent rendering approach, AllRAD, and a parametric rendering method, COMPASS, were selected for investigation. Two rooms of differing spatial characteristics were designed using a hybrid room acoustic simulation system in order to create sound scenes to be used as a reference. These were then compared to first-, third-, and fifth-order AllRAD renderings of the same sound scenes, as well as first- and third-order COMPASS renderings. Ten bilateral HI listeners and 15 NH listeners were recruited for the study. These participants performed a perceptual listening test to compare the rendering methods, and they did so twice, once in a free-field environment and once in a non-free-field, acoustically treated environment; the latter of which may be more feasibly constructed within a clinical setting.

The results indicate that sound scenes rendered by COMPASS were perceptually more similar to the reference than scenes rendered with the AllRAD method. This was implied by the higher median scores of the COMPASS renderings, while the AllRAD rendered stimuli received lower scores which were found to be significantly different from the reference and COMPASS renderings in most of the test conditions. Individual pairwise contrasts in the *post hoc* analysis should, however, be interpreted with caution—it is, for example, clear that COMPASS was consistently rated below the reference although the differences did not reach statistical significance. Nevertheless, these findings suggest that given the potential advantages of COMPASS in terms of the reduced microphone array requirements, it could be employed in free-field conditions for both NH and HI listeners for the types of sound scenes employed in this study. In non-free-field conditions, it is difficult to form conclusions, given that the reference was affected by the room acoustics. However, the trends in the results for this listening environment for the NH group are similar to the trends in free-field conditions. While this implies that COMPASS may be suitable for rendering sound scenes even in non-free-field environments, further investigations are needed. This is particularly true for HI participants as the large variability in the ratings between the rendering methods are confounded with the variability present among the hearing devices employed within the group.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers and the editor for their helpful comments and suggestions.

[1]S. Hougaard and S. Ruf, "EuroTrak I: A consumer survey about hearing aids in Germany, France, and the UK," Hear. Rev. **18**(2), 12–28 (2011).
[2]L. L. Wong, L. Hickson, and B. McPherson, "Hearing aid satisfaction: What does research from the past 20 years say?," Trends Amplif. **7**(4), 117–161 (2003).
[3]J. B. Nielsen and T. Dau, "The Danish hearing in noise test," Int. J. Audiol. **50**(3), 202–208 (2011).
[4]M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**(2), 1085–1099 (1994).
[5]ISO 8253-2: "Acoustics–Audiometric test methods–Part 2: Sound field audiometry with pure-tone and narrow-band test signals" (ISO, Geneva, Switzerland, 2009).
[6]ISO 8253-3: "Acoustics. Audiometric test methods–Part 3: Speech audiometry" (ISO, Geneva, Switzerland, 1996).
[7]ISO, "Pure-tone air and bone conduction audiometry" (ISO, Geneva, Switzerland, 2010).
[8]M. Cord, D. Baskent, S. Kalluri, and B. Moore, "Disparity between clinical assessment and real-world performance of hearing aids," Hear. Rev. **14**(6), 22 (2007).
[9]T. Ricketts, "Impact of noise source configuration on directional hearing aid benefit and performance," Ear Hear. **21**(3), 194–205 (2000).
[10]B. E. Walden, R. K. Surr, M. T. Cord, B. Edwards, and L. Olson, "Comparison of benefits provided by different hearing aid technologies," J. Am. Acad. Audiol. **11**(10), 540–560 (2000).
[11]C. L. Compton-Conley, A. C. Neuman, M. C. Killion, and H. Levitt, "Performance of directional microphones for hearing aids: Real-world versus simulation," J. Am. Acad. Audiol. **15**(06), 440–455 (2004).
[12]C. Valzolgher, J. Gatel, S. Bouzaid, S. Grenouillet, M. Todeschini, G. Verdelet, R. Salemme, V. Gaveau, E. Truy, A. Farnè, and F. Pavani, "Reaching to sounds improves spatial hearing in bilateral cochlear implant users," Ear Hear. **44**(1), 189–198 (2023).
