Applied Acoustics 188 (2022) 108523

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Compensating first reflections in non-anechoic head-related transfer function measurements

Jose J. Lopez [a,*], Pablo Gutierrez-Parera [a], Maximo Cobos [b]

[a] iTEAM Institute, Universitat Politècnica de València, 46022 Valencia, Spain
[b] Departament d'Informàtica, Universitat de València, 46100 Burjassot, Spain

ARTICLE INFO

Article history:
Received 28 June 2021
Received in revised form 1 October 2021
Accepted 5 November 2021

Keywords:
HRTF
Acoustic channel measurement
Echo removal
Plane wave decomposition
Spherical arrays
Spatial audio

ABSTRACT

Personalized Head-Related Transfer Functions (HRTFs) are needed as part of the binaural sound individualization process in order to provide a high-quality immersive experience for a specific user. Signal processing methods for performing HRTF measurements in non-anechoic conditions are of high interest to avoid the complex and inconvenient access to anechoic facilities. Non-anechoic HRTF measurements capture the effect of room reflections, which should be correctly identified and eliminated to obtain HRTFs estimates comparable to ones acquired in an anechoic setup. This paper proposes a sub-band frequency-dependent processing method for reflection suppression in non-anechoic HRTF signals. Array processing techniques based on Plane Wave Decomposition (PWD) are adopted as an essential part of the solution for low frequency ranges, whereas the higher frequencies are easily handled by means of time-crop windowing methods. The formulation of the model, extraction of parameters and evaluation of the method are described in detail. In addition, a validation case study is presented showing the suppression of reflections from an HRTF measured in a real system. The results confirm that the method allows to obtain processed HRTFs comparable to those acquired in anechoic conditions.

1. Introduction

During the last decades, spatial audio has become a very important research topic with applications in different fields, including Virtual and Augmented Reality (VR, AR), Audiology or Psychology, and also in the musical industry [1–3]. According to [4], audiovisual content is increasingly consumed through headphones, resulting in a remarkable expansion of headset-based audio playback in recent years. In addition, headphones' portability and individuality make them the ideal complement to the head-mounted displays used in VR and AR. In this context, headphone-based binaural rendering is particularly interesting, allowing spatial audio scenarios to be simulated at a relatively low cost. Binaural hearing refers to the ability of the auditory system to analyze the sound at the two ears, integrate the information embedded in the acoustic stimuli, and perceive sound as coming from a three-dimensional space [5]. When binaural rendering is done in real time, the simulation can be interactive, adding a greater sense of realism and presence to multimedia content, which has applications in VR and AR [6].

In order to create a spatialized sound experience, binaural audio uses Head-Related Impulse Responses (HRIRs) or their frequency-domain equivalent, known as Head-Related Transfer Functions (HRTF) [7]. For a specific position in space, the HRTF captures the effect that a sound source has on a subject's ear canals, under free-field conditions. The HRTF includes the contributions of the human body, which influence the way individuals perceive sound. These contributions come mainly from the head, torso and pinna, all of which are described by the HRTF [8]. Since the anthropometric characteristics differ among individuals and have a strong influence on sound perception, HRTFs present tailored features that make them specific for each subject. Therefore, individualizing binaural sound by means of personal HRTFs is important for providing a better, immersive and natural listening experience [9], as well as for improving localization and perception of elevated sound sources.

Different techniques have been proposed to obtain individualized HRTFs [10]. These can be classified into three main families: acoustical measurements, anthropometric data and perceptual feedback.

[*] Corresponding author.
E-mail address: jjlopez@dcom.upv.es (J.J. Lopez).

https://doi.org/10.1016/j.apacoust.2021.108523
0003-682X/© 2021 The Author(s). Published by Elsevier Ltd.

- Acoustical measurement: It is the most straightforward method to obtain individual HRTFs [8,11]. The acoustic transfer path between the loudspeaker and two microphones inserted in the ears of the subject is captured for different discrete positions, creating a virtual sphere of measured points around the listener. It is considered the ideal solution because of its fidelity, but it is tedious as it requires long sessions of time that can cause fatigue to the subjects, as well as requiring special facilities such as anechoic chambers.
- Anthropometric data: Based on the morphology of the subject, some methods try to compute or approximate the HRTF of individuals. They usually employ optical descriptors such as 3D meshes or 2D images [12,13] and numerical methods (PCA, FEM, BEM, artificial neural networks) to calculate a solution, or structural models of the HRTF to adapt some acoustic parameters or choose a best-fit option [11]. These technologies are based on acoustic principles and study the effects of independent elements of the morphology, but usually require big amounts of data and/or high resolution images or 3D scans.
- Perceptual feedback: By means of perceptual tests, subjects can choose a suitable HRTF from a set, or modify some parameters according to their responses [14,15]. These techniques are exempt from complex installations and expensive equipment, looking for an easy-to-apply method. However, the design of the listening test is not obvious and is often time-consuming to perform, limiting its use to expert or trained listeners.

The family of acoustical measurement methods is considered the reference for all individualization techniques, as it provides the direct acoustical HRTF. The traditional acoustical measurement implementation [16] employs an anechoic chamber and a complex motorized positioning system to vary the loudspeaker-listener relative position. While human beings are able to distinguish frontal sources separated 1–2° [17], measurements are usually acquired every 5° in the horizontal plane and every 10° in the vertical plane. This leads to a measured set of HRTFs with more than 1,000 measurements, which results in long HRTF recording sessions that produce fatigue and discomfort in the subjects, as no movement is allowed to avoid measurement errors.

With the aim of overcoming the drawbacks of traditional acoustical measurement methods, techniques for reducing the total measurement time have also been actively pursued. The most straightforward is to install a set of loudspeakers in multiple positions to reduce the time required for loudspeaker positioning. Setting up a loudspeaker in each measurement point is impractical, since it would require more than a thousand units. Therefore, hybrid systems have been proposed, which use multiple loudspeakers to cover a polar angle and a single-axis positioning system to cover the other. The most common combination employs a loudspeaker array in an arc structure on the vertical plane, containing as many units as elevation angles are being measured. The rotation of either the arc around the listener or the listener on a turntable [18] presents the real asset of this method. Moreover, there exist methods for simultaneous HRTF measurement using multiple loudspeakers which save even more time, such as the multiple exponential sweep method (MESM) where the reproduced sweeps overlap in time [19,20].

Despite the improvements in the reduction of measurement time, the HRTF measurement procedure still requires anechoic conditions, as HRTFs are natively referred to free-field measurements. Conducting the process within a non-anechoic room will introduce reflections that do not really belong to the actual HRTF. Although processing of the measured HRIR can be performed to crop contributions before the first room reflection, depending on the loudspeaker-listener distance and the room geometry, the result may reduce the resolution at low frequencies, affecting the measurement of the real HRTF at frequencies where torso and shoulder reflections assist localization (Section 2.3). Therefore, ideal full-band HRTF measurements require anechoic chambers, restricting high-quality HRTF acquisition to research laboratories and creating a barrier that affects the adoption of binaural technology by the general interested public or small companies.

In this paper, a method for measuring full-band HRTFs in non-anechoic rooms based on the suppression of main room reflections is proposed and validated. In Section 2, the background and problem formulation is presented. Section 3 describes in detail the steps making up the proposed method. Section 4 presents the experimental setup used to validate the proposal. In Section 5, the results obtained at each experimental stage are discussed, comparing with the ones obtained in anechoic conditions. Finally, the extracted conclusions, as well as future work, are outlined in Section 6.

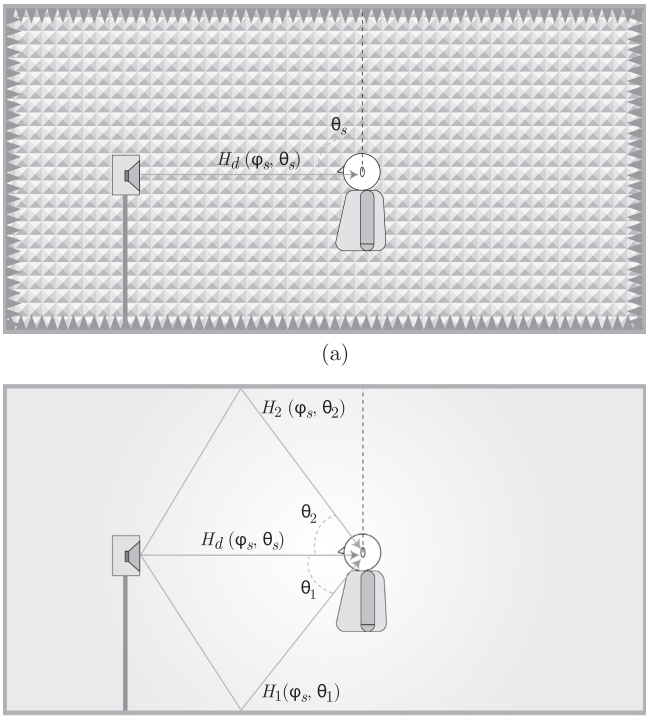
Throughout this paper, the following notation conventions will be used. Symbols with tilde, such as $\tilde{h}(t)$, denote measurements of the actual quantities, i.e. $h(t)$. Symbols with hat, such as $\hat{h}(t)$, denote estimates of the actual quantities. The superscript ^{T_0}, as in $\tilde{h}^{T_0}(t)$, denotes a signal that has been cropped to a maximum time T_0.

2. Background and problem formulation

This section presents the analytical model to explain the problem, discusses the main drawbacks appearing in non-anechoic scenarios, and explains the expected frequency limitations as a consequence of signal cropping techniques.

2.1. Signal model

Considering an ideal anechoic environment, as in Fig. 1(a), the sound pressure measured at one of the ears resulting from a sound


θ_s
$H_d\ (\varphi_s,\ \theta_s)$
$H_2\ (\varphi_s,\ \theta_2)$
θ_2
$H_d\ (\varphi_s,\ \theta_s)$
θ_1
$H_1(\varphi_s,\ \theta_1)$

(a)

(b)

Fig. 1. Anechoic and non-anechoic HRTF measurement. (a) Anechoic measurement with no reflections. (b) Non-anechoic measurement with relevant early reflections.

source signal, $s(t)$, coming from the spatial direction defined by the azimuth and elevation angles (ϕ_s, θ_s), can be expressed as:

$$y_d(t) = s(t) * h_d(t, \phi_s, \theta_s) * h_{hrir}(t, \phi_s, \theta_s), \tag{1}$$

where $*$ denotes convolution and $h_d(t, \phi_s, \theta_s)$ is the direct-path acoustic channel, which can be modeled as a simple delay when ideal transducers are used. The term $h_{hrir}(t, \phi_s, \theta_s)$ represents the head-related impulse response (HRIR) corresponding to direction (ϕ_s, θ_s).

