OPEN ACCESS Freely available online

CC BY

H. Bahu, J.-M. Jot, T. Carpentier, M. Noisternig, M. Mihocic, P. Majdak, and O. Warusfel, "Toward Improved Consistency Between Databases of Head-Related Transfer Functions," *J. Audio Eng. Soc.*, vol. 74, no. 1/2, pp. 35–49 (2026 Jan./Feb.). https://doi.org/10.17743/jaes.2022.0247.

Toward Improved Consistency Between Databases of Head-Related Transfer Functions

HÉLÈNE BAHU,[1,*] **JEAN-MARC JOT,**[2]
(helenebahu@gmail.com) (jmmjot@gmail.com)

THIBAUT CARPENTIER,[1] **MARKUS NOISTERNIG,**[1]
(thibaut.carpentier@ircam.fr) (markus.noisternig@ircam.fr)

MICHAEL MIHOCIC,[3] **PIOTR MAJDAK,**[3] **AND**
(michael.mihocic@oeaw.ac.at) (piotr.majdak@oeaw.ac.at)

OLIVIER WARUSFEL[1]
(olivier.warusfel@ircam.fr)

[1] *STMS, Sorbonne Université, CNRS, IRCAM, Ministère de la Culture, 75004 Paris, France*
[2] *Virtuel Works LLC, Aptos, CA*
[3] *Acoustics Research Institute, Austrian Academy of Sciences, 1010 Vienna, Austria*

Since the creation of the spatially oriented format for acoustics (SOFA, Audio Engineering Society standard AES69), numerous databases of head-related transfer functions (HRTFs) are now available as standardized SOFA files. However, the methodologies for measuring and postprocessing HRTFs vary significantly across laboratories. This leads to objective and perceptual inconsistencies between HRTF databases and makes it challenging to integrate multiple databases into a single repository to facilitate wide-scale research and application. This paper introduces a normalization procedure, applicable to any HRTF data set, aimed at enhancing the consistency across HRTF data sets obtained from different laboratories while preserving the spatial information essential to HRTFs. The proposed approach consists of six processing steps: low-pass filtering, temporal alignment, temporal windowing, diffuse-field equalization, low-frequency extrapolation, and far-field correction. The normalization was evaluated on 17 HRTF data sets of the same dummy head by means of acoustic analyses and auditory simulations and further validated with respect to a database of 54 human subjects. Results show that the proposed normalization improves data set applicability and consistency while maintaining the directional cues within each data set.

Keywords: head-related transfer function (HRTF), binaural rendering, normalization procedure, database signature, postprocessing

1 INTRODUCTION

Head-related transfer functions (HRTFs) are filters characterizing how a listener's morphology affects a sound wave as it travels from a source to the entrances of the listener's ear canals. HRTFs encode the directional cues for sound localization, including the interaural time differences (ITDs), interaural level differences (ILDs), and spectral cues. These filters can be used in binaural technology to simulate virtual sound sources over headphones. For a compelling listening experience, it may be necessary to apply HRTFs of the individual listener [1, 2]. HRTFs can be measured acoustically in an anechoic chamber by placing miniature microphones at the entrances of the listener's ear canals (ear canal entrance point [3]) and loudspeakers at various spatial locations. However, such measurements are complex and time consuming. Alternatively, various methods have been developed to individualize HRTFs for the listener [4, 5].

Theoretically, HRTFs capture only listener-specific filtering effects, independent of the measurement environment. Consequently, HRTF sets measured on the same head in different facilities should be identical. However, this is not the case [6], even after postprocessing aimed at removing the effects of the measurement system. In practice,

*To whom correspondence should be addressed, e-mail: helenebahu@gmail.com.

HRTFs acquired in different facilities exhibit significant discrepancies due to variations in the acoustic environment, equipment, measurement techniques, and postprocessing methodologies.

The first source of discrepancies among HRTF databases stems from differences in the measurement system. The spatial measurement grid (defined by the distribution and density of directions, elevation range, and measurement distance) varies across databases, often constrained by apparatus-specific limitations. Acoustic environments also differ, ranging from fully anechoic to semianechoic, with residual reflections shaped by wall structures and the measurement apparatus. Also, hardware equipment choices vary between laboratories, as do system identification methods. These choices include technical parameters, such as the excitation signal (e.g., swept sine, maximum length sequence, and multiple exponential sweeps [7]) and its properties (e.g., amplitude, duration, and repetition), the audio sampling rate, and measurement procedure (e.g., continuous rotation versus discrete angular steps). Additional factors include subject position monitoring during the measurement, in-ear microphone placement, the number of loudspeakers used, and equipment calibration, all of which contribute to inconsistencies across data sets.