[13]V. Pulkki, A. Politis, M.-V. Laitinen, J. Vilkamo, and J. Ahonen, "First-order directional audio coding (DirAC)," in *Parametric Time-Frequency Domain Spatial Audio*, edited by V. Pulkki, S. Delikaris-Manias, and A. Politis (John Wiley & Sons, New York, 2017), pp. 89–138.
[14]A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," IEEE J. Sel. Top. Signal Process. **9**(5), 852–866 (2015).
[15]A. Politis, S. Tervo, and V. Pulkki, "COMPASS: Coding and multidirectional parameterization of Ambisonic sound scenes," in *Proceedings of the 2018 ICASSP*, Calgary, Canada (April 15–20, 2018)., pp. 6802–6806.
[16]L. McCormack and A. Politis, "Estimating and reproducing ambience in Ambisonic recordings," in *Proceedings of the 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, Serbia (August 29–September 3, 2022), pp. 314–318.
[17]O. Santala, H. Vertanen, J. Pekonen, J. Oksanen, and V. Pulkki, "Effect of listening room on audio quality in Ambisonics reproduction," in *Proceedings of the Audio Engineering Society Convention 126*, Munich, Germany (May 7–10, 2009).
[18]V. Hohmann, R. Paluch, M. Krueger, M. Meis, and G. Grimm, "The virtual reality lab: Realization and application of virtual sound environments," Ear Hear. **41**(Suppl 1), 31S–38S (2020).
[19]S. Braun and M. Frank, "Localization of 3D Ambisonic recordings and Ambisonic virtual sources," in *Proceedings of the 1st International Conference on Spatial Audio*, Detmold, Germany (November 10–13, 2011).
[20]S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources," Acta Acust. united Ac. **99**(4), 642–657 (2013).
[21]T. Koski, V. Sivonen, and V. Pulkki, "Measuring speech intelligibility in noisy environments reproduced with parametric spatial audio," in *Proceedings of the Audio Engineering Society Convention 135*, New York, NY (October 17–20, 2013).
[22]M. A. Gerzon, "Periphony: With-height sound reproduction," J. Audio Eng. Soc. **21**(1), 2–10 (1973).
[23]F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer Nature, New York, 2019).
[24]B. Rafaely, *Fundamentals of Spherical Array Processing* (Springer, New York, 2015), Vol. 8.
[25]S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with higher order Ambisonics–objective measurements and validation of a 4th order spherical microphone," in *Proceedings of the 120th Convention of the AES*, Paris, France (May 20–23, 2006), pp. 20–23.
[26]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999).
[27]F. Zotter and M. Frank, "All-round Ambisonic panning and decoding," J. Audio Eng. Soc. **60**(10), 807–820 (2012).
[28]S. Favrot and J. M. Buchholz, "Validation of a loudspeaker-based room auralization system using speech intelligibility measures," in *Proceedings of the Audio Engineering Society Convention 126*, Munich, Germany (May 7–10, 2009).
[29]J. Cubick and T. Dau, "Validation of a virtual sound environment system for testing hearing aids," Acta Acust. united Ac. **102**(3), 547–557 (2016).
[30]L. Savioja, "Modeling techniques for virtual acoustics," Thesis, Helsinki University of Technology, Espoo, Finalnd (1999).
[31]S. E. Favrot, *A Loudspeaker-Based Room Auralization System for Auditory Research* (Technical University of Denmark, Lyngby, Denmark, 2010).
[32]S. Favrot and J. M. Buchholz, "LoRA: A loudspeaker-based room auralization system," Acta Acust. united Ac. **96**(2), 364–375 (2010).
[33]C. Oreinos and J. Buchholz, "Validation of realistic acoustic environments for listening tests using directional hearing aids," in *Proceedings of the 2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Juan-les-Pins, France (September 8–11, 2014), pp. 188–192.