In non-anechoic conditions, as represented in Fig. 1(b), multiple reflections coming from the different surfaces occur inside the room and the measured acoustic pressure now contains a non-desired component as:

$$y(t) = y_d(t) + y_r(t), \tag{2}$$

with

$$y_r(t) = s(t) * \left(\sum_{m=1}^{M} h_m(t, \phi_m, \theta_m) * h_{hrir}(t, \phi_m, \theta_m) \right), \tag{3}$$

where M is the total number of significant reflections and $h_m(t, \phi_m, \theta_m)$ denotes the acoustic impulse response for the acoustic path of the m-th reflection coming from direction (ϕ_m, θ_m). As a result, the measured HRIR in a non-anechoic measurement setup will include the anechoic HRIR plus the contribution of the non-desired acoustic path components:

$$\begin{aligned}\tilde{h}_{hrir}(t, \phi_s, \theta_s) = & \underbrace{h_d(t, \phi_s, \theta_s) * h_{hrir}(t, \phi_s, \theta_s)}_{\tilde{h}_{hrir(d)}(t)} \\ & + \underbrace{\sum_{m=1}^{M} h_m(t, \phi_m, \theta_m) * h_{hrir}(t, \phi_m, \theta_m)}_{\tilde{h}_{hrir(r)}(t)}.\end{aligned} \tag{4}$$

In the frequency domain, the measured HRTF translates to the addition of the anechoic HRTF plus multiple reflections coming from different directions, filtered by their corresponding traveled acoustic channel, i.e.

$$\begin{aligned}\tilde{H}_{hrtf}(\omega, \phi_s, \theta_s) = & H_d(\omega, \phi_s, \theta_s) H_{hrtf}(\omega, \phi_s, \theta_s) \\ & + \sum_{m=1}^{M} H_m(\omega, \phi_m, \theta_m) H_{hrtf}(\omega, \phi_m, \theta_m),\end{aligned} \tag{5}$$

where $H_m(\omega, \phi_m, \theta_m)$ represents the Fourier transform of the reflection acoustic path $h_m(t, \phi_m, \theta_m)$ and $H_{hrtf}(\omega, \phi, \theta)$ is the Fourier transform of the head-related impulse response $h_{hrir}(t, \phi, \theta)$.

2.2. HRIR temporal truncation or cropping

Consider a non-anechoic measurement scenario, as in Fig. 1(b). At the listening position, the arriving signal is a mixture of the direct sound and two typical significant room reflections. In this example they are mainly originated on the floor and the ceiling. The lateral and opposite walls also produce reflections, but they are not considered here for the sake of simplicity. A typical measured impulse response in such type of scenario is shown in Fig. 2, corresponding to a medium-size room with a standard ceiling height of 2.6 meters. For a source to listener distance of 1 meter and a source to ceiling distance of 2 meters, the floor reflection arrives at 4.5 ms, while the ceiling reflection arrives at 12 ms approximately. The direct signal and the two main reflections can be identified arriving after the main signal. A first attempt to reflection suppression would naturally come by cropping the measured response before the arrival of the first echo, thereby preserving just the direct path and eliminating subsequent reflections.

Let us assume that each reflected path is given by

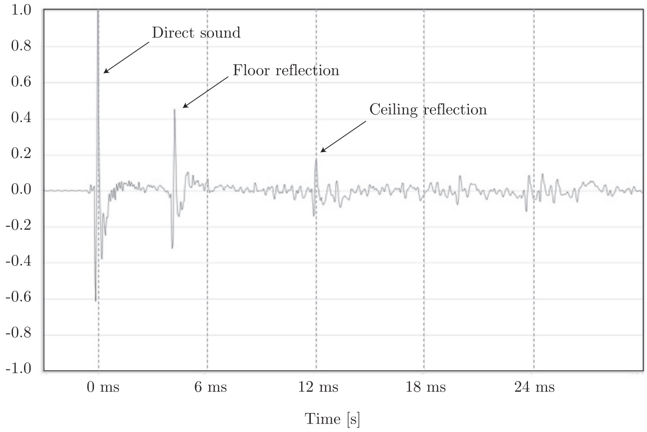


Fig. 2. Typical impulse response in non-anechoic measurement room with large floor and ceiling reflections.

$$h_m(t, \phi_m, \theta_m) \approx a_m(t) * \delta(t - \tau_m), \quad m = 1, \ldots, M. \tag{6}$$

where τ_m is the time-delay corresponding to the m-th reflection, and $a_m(t)$ encodes the filtering effect of such path. Cropping would create a new measured HRIR as follows:

$$\tilde{h}^{T_0}_{hrir}(t, \phi_s, \theta_s) = \begin{cases} \tilde{h}_{hrir}(t, \phi_s, \theta_s), & t < T_0 \\ 0, & t \geqslant T_0, \end{cases} \tag{7}$$

where T_0 is selected as the delay corresponding to the first echo, i.e. $T_0 = \min(\{\tau_m\})$, with $\{\tau_m\} = \{\tau_0, \tau_1, \ldots, \tau_M\}$.

However, cropping the signal in time with such a short window implies a significant loss in frequency resolution, which at lower bands can be considerably critical. Insufficient resolution at low frequencies implies a lack of information in spectral ranges that affect significantly localization accuracy. As an example, consider the response of Fig. 2. The first reflection arrives approximately 4.5 ms after the direct signal. Given the frequency resolution allowed by a time segment of duration T_0, i.e.

$$f_q = \frac{F_s}{N} = \frac{F_s}{F_s T_0} = \frac{1}{T_0}, \tag{8}$$

if the impulse response is cropped to a duration of $T_0 = 4.0$ ms, the resulting frequency resolution is 250 Hz, independently of the selected sampling frequency F_s. Moreover, when applying a non-rectangular smoothing window, more restricting values are expected due to the increased width of the main lobe (e.g. 500 Hz for a Hamming window, but in practice it extends up to 1 kHz as it will be seen in the next section). This windowing procedure can be employed as part of techniques for measuring loudspeaker responses [21–23], but it turns out to be incomplete for measuring HRTFs.

If only the high frequency part of the HRTF is intended to be obtained, this procedure might be sufficient. For the low frequency part, there exist methods to reconstruct a flat low frequency response maintaining interaural time differences (ITDs) [24]. However, important information of the HRTF at low frequencies, such as the torso and shoulders reflections, are not properly acquired due to the inaccurate information at frequencies below 1 kHz [25].

2.3. Other techniques for reflection removal

As presented in the previous section, HRIR cropping is the most straightforward technique to remove reflections from non-anechoic HRIR measurements. However, other solutions have been previously proposed in the literature to deal with unwanted reflections in the measured responses. This section elaborates on some existing well-known approaches for reflection removal, discussing their advantages and disadvantages through a reproducible experiment.

- Frequency-dependent cropping [26,27]: Corresponds to a frequency-dependent truncation approach that processes mid to high frequencies by adapting the cropping length in each band. However, the first window considered by this method still includes effects of the first reflections as in the simple temporal cropping technique (Section 2.2), which create comb-filter-like distortions in the response at mid and low frequencies.
- Reconstruction of low frequency: It is a well-known fact that HRTFs show a flat response along the lowest frequency band, as the human physical features do not have any impact at such low frequencies. There are methods for reconstructing the low-frequency response while preserving ITDs. The missing information could be completed by means of geometric models of head and torso [28], with results from the boundary element method [29], by cross-over filtering with an adequate low-frequency response [30] or modeling a constant magnitude and a linearly extrapolated phase [24]. However, the reconstruction is usually dependent on information from the upper neighboring band, in order to cross-fade it properly. We may usually find two situations. First, when cropped measured responses are used, the neighboring band may be too high and it is not guaranteed that frequencies below show a flat response (e.g. shoulder reflection effects). Second, if non-anechoic responses are used, the upper frequency band will contain spectral distortions caused by the reflected paths, leading to non-accurate reconstructions.
- Dereverberation: While one might think that dereverberation methods can be used to fix spectral distortions on non-anechoic HRTFs, it should be noted that reflections arrive from different spatial directions. Thus, the reflected sound will be filtered by different HRTFs corresponding to different incidence angles, as shown by the sum making up the $\tilde{h}_{hrir^{(r)}}(t)$ term in Eq. (4). Classical dereverberation methods [31–33] do not take into account the direction-dependent filtering of HRTFs, which render these methods not suitable for isolating a target HRTF from a non-anechoic measurement.

A practical case considering a common measuring setup, such as the one in Fig. 1(b), is analyzed in the following to illustrate why the methods above are not completely effective. Let us assume that the subject is two meters away from the loudspeaker, which is placed at a height of 1 meter, therefore creating a floor reflection that arrives from a direction $\theta_1 = -45^o$. For reproducibility purposes, HRTFs extracted from the KEMAR public dataset [34] measured at MIT's anechoic chamber are used. Specifically, we take for the experiment the left-ear responses corresponding to $(\phi = 0^\circ, \theta = 0^\circ)$ and $(\phi = 0^\circ, \theta = -45^\circ)$ from the so-called "full" measurement set.

According to Eqs. (4) and (6), the measured response for a single (floor) reflection in the above setup is modeled as:

$$\begin{aligned}\tilde{h}_{hrir}(t,0^\circ,0^\circ) = h_d(t,0^\circ,0^\circ) * h_{hrir}(t,0^\circ,0^\circ) + \\ a_1(t) * \delta(t-\tau_1) * h_{hrir}(t,0^\circ,-45^\circ),\end{aligned} \tag{9}$$

which can be further simplified to

$$\tilde{h}_{hrir}(t,0^\circ,0^\circ) = h_{hrir}(t,0^\circ,0^\circ) + a_1 h_{hrir}(t-\tau_{10},0^\circ,-45^\circ), \tag{10}$$

if only a relative delay, $\tau_{10} = \tau_1 - \tau_d$, and a relative scalar attenuation factor, a_1, are considered for the reflected path. Then, for such setup, we can directly simulate a measured response by adding to the target HRIR at $(\phi = 0^\circ, \theta = 0^\circ)$ a delayed and attenuated version of the HRIR at $(\phi = 0^\circ, \theta = -45^\circ)$. The temporal delay corresponding to the path difference $(0.82\ \mathrm{m})$ is $\tau_{10} = 115$ samples $(c = 343\ \mathrm{m/s})$ for F_s = 48 kHz. Also, we can consider a representative attenuation factor of $a_1 = 0.3$, which is typical for a rigid floor.

Let us now apply the described methods to evaluate their impact on the processed response at low frequencies. We apply the first 3 since, as already explained, dereverberation does not make sense in the considered context.

Fig. 3(a) shows the HRIRs used from the KEMAR public database [34] and the composite HRIR obtained according to Eq. (10), where the direct signal at 0° is mixed with the attenuated and delayed signal at -45°. The rest of Fig. 3 compares the results after the processing, where the dashed line always indicates the true (target) anechoic HRTF for $(\phi = 0^\circ, \theta = 0^\circ)$.

Fig. 3(b) compares the anechoic HRTF with the non-anechoic one including the floor reflection and without applying any processing method. Note that the anechoic HRTF shows some remarkable oscillating effects (2 or 3 dB) between 400 Hz and 1 kHz, which are produced by reflections on the shoulders [25,35]. The comb filtering effect produced by the reflection degrades significantly the response within the represented frequency range, distorting the slight shoulders oscillations and adding more pronounced deeps in the response.

Fig. 3(c) shows the result after applying temporal cropping just before the arrival of the floor reflection. Although the comb filtering effect is avoided in this case, unfortunately the shoulder reflections also disappear.

Finally, Fig. 3(d) depicts the result after applying frequency-dependent cropping. Besides the window applied in the above case, another window with two times its duration has been used to increase the resolution at low frequencies. As observed, the effects of the shoulder are neither preserved in this case and other unwanted distortions appear such as the peak at 400 Hz.

The low-frequency reconstruction method would simply take the response of Fig. 3(c) and would create a flat response below one desired frequency (e.g. 400 Hz). Clearly, that would not solve the problem since the shoulder and torso reflections would still be missing and it is not worth representing it. Nonetheless, it would have the benefit of preserving the group delay between 20 Hz and 400 Hz and the almost constant response shown within the lowest part of the graph.

With this simple example we have tried to show that these methods do not completely work in the specific case of HRTF measurements when it is desired to preserve the effects at low frequencies due to the torso and shoulders. Therefore, one of the main objectives of this work is to obtain a good estimate of the low frequency range of the HRTF by performing a proper compensation of the most significant reflections, outperforming the methods discussed above.

3. Proposed method

This section presents the general measurement and processing framework proposed in this paper for correcting HRTFs measured in non-anechoic scenarios. As already discussed, simple cropping can be applied to extract the information related to the high-frequency range of the desired response. However, additional processing is necessary to preserve low-frequency information while minimizing the effect of room reflections.

The general processing scheme of the proposed sub-band method is shown in Fig. 4. The bottom branch is aimed at processing the low-frequency part of the input HRIR signal. The effect of room reflections is cancelled based on the directional information extracted by a spherical microphone array and sound field analysis techniques. The top branch extracts the high-frequency informa-

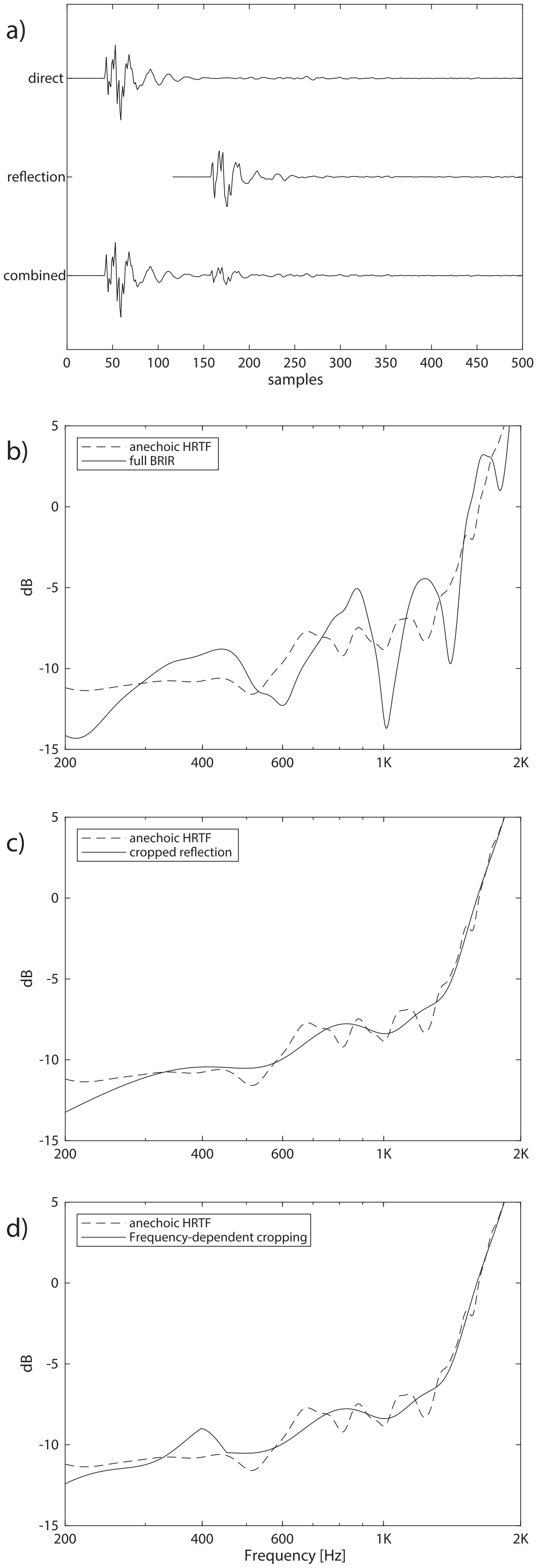


Fig. 3. Comparison of methods to overcome reflections (a) HRIRs direct, from reflection direction and combined to simulate $\tilde{h}_{hrir}(t, 0^\circ, 0^\circ)$ which includes the main reflection. (b) BRIR, considering the reflection. (c) Cropping the reflection. (d) Frequency-dependent cropping.

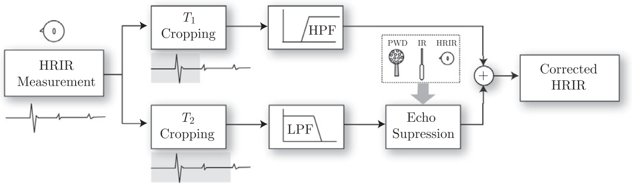


Fig. 4. Block diagram of the proposed method.

tion of the desired signal. The input HRIR is cropped before the arrival of the first reflection to keep only the direct path, leaving only the high-frequency information.

An important aspect of the proposed approach is the lowest frequency limit to be achieved, which is set at 100 Hz. For frequencies below this limit, the magnitude of the HRTF is practically flat and can be reconstructed. In addition, very low frequencies have practically no contribution to spatial localization [16,30]. A frequency limit of 100 Hz implies a time interval of 10 ms after the arrival of the direct sound (or 480 samples for a sampling frequency $F_s = 48$ kHz). Thus, reflections occurring within such time window are the ones that must be properly handled. Note that the proposed technique effectively extends the frequency range of non-anechoic HRTF measurements to cover an important part of the low-frequency response, between 100 and 500 Hz or even 1 kHz depending on the time of arrival of reflections.

The following sections present the processing and measurement steps proposed to achieve the suppression of important room reflections in non-anechoic HRTF measurements. The methodology involves the use of different sound capture techniques and loudspeaker setups. For the sake of simplicity, only the removal of the floor reflection is being considered in the following explanation. The suppression of the ceiling reflection or other main reflections can be addressed by extending the same procedure taking into account the direction of origin of the reflection. Furthermore, the case of suppression of a floor reflection is especially relevant because in non-anechoic rooms it will occur in almost any measurement scenario. Unlike the distance to the ceiling or walls, the measurement distance to the floor will always be quite similar due to human dimensions. This results in main reflections coming from the floor in a similar time range and close to the main peak of the impulse response, which makes them difficult to remove by windowing methods. On the other hand, the ceiling or walls are easier to treat with acoustically absorbent material simply because they are not walked on and the floor is.

3.1. Step 1: cropping

Let us assume the non-anechoic HRTF measurement setup depicted in Fig. 5(a), corresponding to a source direction (ϕ_s, θ_s). As an example, the frontal direction $\phi_s = 0$ and $\theta_s = 0$ is considered, obtaining the measurement $\tilde{h}_{hrir}(t, 0, 0)$, which follows the model of Eq. (4). The first and most relevant reflection affecting the measurement is that coming from the floor surface, with direction (ϕ_s, θ_1). Obviously, there are many other contributions arriving to the listener, but since they arrive considerably later, they can be discarded by cropping. The measured response can be cropped before the arrival of this first room reflection by selecting a cropping time T_1 (e. g. $T_1 = 4$ ms). However, depending on the length and type of the cropping window used, a short time results in poor frequency resolution. Unfortunately, such low resolution affects frequency bands where there are relevant directional aspects

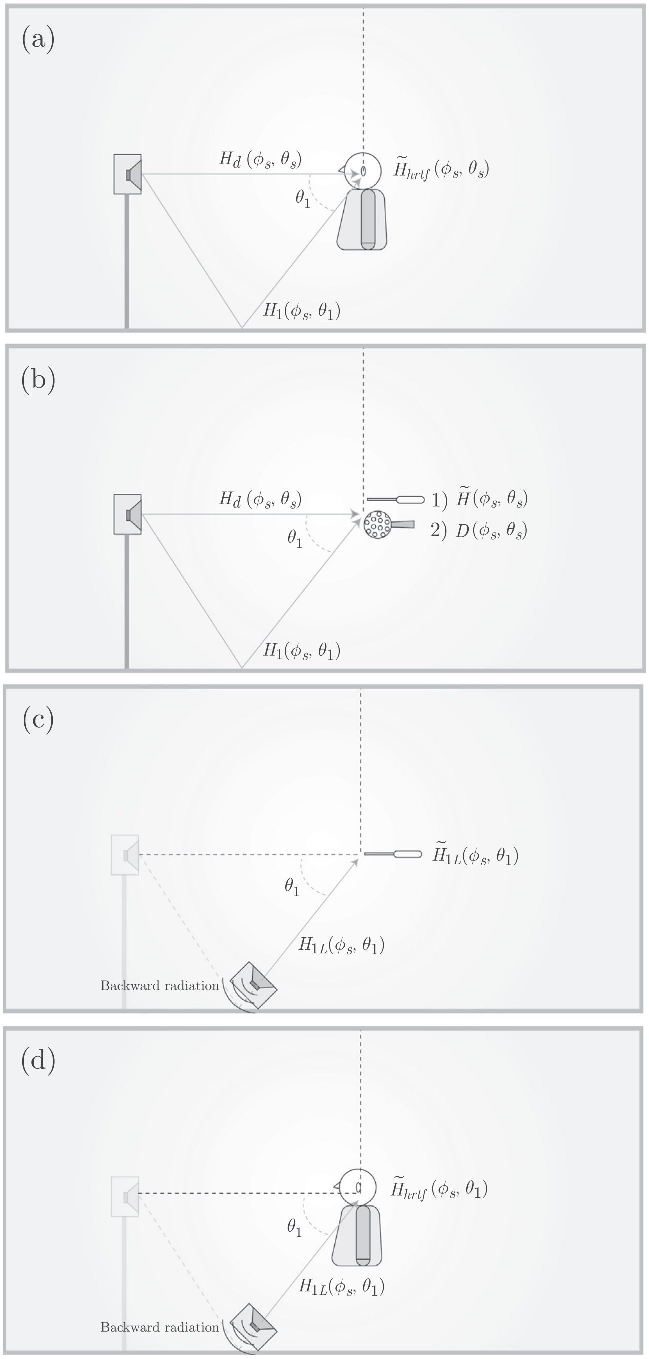

(a)
$H_d(\phi_s, \theta_s)$
$\widetilde{H}_{hrtf}(\phi_s, \theta_s)$
θ_1
$H_1(\phi_s, \theta_1)$
(b)
$H_d(\phi_s, \theta_s)$
1) $\widetilde{H}(\phi_s, \theta_s)$
2) $D(\phi_s, \theta_s)$
θ_1
$H_1(\phi_s, \theta_1)$
(c)
$\widetilde{H}_{1L}(\phi_s, \theta_1)$
θ_1
$H_{1L}(\phi_s, \theta_1)$
Backward radiation
(d)
$\widetilde{H}_{hrtf}(\phi_s, \theta_1)$
θ_1
$H_{1L}(\phi_s, \theta_1)$
Backward radiation