[34]C. Oreinos and J. M. Buchholz, "Objective analysis of Ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones," J. Acoust. Soc. Am. **137**(6), 3447–3465 (2015).
[35]G. Grimm, S. Ewert, and V. Hohmann, "Evaluation of spatial audio reproduction schemes for application in hearing aid research," Acta Acust. united Ac. **101**(4), 842–854 (2015).
[36]L. S. Simon, H. Wuethrich, and N. Dillier, "Comparison of higher-order Ambisonics, vector- and distance-based amplitude panning using a hearing device beamformer," in *4th International Conference on Spatial Audio*, Graz, Austria, September 7–10 (2017).
[37]N. Mansour, M. Marschall, T. May, A. Westermann, and T. Dau, "Speech intelligibility in a realistic virtual sound environment," J. Acoust. Soc. Am. **149**(4), 2791–2801 (2021).
[38]V. Best, G. Keidser, J. M. Buchholz, and K. Freeston, "An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment," Int. J. Audiol. **54**(10), 682–690 (2015).
[39]K. Sun, N. H. Pontoppidan, D. Wendt, and L. Bramsløw, "Perception of virtual reality based audiovisual paradigm for people with hearing impairment," BNAM **1**, 95–104 (2022).
[40]A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," J. Acsout. Soc. Am. **133**(5), 2711–2721 (2013).
[41]P. Stitt, S. Bertet, and M. van Walstijn, "Off-centre localisation performance of Ambisonics and HOA for large and small loudspeaker array radii," Acta Acust. united Ac. **100**(5), 937–944 (2014).
[42]D. Thery and B. F. Katz, "Auditory perception stability evaluation comparing binaural and loudspeaker Ambisonic presentations of dynamic virtual concert auralizations," J. Acoust. Soc. Am. **149**(1), 246–258 (2021).
[43]S. Berge and N. Barrett, "High angular resolution planewave expansion," in *Proceedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France (May 6–7, 2010), pp. 6–7.
[44]A. Wabnitz, N. Epain, A. McEwan, and C. Jin, "Upscaling Ambisonic sound scenes using compressed sensing techniques," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (October 16–19, 2011), pp. 1–4.
[45]A. Politis, L. McCormack, and V. Pulkki, "Enhancement of Ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY (October 15–18, 2017), pp. 379–383.
[46]R. H. Hardin and N. J. Sloane, "McLaren's improved snub cube and other new spherical designs in three dimensions," Discrete Comput. Geom. **15**, 429–441 (1996).
[47]ISO 3382: "Acoustics—Measurement room acoustic parameters–Part 1: Performance spaces" (ISO, Geneva, Switzerland, 2009).

[48]A. A. Kressner, A. Westermann, and J. M. Buchholz, "The impact of reverberation on speech intelligibility in cochlear implant recipients," J. Acoust. Soc. Am. **144**(2), 1113–1122 (2018).
[49]L. McCormack and A. Politis, "SPARTA & COMPASS: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Proceedings of the Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, York, UK (March 27–29, 2019).
[50]An open-source MATLAB reference implementation of the COMPASS method may also be found at https://github.com/polarch/COMPASS-ref. The VST audio plugin implementation is downloadable at https://leomccormack.github.io/sparta-site/.
[51]D. Rudrich, "IEM plug-in suite," https://plugins.iem.at/ (Last viewed September 27, 2022).
[52]B. Bechtold, "Violin plots for MATLAB," Github Project 10, https://github.com/bastibe/Violinplot-Matlab (Last viewed December 7, 2022).
[53]V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," J. Audio Eng. Soc. **45**(6), 456–466 (1997).
[54]M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of Ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," J. Acoust. Soc. Am. **143**(6), 3616–3627 (2018).
[55]C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of Ambisonic signals via magnitude least squares," in *Proceedings of the DAGA*, Munich, Germany (March 19–22, 2018), pp. 339–342.