Fig. 5. HRTF correction procedure. (a) Non-anechoic HRTF measurement. (b) Impulse response measurement with ominidirectional microphone and acoustic channel estimation based on spherical array processing. (c) Echo path impulse response measurement. (d) Echo path HRTF measurement.

related to shoulder reflection and torso diffraction, as commented at the end of Section 2.2.

An alternative to increase the frequency resolution is to use a longer window with cropping time T_2. As an example based on Fig. 2, a cropping time $T_2 = 11$ ms would increase the frequency resolution to 90 or 180 Hz, extending the frequency range to a factor close to 1.5 octaves with respect to T_1. The cropping time T_2 is selected slightly before the arrival time of the second relevant reflection (see Step 2). Thus, the cropped HRIR can be written as

$$\begin{aligned}\tilde{h}^{T_2}_{hrir}(t,\phi_s,\theta_s) = h_d(t,\phi_s,\theta_s) * h_{hrir}(t,\phi_s,\theta_s) + \\ h_1(t,\phi_s,\theta_1) * h_{hrir}(t,\phi_s,\theta_1),\end{aligned} \tag{11}$$

leading, in the frequency domain, to the HRTF

$$\begin{aligned}\widetilde{H}^{T_2}_{hrtf}(\omega,\phi_s,\theta_s) = H_d(\omega,\phi_s,\theta_s) \cdot H_{hrtf}(\omega,\phi_s,\theta_s) + \\ H_1(\omega,\phi_s,\theta_1) \cdot H_{hrtf}(\omega,\phi_s,\theta_1).\end{aligned} \tag{12}$$

The process employed is a variation of the Frequency-Dependent Windowing (FDW) [26], also referred to as Frequency Dependent Truncation [27] for the specific case of removing reflections of impulse responses (cf. Section 2.3).

It becomes apparent that to estimate the actual frontal HRTF with increased frequency resolution, the effect from the reflection must be suppressed. This implies estimating not only the acoustic path corresponding to such reflection, but also the HRIR for its direction. The following steps address such estimation process.

3.2. Step 2: acoustic channel estimation

The second step involves two measurements, as depicted in Fig. 5(b). These are conducted as follows.

3.2.1. Impulse response measurement

First, the listener is substituted by a flat-response omnidirectional microphone to measure the impulse response between the source and the listener's position, $\tilde{h}(t,\phi_s,\theta_s)$. By inspecting this response, the cropping times T_1 and T_2 can be determined as those right before the first and second relevant reflections, respectively. Cropping the impulse response at time T_2 results in the following model for the measured signal

$$\tilde{h}^{T_2}(t,\phi_s,\theta_s) = h_d(t,\phi_s,\theta_s) + h_1(t,\phi_s,\theta_1), \tag{13}$$

or, in the frequency domain, to the transfer function

$$\widetilde{H}^{T_2}(\omega,\phi_s,\theta_s) = H_d(\omega,\phi_s,\theta_s) + H_1(\omega,\phi_s,\theta_1). \tag{14}$$

3.2.2. Spherical array measurement and PWD

A theoretical background of the Plane Wave Decomposition (PWD) and the notation employed here is described in A.

To separate the contribution of the two acoustic paths defined in the previous section, we obtain an M-channel impulse response recorded with a spherical microphone array positioned at the location of the previous omnidirectional microphone. We denote such multi-channel impulse response as $\widetilde{P}^{T_2}(\omega,\phi_j,\theta_j), j = 1,\ldots,M$, where the T_2 superscript indicates that temporal cropping is also applied to avoid reflections. The cropped array response is analyzed by means of PWD, using Eqs. (A3) and (A6), to obtain the desired responses at directions (ϕ_s,θ_s) and (ϕ_s,θ_1). The analysis would provide as result the two plane-wave components $D(\omega,\phi_s,\theta_s)$ and $D(\omega,\phi_s,\theta_1)$.

It is important to note that, while the signals obtained by PWD are related to the actual acoustic channels, they may contain some amplitude differences that are modeled here by an unknown filter $Q(\omega)$ as follows:

$$D(\omega,\phi_s,\theta_s) = Q(\omega)H_d(\omega,\phi_s,\theta_s), \tag{15}$$

$$D(\omega,\phi_s,\theta_1) = Q(\omega)H_1(\omega,\phi_s,\theta_1). \tag{16}$$

The filter $Q(\omega)$ takes into account both the frequency response effect of the PWD and that of the microphones making up the spherical array. By substituting Eqs. (15) and (16) into Eq. (14), and omitting the variables ω and ϕ_s to simplify notation, the measured response can be written as

$$\widetilde{H}^{T_2}(\theta_s) \approx \frac{1}{Q} \cdot (D(\theta_s) + D(\theta_1)), \tag{17}$$

so that the filter $Q(\omega)$ is estimated as

$$\widehat{Q} = \frac{D(\theta_s) + D(\theta_1)}{\widetilde{H}^{T_2}(\theta_s)}. \tag{18}$$

The acoustic paths can therefore be estimated as

$$\widehat{H}_d(\theta_s) = \frac{1}{\widehat{Q}} D(\theta_s) = \frac{\widetilde{H}^{T_2}(\theta_s)}{D(\theta_s) + D(\theta_1)} D(\theta_s), \tag{19}$$

$$\widehat{H}_1(\theta_1) = \frac{1}{\widehat{Q}} D(\theta_1) = \frac{\widetilde{H}^{T_2}(\theta_s)}{D(\theta_s) + D(\theta_1)} D(\theta_1). \tag{20}$$

At this point, an estimate of the main acoustic paths have been obtained, but the effect of the HRTF for the direction of the main reflection is still completely unknown. The third step addresses such issue.

The measurements should ideally be performed using long enough sweep signals in order to increase the resulting signal-to-noise ratio [19]. This avoids estimation problems at low frequencies, allowing to obtain a good plane wave decomposition and avoiding problems in the denominator of the Eqs. (19, 20).

3.3. Step 3: Reflection Path Measurements and Suppression

In the third step, the acoustic transfer function is again measured with the omnidirectional microphone, but this time using a loudspeaker placed on the floor and oriented towards the direction of the reflection θ_1, as shown in Fig. 5(c). It is important to note that the measured acoustic channel, denoted as $\tilde{h}_{1L}(t, \phi_s, \theta_1)$, includes the bass enhancement effect resulting from the floor placement of the loudspeaker. Indeed, due to the speaker boundary interference response [36], the loudspeaker behavior in Fig. 5(b) for frontal radiation is different from the one observed when the loudspeaker is placed on the floor. Note that, due to the orientation of the loudspeaker, the reflections will arrive later in time than T_2. Then, as with $\tilde{h}^{T_2}(t, \phi_s, \theta_s)$, this new response is also cropped with the same window length to discard reflections, resulting in $\tilde{h}^{T_2}_{1L}(t, \phi_s, \theta_1)$.

Subsequently, the microphone is substituted with the subject, measuring and cropping the HRIR from the same direction (see Fig. 5(d)). The resulting HRTF can be written as:

$$\widetilde{H}^{T_2}_{hrtf}(\omega, \phi_s, \theta_1) = H_{hrtf}(\omega, \phi_s, \theta_1) \cdot H_{1L}(\omega, \phi_s, \theta_1), \tag{21}$$

where $H_{1L}(\theta_1)$ is the frequency-domain equivalent of $h_{1L}(\theta_1)$, already known from the microphone measurement. Thus, the HRTF for the echo direction can be estimated as:

$$\widehat{H}_{hrtf}(\omega, \phi_s, \theta_1) = \frac{\widetilde{H}^{T_2}_{hrtf}(\omega, \phi_s, \theta_1)}{\widetilde{H}^{T_2}_{1L}(\omega, \phi_s, \theta_1)}. \tag{22}$$

Finally, according to Eq. (14), an estimate of the echo-free HRTF for the desired source direction can be obtained as:

$$\widehat{H}_{hrtf}(\theta_s) = \frac{\widetilde{H}^{T_2}_{hrtf}(\theta_s) - \widehat{H}_1(\theta_1)\widehat{H}_{hrtf}(\theta_1)}{\widehat{H}_d(\theta_s)}, \tag{23}$$

where the azimuth angle ϕ_s and frequency ω have been again omitted for notation simplicity.

4. Experimental setup

This section introduces the real experiments that have been carried out to validate the proposed reflection suppression method. First, the experimental setup is described in detail, providing as well practical considerations related to the implementation of the proposed method. To analyze the performance, a measured and corrected HRTF is compared to the corresponding HRTF obtained in an anechoic chamber using the same measuring equipment.

4.1. Array setup and measurement room

The setup employed to evaluate the proposed method is based on a set of loudspeaker arrays at different heights. The system has been designed to perform fast HRTF measurements by avoiding the usual traditional drawbacks, bringing about a reduction of measurement time, hardware complexity and cost.

The system, previously introduced in [37,38] is shown in Fig. 6. It is composed by a 72-loudspeaker circular array of 2 meters of radius at an elevation of 0° with a resolution of 5° in azimuth. Two additional 8-loudspeaker circular arrays of 1 meter of radius are placed at the ceiling and on the floor, with a lower resolution of 45° and oriented towards the listener position at 45° and −45°, respectively. The loudspeakers used in all the arrays are all self-powered M-Audio model BX5 D2 audio monitors. The system was installed inside a non-anechoic low reverberant room with minimum acoustic conditioning. Table 1 summarizes the main characteristics of the room.

The described configuration allows to perform HRTF measurements for three different planes. Note that, although three elevation planes do not provide a dense spatial sampling, the measurement setup is highly simplified and interpolation techniques can be used to extend the amount of measurements in the median plane [39,40]. In any case, the number of loudspeakers can be extended if necessary. Nonetheless, an important feature is that the elevated arrays are strategically placed to be at the approximate point of reflection

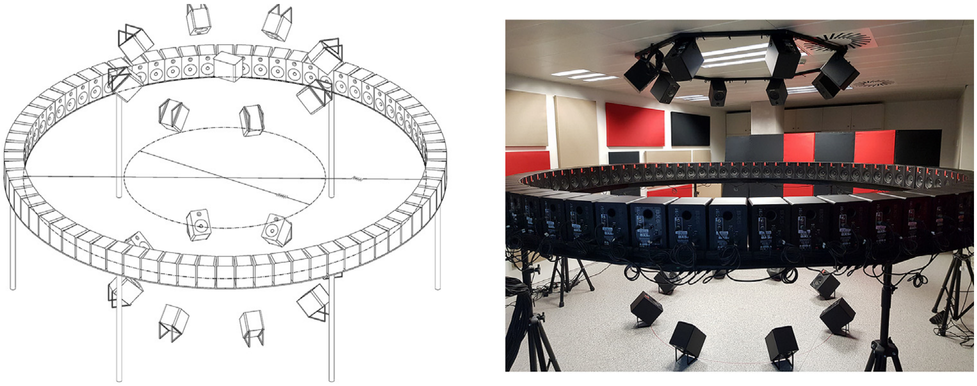


Fig. 6. Measurement setup. On the left, elevation view of the 3D model of the system. On the right, deployed setup inside room.

Table 1
Characteristics of the experimental room.

Dimensions	D 9.88 × W 4.80 × H 2.65 m
Volume	128 m^3
Surface Materials	Walls: 70% conditioned Floor: vinyl flooring (reflective) Ceiling: low quality acoustic panels
RT_{60}	169 ms
Number of loudspeakers	88

on those surfaces, which will be useful in practice for applying the proposed echo suppression technique. Thus, the proposed setup not only allows to avoid costly and complex positioning systems, but also facilitates the proposed measurement process to obtain an extended low-frequency response.

4.2. Microphones

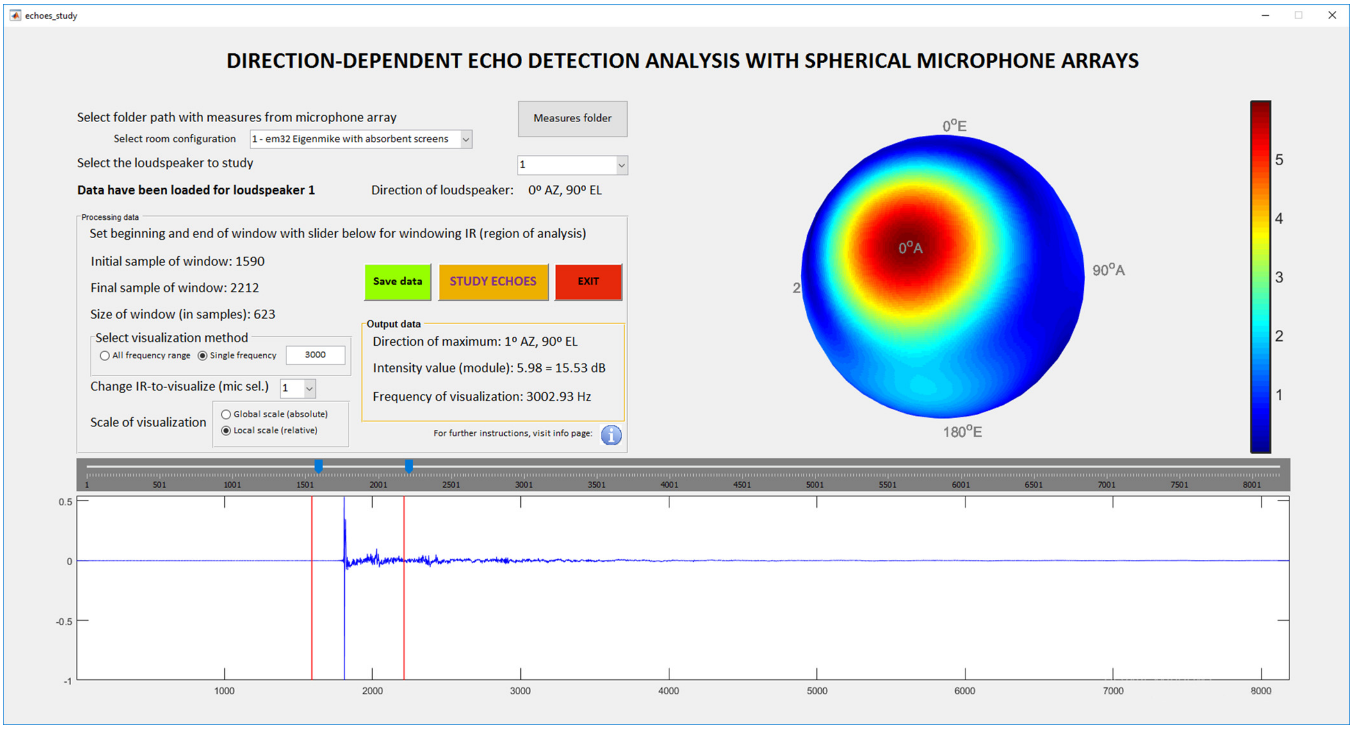
The Earthworks M30 [41] microphone was used to measure the required omnidirectional impulse responses at the different stages of the method. The spherical array recordings were captured by an em32 Eigenmike from mh acoustics [42], composed by 32 individual electret capsules inserted on a rigid sphere of radius 4.2 cm, following the shape of a pentakis dodecahedron with the transducers placed at its vertices. The HRIR measurements of this validation experiment were acquired with the Brüel & kJær Head And Torso Simulator (HATS) 4100 dummy head [43]. The multiple exponential sweep method (MESM) [19] was used, with measurements taking 2 min and 28 s for all the 88 loudspeakers. All the recordings considered a sampling frequency of $F_s = 48$ kHz.

4.3. PWD processing

Finally, the PWD processing was performed by means of the open-source SOFiA MATLAB library [44]. The Eigenmike array allows a spherical harmonic decomposition up to the order $N = 4$. The reflection from the back and lateral walls are not considered in the low-frequency processing, since they are highly absorbed by acoustic materials placed on these surfaces. The reflection on the ceiling can be handled following the same procedure as the one on the floor but using the elevated loudspeaker array. In order to make easier the whole measuring process, which includes the identification and location of the main reflections in the measured response, a specific software with a GUI was designed, as shown in Fig. 7. The GUI allowed for a convenient visual inspection and cropping of the responses to properly set the times T_1 and T_2, and supports a PWD analysis that facilitates the identification of the direction of arrival corresponding to the main reflections occurring within the room.

5. Measurements and results

This section presents the results obtained at the different stages of the proposed method on a real validation case. In order to maintain the coherence of the explanation, an HRTF measurement in which a main reflection from the floor will be suppressed has been chosen as a real example. In the case of real measurements in another room and/or HRTFs with different source directions, other main reflections coming from different surfaces could be generated. In such cases the same method would have to be applied but taking into account the HRTF of the measured direction and obtaining the corresponding main reflection by PWD. Depending on the case, this may require different loudspeaker setups and the casuistry is infinite depending on the characteristics of the measurement room. In the experiment presented below, the selected direction was the frontal ($\phi_s = 0^\circ, \theta_s = 0^\circ$). The processing dealing with echo suppression will consider the frequency range between 100 and 1000 Hz. As already described, frequencies above 1000 Hz can be reliably measured by cropping the measured HRIRs. On the other hand, the magnitude response of HRTFs below 100 Hz is completely flat, and its phase can be easily reconstructed by ensuring that the ITD is properly preserved at low frequencies.



echoes_study

DIRECTION-DEPENDENT ECHO DETECTION ANALYSIS WITH SPHERICAL MICROPHONE ARRAYS

Select folder path with measures from microphone array

Select room configuration: 1 - em32 Eigenmike with absorbent screens

Select the loudspeaker to study: 1

Data have been loaded for loudspeaker 1

Direction of loudspeaker: 0º AZ, 90º EL

Processing data

Set beginning and end of window with slider below for windowing IR (region of analysis)

Initial sample of window: 1590

Final sample of window: 2212

Size of window (in samples): 623

Select visualization method: ○ All frequency range ● Single frequency 3000

Change IR-to-visualize (mic sel.): 1

Scale of visualization: ○ Global scale (absolute) ● Local scale (relative)

Output data

Direction of maximum: 1º AZ, 90º EL

Intensity value (module): 5.98 = 15.53 dB

Frequency of visualization: 3002.93 Hz

For further instructions, visit info page:

Fig. 7. Software GUI used for the spatial and temporal analysis of the measured responses.

5.1. HRIR and impulse response cropping

As a first step, both the HRIR to be compensated and the room impulse response, are measured with the dummy-head device and the flat-response microphone, respectively. Thus, the signals $\tilde{h}_{hrir}(t, 0^\circ, 0^\circ)$ and $\hat{h}(t, 0^\circ, 0^\circ)$ are acquired. Since all the measurements correspond to the same azimuth $\phi_s = 0^\circ$, the azimuth angle will be omitted for simplicity in what follows. Given the specific geometry of the measurement setup, the reflection from the floor arrives approximately at 4.4 ms (211 samples after the direct sound, at $F_s = 48$ kHz), leading to a low-frequency limit between $1/0.0044 = 227$ Hz to 454 Hz depending on the windowing. As described in Section 3.1, T_1 is fixed to 200 samples in order to avoid this reflection. The acquired impulse response is shown in Fig. 8, where such reflection can be clearly identified, as well as the second one. Therefore, T_2 is fixed just before this second reflection at 420 samples. Note that the longer window T_2 contains the effect of the first reflection that must be suppressed with the proposed compensation method. The windows used to crop the measured temporal responses are also depicted in Fig. 8, which have the shape of a soft decay rectangular window having a Hann profile.

5.2. Acoustic channel estimation

In the next step, the multichannel impulse response at the same location is measured with the Eigenmike microphone array, extracting by means of PWD the signals $D(\omega, 0^\circ)$ and $D(\omega, -45^\circ)$. Both are represented in Fig. 9 for the frequency range of interest. The result after performing the equalization operation of Eqs. (19) and (20) are also represented as $\widehat{H}_d(0^\circ)$ and $\widehat{H}_1(-45^\circ)$. The sum of both corrected responses $\left(\widehat{H}_d(0^\circ) + \widehat{H}_1(-45^\circ)\right)$ is shown to match perfectly the original frequency response measured with the microphone, $\widetilde{H}^{T_2}(0^\circ)$, which includes the combination of the two paths. The need for such equalization is also demonstrated by showing the addition of the two PWD signals

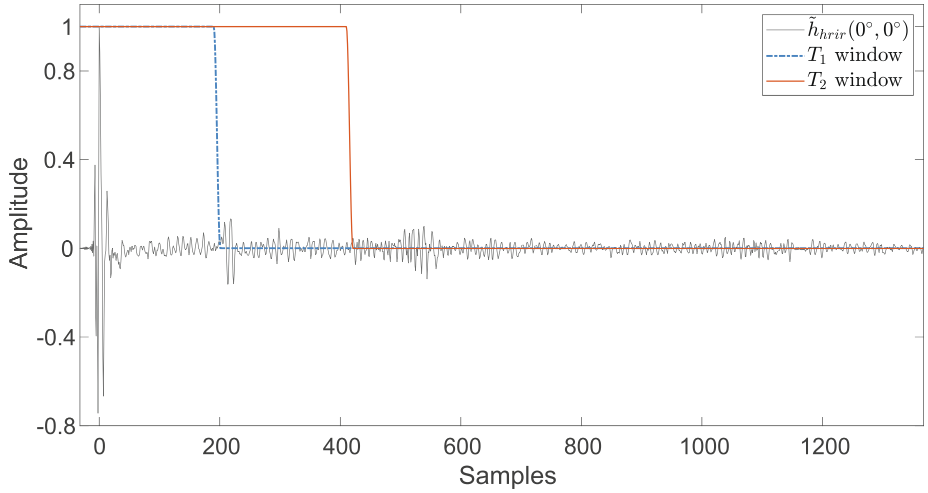


Fig. 8. Measured impulse response from the frontal direction, $\tilde{h}(t, 0^\circ, 0^\circ)$, and selected cropping windows.

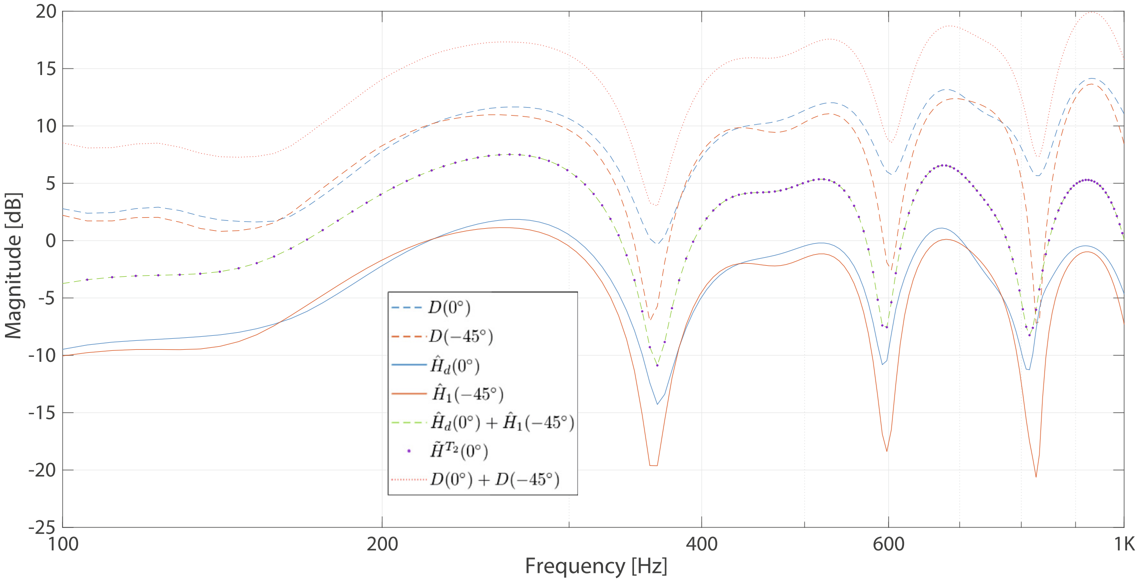


Fig. 9. PWD Components, $D(0^\circ)$ and $D(-45^\circ)$, and estimated acoustic channels, $\widehat{H}_d(0^\circ)$ and $\widehat{H}_1(-45^\circ)$. The addition of the two estimated acoustic paths is shown to match the measured omnidirectional response $\widetilde{H}^{T_2}(0^\circ)$.

$(D(0^\circ) + D(-45^\circ))$, which results in a magnitude difference between 10 and 12 dB due to factors such as the frequency response of the array capsules and other effects derived from the PWD processing.

5.3. Reflection path measurements and compensation

For the next measurement step, the HRTF for the reflection direction must be properly estimated. To this end, a loudspeaker having the same azimuth angle, $\phi_s = 0^\circ$, but from the lower loudspeaker array, is selected. This loudspeaker has the elevation angle of the reflection $(\theta_s = -45^\circ)$ and points towards the listener position. The response is measured both with the flat-response microphone and with the dummy-head device, resulting in the signals $\tilde{h}_{hrir}(t, -45^\circ)$ and $\tilde{h}_{1L}(t, -45^\circ)$, respectively. Due to the back radiation of the loudspeaker on the floor, the original room transfer function presents a relevant boost in the range 100–250 Hz. The compensation is performed by applying Eq. (22). To this end, a regularized inverse filter [45] with $\beta = 0.05$ is considered. Fig. 10 shows the measured $\widetilde{H}_{hrtf}(-45^\circ)$, the inverse filter $\widetilde{H}_{1L}(-45^\circ)^{-1}$ and the compensated version $\widehat{H}_{hrtf}(-45^\circ)$, in the frequency domain for the frequency range of interest. As observed, the result is the cancellation of the low-frequency boost at the frequencies of interest, with a reduction in magnitude close to 5 dB.

5.4. Final HRTF estimation

The last step corresponds to the final estimation of the reflection-free HRTF for the frontal direction, as given by Eq. (23). Following our experimental setup, the compensated low-frequency response obtained from the T_2 cropping window is crossed-over at 1 kHz with the high frequency response obtained by the T_1 cropping window. In this way, after processing low and high frequencies separately and adding them back together, we obtained the final processed and compensated HRTF, free from reflections. For comparison and checking purpose, an anechoic chamber measurement was acquired under the replication of the measurement conditions of the non-anechoic setup, as shown in Fig. 11. The same source-to-listener distance and same measurement equipment was considered in this checking anechoic measurement in order to avoid changes due to different loudspeaker responses. The final compensated HRIR $\hat{h}_{hrtf}(0^\circ, 0^\circ)$ and the equivalent anechoic HRIR measurement $h_{hrtf}(0^\circ, 0^\circ)$ are shown together in Fig. 12 to evaluate the temporal error.

5.5. Discussion

In order to properly interpret the results and be able to compare the performance of the method, different frequency responses are

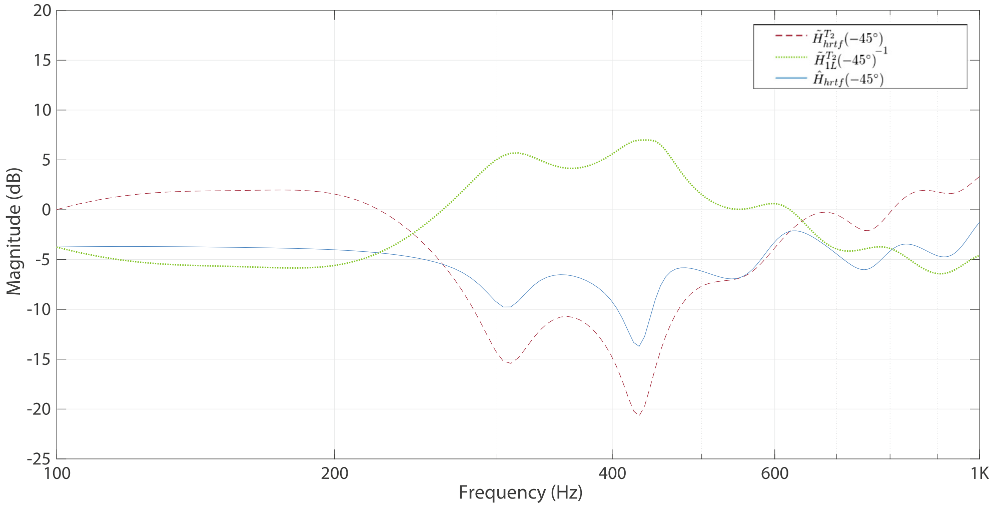


Fig. 10. Measured $\left(\widetilde{H}_{hrtf}(-45^\circ)\right)$ and compensated $\left(\widehat{H}_{hrtf}(-45^\circ)\right)$ HRTFs from the echo direction, together with the inverse of the measured transfer function $\widetilde{H}_{1L}(-45^\circ)$.

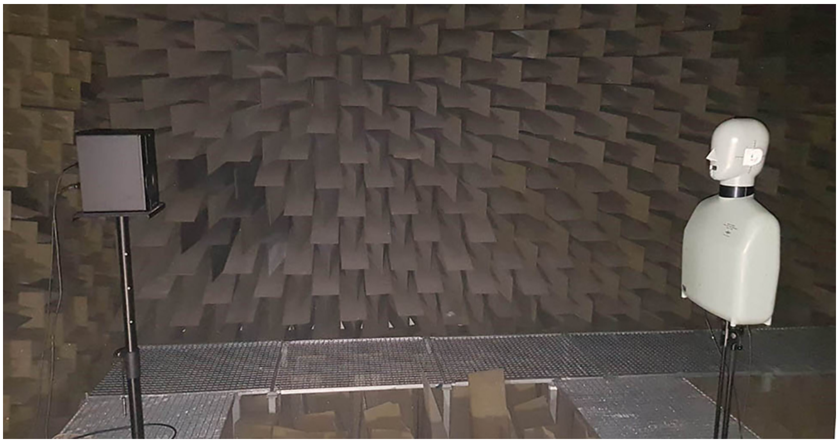


Fig. 11. Measurements in the anechoic chamber.

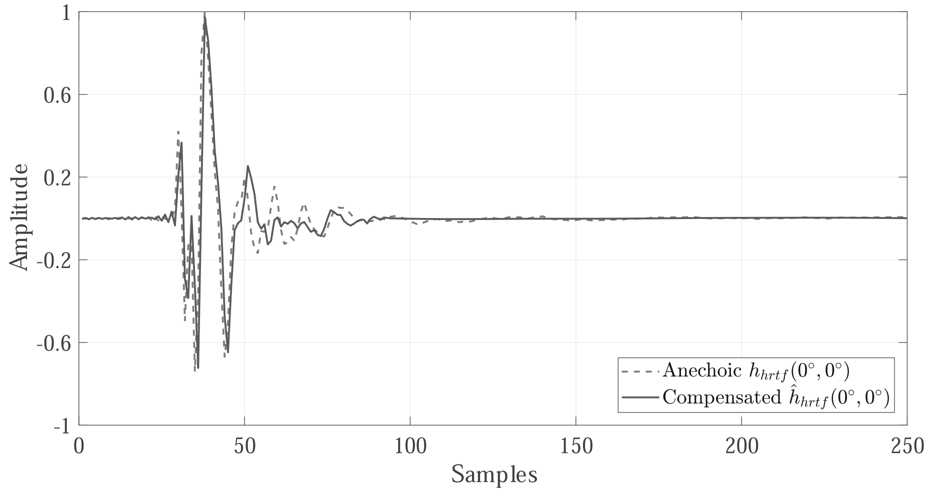


Fig. 12. Comparison between the compensated HRIR and the corresponding measurement in anechoic chamber.

represented in Fig. 13. Firstly, the HRTF measured in anechoic chamber $H_{hrtf}(0^\circ,0^\circ)$, is shown as a reference and target. In addition, the figure also shows the non-compensated HRTFs obtained in non-anechoic conditions using the cropping windows T_1 and T_2, i.e. $\widetilde{H}^{T_1}_{hrtf}(0^\circ,0^\circ)$ and $\widetilde{H}^{T_2}_{hrtf}(0^\circ,0^\circ)$. Finally, the HRTF estimated with the proposed compensation method, $\widehat{H}_{hrtf}(0^\circ,0^\circ)$, is represented.

The responses corresponding to all the non-anechoic curves overlap in the high-frequency range (above 1 kHz), as expected from the sub-band processing scheme. In the low-frequency range, the two non-compensated responses show undesired effects. The response obtained from the longer window T_2 presents a remarkable comb filtering effect resulting from the main reflection on the floor. This effect is highly mitigated in the response corresponding to the shorter T_1 windowing, although other inaccuracies appear derived from the loss in frequency resolution and, to a less extent, from other effects related to the neighboring loudspeakers of the array. In contrast, the HRTF obtained with the proposed compensation method is free from the low-frequency ripples produced by the room reflections and shows a response that is closer to the one measured in anechoic chamber.

The most notable differences between the compensated signal and the anechoic one appear in the 100–150 Hz band, which are probably due to the effects caused by the T_2 cropping window. However, above this frequency, the results are very well aligned with the objective of the proposed method, which aimed at the full-band suppression of room reflections in non-anechoic HRTFs.

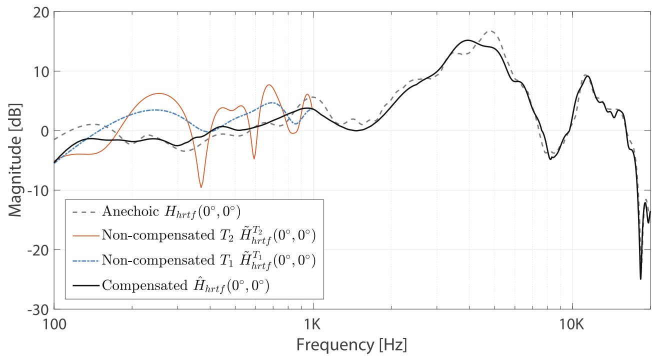


Fig. 13. Comparison between the anechoic, non-compensated and compensated HRTFs.

Table 2
Average error per 1/2 octave band for the non-compensated measurement and the proposed method with respect to the anechoic measurement.

ISO 1/2 octave band	Non-compens T_2	Non-compens T_1	Compensated
125–175 Hz	3.88 dB	1.36 dB	2.21 dB
175–200 Hz	3.13 dB	3.84 dB	0.47 dB
200–250 Hz	6.50 dB	4.44 dB	0.44 dB
250–350 Hz	6.45 dB	5.16 dB	0.77 dB
350–500 Hz	3.80 dB	1.81 dB	1.18 dB
500–700 Hz	4.20 dB	3.10 dB	0.42 dB
700–1000 Hz	2.37 dB	2.28 dB	1.09 dB

This can be more clearly observed in Table 2, which shows a more detailed evaluation of the error. It contains the average error with respect to the anechoic response for the estimated and non-compensated signals in half octave bands (according to ISO standard frequencies). The error has been computed by averaging the absolute value of the error in dB for each frequency bin within each band, i.e.

$$\text{Error}_i[\text{dB}] = \frac{1}{N_i}\sum_{\omega_k \in B_i} |\overline{H}_{hrtf}(\omega_k)[\text{dB}] - H_{hrtf}(\omega_k)[\text{dB}]|, \tag{24}$$

where N_i denotes the number of frequency bins in band B_i and $\overline{H}_{hrtf}(\omega_k)$ the response under evaluation at the k-th frequency bin. For the compensated version, the error is always below or around 1 dB, except for the first band due to the reason previously commented. However, for the non-compensated version T_1, the error is higher than 5 dB for one band and is around 3 or 4 dB for other two bands. An important result is that of the 500–700 Hz band, which contains the effect of shoulder reflections. The numerical analysis of the measurements confirms both the validity of the proposed method and its advantages.

Finally, it is worth commenting on an interesting aspect that, although not related to the low-frequency correction of the response, is important to take into account. By looking at the HRTF results, some differences are observed around 4–5 kHz. There is a level mismatch between the anechoic signal and the estimated HRTF. Such undesired effect is thought to be related to the acoustic diffraction caused by loudspeaker edges. In fact, when measuring inside the non-anechoic room, speakers followed a circular array arrangement, working as an infinite screen. As a result, diffraction is minimized and it might occur at a much lower amplitude. On the other hand, the loudspeaker in the anechoic chamber was set alone, producing wave diffraction on the edges. As a result, the frontal radiation due to this effect appears as peaks in the temporal response after the arrival of the direct sound. Note that the array-based measurement setup improves this issue, obtaining an even clearer HRTF.

6. Conclusion

This paper presents a measurement methodology and processing method that allows to obtain quasi-anechoic HRTFs in non-anechoic measurement setups. The method is based on a sub-band approach that treats separately the low-frequency and high-frequency ranges resulting from different croppings of the measured response. The low-frequency part of the estimated HRTF is obtained by means of several measurement and processing steps that involve acquiring the HRIR, the room impulse response and plane-wave decomposition (PWD) signals extracted from a spherical array. Similarly, the method also needs to estimate and compensate HRTFs measured from the most prominent room reflections. The model formulation for the proposed method also includes additional equalization procedures, which are aimed at cancelling the effect of the different transducer and array responses. The different steps of the method have been validated by a case study that eliminates the main floor reflection from a real HRTF measurement obtained in a non-anechoic room. The results of the proposed method have been shown to be completely comparable to the ones obtained in anechoic conditions, with estimated signals that are free from reflections and preserve the shape of the anechoic HRTF, i.e. without altering important localization cues.

The final aim of this work is to make out of this novel system a potential alternative to traditional anechoic methods. The method facilitates the acquisition of HRTFs by avoiding the need for very complex setups and long measurement times. Indeed, measurements over real subjects have already been performed for spatial sound personalization purposes. However, there is still work to carry out to further refine and improve the current system. Additional loudspeaker arrays at intermediate elevation positions will be set to increase resolution in the measurement process. Alternatively, the development of intelligent interpolation and correction methods based on deep neural networks will be considered to relax the hardware needs of the system. To this end, a large HRTF dataset extracted with our proposed method will be created with the aim of bringing personalized spatial audio closer to the general public.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has received funding from the Spanish Ministry of Science, Innovation and Universities, through projects RTI2018-097045-B-C21 and RTI2018-097045-B-C22, and Generalitat Valenciana under the AICO/2020/154 project grant.

Appendix A. Plane wave decomposition (PWD)

A given sound field can be decomposed into its plane wave components according to the principle of superposition. Assuming a continuous pressure distribution $P(\omega,\phi,\theta,r_0)$ on an open sphere, and its corresponding spatial Fourier coefficients, $\mathring{P}_{nm}(\omega,r_0)$, the PWD would return the plane wave components D for a specific spatial decomposition direction and angular frequency $(\omega,\phi_\mathrm{d},\theta_\mathrm{d})$:

$$D(\omega,\phi_\mathrm{d},\theta_\mathrm{d},)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{1}{i^n j_n\left(\frac{\omega}{c}r_0\right)}\mathring{P}_{nm}(\omega,r_0)Y_n^m(\phi_\mathrm{d},\theta_\mathrm{d}). \tag{A1}$$

The angular functions $Y_n^m(\phi,\theta)$ are referred to as spherical harmonics:

$$Y_n^m(\phi,\theta)=\sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}}P_n^m(\cos\theta)e^{im\phi}, \tag{A2}$$

where $P_n^m(\cos\theta)$ are Legendre functions of the first kind of order n and mode m. The radial part in Eq. (A1) depending on ω and r_0 is written in terms of j_n, which is the nth-order spherical Bessel function of the first kind.

In practice, the pressure distribution on the sphere is sampled at a limited amount of discrete spatial sampling nodes. As a consequence, discrete sampling schemes resolve spherical harmonics up to a maximum order N. While ideal PWD for $N\to\infty$ corresponds to a spatial Dirac pulse, order truncation results in a widened main lobe and additional side-lobes, decreasing substantially spatial resolution. In fact, practical signal processing applications limit the maximum order N, so that $N<\frac{\omega}{c}r_0$ to reduce aliasing contributions arising from discrete spatial sampling. For M discrete microphone positions defined by a quadrature grid on the sphere, the spatial Fourier coefficients in the spherical wave domain are given by the summation:

$$\mathring{P}_{nm}(\omega,r_0)=\sum_{j=1}^{M}\beta_j P(\omega,\phi_j,\theta_j,r_0)Y_n^m(\theta_j,\phi_j)^*, \tag{A3}$$

where β_j are weighting factors that account for the selected spatial grid.

A.1. Radial filters

Radial filters compensate for the radial portion of the Helmholtz equation, scaling the amplification gain of spherical harmonic modes. Radial filters depend on the sphere configuration, which describes whether sensor nodes on the sphere are in free field or mounted on a rigid body. For an open measurement sphere with pressure transducers, as the radial part in Eq. (A1), the radial filters are directly given by

$$d_n\left(\frac{\omega}{c}r_0\right)=\left[4\pi i^n j_n\left(\frac{\omega}{c}r_0\right)\right]^{-1}. \tag{A4}$$

For pressure transducers mounted on a rigid sphere, radial filters take the following form:

$$d_n\left(\frac{\omega}{c}r_0\right)=\left[4\pi i^n\left(j_n\left(\frac{\omega}{c}r_0\right)-\frac{j_n'\left(\frac{\omega}{c}r_0\right)}{h_n'^{(2)}\left(\frac{\omega}{c}r_0\right)}h_n^{(2)}\left(\frac{\omega}{c}r_0\right)\right)\right]^{-1}, \tag{A5}$$

where $h_n^{(2)}$ denotes the spherical Hankel function of the second kind.

In a practical scenario, radial filters are not assumed to cover the entire frequency range. The modal amplification demanded by such filters is too high at low $(\frac{\omega}{c}r_0)$, leading to unstable array responses at lower frequencies due to noise amplification. Thus, the amplification of higher modes is limited in practice to a reasonable value, although the limiting operation results in a loss of spatial resolution at lower frequencies. Taking into account both the limited order imposed by discrete spatial sampling and the effect of radial filters, the response signal for a frequency/direction $(\omega,\phi_\mathrm{d},\theta_\mathrm{d})$ is obtained by

$$D(\omega,\phi_\mathrm{d},\theta_\mathrm{d})=4\pi\sum_{n=0}^{N}\sum_{m=-n}^{n}d_n\left(\frac{\omega}{c}r_0\right)\mathring{P}_{nm}(\omega,r_0)Y_n^m(\phi_\mathrm{d},\theta_\mathrm{d}). \tag{A6}$$

Despite the use of non-critical radial filters have an impact on the effective operational bandwidth of spherical arrays and the ideal constant directivity PWD response is distorted for very low frequencies, such limitation is not really relevant for the application at hand.

References

[1] Serafin S, Geronazzo M, Erkut C, Nilsson NC, Nordahl R. Sonic interactions in virtual reality: state of the art, current challenges, and future directions. IEEE Comput Graph Appl 2018;38(2):31–43. https://doi.org/10.1109/MCG.2018.193142628.

[2] Dolby Laboratories Inc., Dolby Atmos Music—Immersive audio experiences that move you (Accessed: 2021-09-27). url: https://www.dolby.com/music/.

[3] Sony Corporation, 360 Reality Audio—So Immersive. So Real. (Accessed: 2021-09-27). url: https://electronics.sony.com/360-reality-audio.

[4] Global Market Insights Inc., Earphones And Headphones Market Size By Technology, By Application, Industry Analysis Report, Regional Outlook, Growth Potential, Price Trends, Competitive Market Share & Forecast, 2017–2024, Tech. rep., Global Market Insights Inc., Ocean View, USA; 2017.

[5] Blauert J (Ed.). The technology of binaural listening, Springer; 2013.https://doi.org/10.1007/978-3-642-37762-4.

[6] Geronazzo M, Kleimola J, Sikstroöm E, de Götzen A, Serafi S, Avanzini F. HOBA-VR: HRTF On Demand for Binaural Audio in immersive virtual reality environments. In: Audio Engineering Society 144th Convention, Milan, Italy; 2018. p. e-Brief 433.

[7] Roginska A, Geluso P. Immersive sound: the art and science of binaural and multi-channel audio. Focal Press; 2017.

[8] Møller H, Sorensen MF, Hammershoi D, Jensen CB. Head-related transfer functions of human subjects. J Audio Eng Soc 1995;43(5):300–21.

[9] Nicol R. Binaural Technology. New York: Audio Engineering Society Inc.; 2010.

[10] Sunder K, He Jianjun, Tan Ee Leng, Gan Woon-Seng. Natural Sound Rendering for Headphones: Integration of signal processing techniques. IEEE Signal Process Mag 2015;32(2):100–13. https://doi.org/10.1109/MSP.2014.2372062.

[11] Xu S, Li Z, Salvendy G. Individualization of head-related transfer function for three-dimensional virtual auditory display: a review. In: 12th International Conference on Human-Computer Interaction (HCI International 2007), vol. 4563; 2007. p. 397–407.https://doi.org/10.1007/978-3-540-73335-5_44.

[12] Katz BFG. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. J Acoust Soc Am 2001;110 (5):2440–8. https://doi.org/10.1121/1.1412440.

[13] Torres-Gallegos EA, Orduña-Bustamante F, Arámbula-Cosío F. Personalization of head-related transfer functions (HRTF) based on automatic photo-anthropometry and inference from a database. Appl Acoust 2015;97:84–95. https://doi.org/10.1016/j.apacoust.2015.04.009.

[14] Poirier-Quinot D, Katz BFG. Assessing the impact of head-related transfer function individualization on task performance: case of a virtual reality shooter game. J Audio Eng Soc 2020;68(4):248–60. https://doi.org/10.17743/jaes.2020.0004.

[15] Pelzer R, Dinakaran M, Brinkmann F, Lepa S, Grosche P, Weinzierl S. Head-related transfer function recommendation based on perceptual similarities and anthropometric features. J Acoust Soc Am 2020;148(6):3809–17. https://doi.org/10.1121/10.0002884.

[16] Blauert J. Spatial hearing: the psychophysics of human sound localization. 2nd Edition. Cambridge: MIT Press; 1997.

[17] Makous JC, Middlebrooks JC. Two-dimensional sound localization by human listeners. J Acoust Soc Am 1990;87(5):2188–200. https://doi.org/10.1121/1.399186.

[18] Richter JG, Fels J. Evaluation of localization accuracy of static sources using HRTFs from a fast measurement system. Acta Acustica united with Acustica 2016;102(4):763–71. https://doi.org/10.3813/AAA.918992.

[19] Majdak P, Balazs P, Laback B. Multiple exponential sweep method for fast measurement of Head-Related Transfer Functions. J Audio Eng Soc 2007;55(7–8):623–36.

[20] Dietrich P, Masiero B, Vorländer M. On the optimization of the multiple exponential sweep method. J Audio Eng Soc 2013;61(3):113–24.

[21] Struck CJ, Temme SF. Simulated free field measurements. J Audio Eng Soc 1994;42(6):467–82.

[22] Vanderkooy J, Lipshitz SP. Can One Perform Quasi-anechoic Loudspeaker Measurements in Normal Rooms? In: Audio Engineering Society 125th Convention, San Francisco, USA. p. 7525.

[23] Bellmann C, Klippel W. Fast Loudspeaker Measurement in Non-Anechoic Environment. In: Audio Engineering Society 143rd Convention, New York, NY, USA. p. 9825.

[24] Xie B. On the low frequency characteristics of head-related transfer functions. Chin J Acoust 2009;28(2):116–28.

[25] Algazi VR, Avendano C, Duda RO. Elevation localization and head-related transfer function analysis at low frequencies. J Acoust Soc Am 2001;109 (3):1110–22. https://doi.org/10.1121/1.1349185.

[26] Karjalainen M, Paatero T. Frequency-dependent signal windowing. In: IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE, New Paltz, New York, USA; 2001. p. 35–38.https://doi.org/10.1109/aspaa.2001.969536.

[27] Denk F, Kollmeier B, Ewert SD. Removing reflections in semianechoic impulse responses by frequency-dependent truncation. J Audio Eng Soc 2018;66 (3):146–53. https://doi.org/10.17743/jaes.2018.0002.

[28] Algazi VR, Duda RO, Duraiswami R, Gumerov NA, Tang Z. Approximating the head-related transfer function using simple geometric models of the head and torso. J Acoust Soc Am 2002;112(5):2053–64. https://doi.org/10.1121/1.1508780.

[29] Gumerov NA, O'Donovan AE, Duraiswami R, Zotkin DN. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. J Acoust Soc Am 2010;127(1):370–86. https://doi.org/10.1121/1.3257598.

[30] Bernschütz B. A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100, in: Fortschritte der Akustik – AIA-DAGA 2013; 2013. p. 592–595.

[31] Wu M, Wang DL. A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Trans Audio Speech Lang Process 2006;14 (3):774–84. https://doi.org/10.1109/TSA.2005.858066.

[32] Naylor PA, Gaubitch ND. Speech Dereverberation, Signals and Communication Technology. London: Springer, London; 2010. https://doi.org/10.1007/978-1-84996-056-4.

[33] Cecchi S, Carini A, Spors S. Room response equalization–a review. Appl Sci 2017;8(1):16. https://doi.org/10.3390/app8010016.

[34] Gardner B, Martin K. HRFT Measurements of a KEMAR Dummy-head Microphone #280, Tech. rep., Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology; 1994.

[35] Avendano C, Algazi VR, Duda RO. A head-and-torso model for low-frequency binaural elevation effects, in: Proceedings of the IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics, WASPAA; 1999. p. 179–182.https://doi.org/10.1109/ASPAA.1999.810879.

[36] Everest FA, Pohlmann KC. The master handbook of acoustics. 6th ed. McGraw-Hill; 2015.

[37] Lopez JJ, Gutierrez-Parera P. Equipment for fast measurement of Head-Related Transfer Functions. In: Audio Engineering Society 142nd Convention, Berlin, Germany; 2017. pp. e-Brief 335.

[38] Lopez JJ, Martinez-Sanchez S, Gutierrez-Parera P. Array processing for echo cancellation in the measurement of Head-Related Transfer Functions. In: EAA Euronoise 2018, Hersonissos, Crete, Greece; 2018. p. 2581–2588.

[39] Ramos G, Cobos M. Parametric head-related transfer function modeling and interpolation for cost-efficient binaural sound applications (L). J Acoust Soc Am 2013;134(3):1735–8. https://doi.org/10.1121/1.4817881.

[40] Ramos G, Cobos M, Bank B, Belloch JA. A Parallel Approach to HRTF Approximation and Interpolation Based on a Parametric Filter Model. IEEE Signal Process Lett 2017;24(10):1507–11. https://doi.org/10.1109/LSP.2017.2741724.

[41] Earthworks Inc., M30 High Definition Measurement Microphone (Accessed: 2021-06-21). url: https://earthworksaudio.com/wp-content/uploads/2018/07/M30-Data-Sheet-2018.pdf.

[42] Mh acoustics, em32 Eigenmike microphone (Accessed: 2021-06-21). url: https://mhacoustics.com/products.

[43] Brüel & Kjær, TYPE 4100 – Brüel & Kjær Sound & Vibration, sound quality Head and Torso Simulator (Accessed: 2021-06-21). url: https://www.bksv.com/en/products/transducers/ear-simulators/head-and-torso/hats-type-4100.

[44] Bernschütz B, Pörschmann C, Spors S, Weinzierl S. SOFiA Sound Field Analysis Toolbox. In: International Conference on Spatial Audio (ICSA). p. 7–15.

[45] Kodrasi I, Gerkmann T, Doclo S. Frequency-domain single-channel inverse filtering for speech dereverberation: Theory and practice. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE; 2014. p. 5177–5181.https://doi.org/10.1109/ICASSP.2014.6854590.