The second issue is the postprocessing methodology, which varies often as well [8]. The head-related impulse responses (HRIRs) need to be truncated to remove possible reflections from the environment, and the window length and type may vary. To eliminate the contribution of the transducers, free-field equalization or diffuse-field equalization is applied. The way the equalization process is conducted has a significant effect on the HRTF data. At low frequencies (below 200 Hz) and high frequencies (above 18 kHz), the loudspeaker's transmitted energy is usually low. Therefore, the HRTF magnitudes are sometimes corrected using various techniques. In the time domain, the filters' length and temporal alignment vary depending on the database. Furthermore, the filters are sometimes represented as a pair of minimum-phase filters combined with a broadband ITD [9]. The ITD extraction method has an effect on the result [10], but no standard method has been established to date.

Prior to the creation of the spatially oriented format for acoustics (SOFA; Audio Engineering Society standard AES69 [11–13]), HRTFs have been stored in various laboratory-specific formats (e.g., MARL-NYU [14], miro [15], and 3DTI [16]). Nowadays, most HRTF data sets are stored as SOFA files, which guarantees consistency in the storage format. SOFA works by means of conventions (i.e., a set of defined data format and metadata). For example, the most-widely used convention "SimpleFreeFieldHRIR" defines the data to be stored as spatially discrete HRIRs accompanied by various metadata describing the measurements [12]. However, these conventions do not specify how the HRTF data should be acquired and processed.

These issues lead to significant consequences. First, HRTF measurements capture not only the acoustic effects of the subject's morphology but also those of the laboratory environment, making it challenging to investigate interindividual variations and develop solutions for individualization. With the increasing use of machine learning (ML) in this field, combining multiple databases has been proposed to expand training data sets [4, 17]. However, Pauwels and Picinali [18] demonstrated that after merging various HRTF databases an ML classifier can easily identify the original measurement facility, revealing environment-specific signatures that can obfuscate the interindividual characteristics under study. This limitation hampers progress in HRTF research. Second, discrepancies between databases lead to perceptual effects during rendering, such as variations in perceived direction, loudness, and timbre. Third, the difficulty in retracing how HRTF data was acquired and processed not only limits its use but also reduces users' confidence in the data when applied to practical scenarios. This issue may help explain why, despite the availability of many publicly available HRTF databases,[1] efforts to collect new databases continue.

To reduce the aforementioned measurement discrepancies, this article proposes a method for normalizing HRTF databases. After briefly reviewing previous efforts to address this issue, SEC. 2 outlines the proposed normalization procedure. SEC. 3 presents an evaluation of the method applied to HRTFs of the same dummy head (Neumann KU100) measured in different laboratories, and SEC. 4 applies it to a subset of an HRTF database of human subjects measured in the same lab. Finally, the results are discussed in SEC. 5.

1.1 Previous Efforts on Normalizing Head-Related Transfer Function Databases

In view of the existing discrepancies among HRTF databases and the limitations that they imply, several studies have attempted to reduce them (in other words, to "normalize" them). They are summarized here.

Andreopoulou and Roginska [14] worked on normalizing four databases through the harmonization of the sampling grid, sampling rate, filter length, and modeling as minimum-phase filters plus ITD. The consistency of the compiled database was evaluated using multidimensional scaling based on the Euclidean distance of HRTF log magnitudes (horizontal plane only). The results indicated that subtracting the mean and normalizing the standard deviation (SD) for each HRTF filter was a key step toward consistency.

The "Club Fritz" project [6] compared HRTF data sets of a unique dummy head measured acoustically in ten different labs. The same postprocessing methodology was applied to all data sets: free-field equalization, frequency resampling, low-pass filtering at 18 kHz, windowing (with a 2.7 ms rectangular window), DC offset removal, gain normalization, and hemisphere imbalances correction. Note that for free-field equalization different microphones were used. Results indicated that significant differences between HRTF data

[1] https://sofacoustics.org/data/database/ provides approximately 20 acoustically measured HRTF databases (last visited on February 28th, 2025).

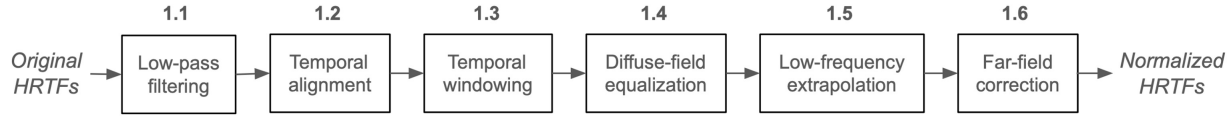


Fig. 1. The six steps proposed for the normalization procedure. Numbers 1.1–1.6 refer to the corresponding sections in the text body.

sets remained after normalization: up to 22 dB in HRTF magnitude (above 6 kHz, at azimuth 0° and 180° on the horizontal plane) and 235 μs in ITD (on the side). Although no listening test was conducted, the ITD variations often exceeded just-noticeable differences (e.g., see [19]) and thus were likely to be perceptible.

Tsui and Kearney [20] developed a MATLAB tool for HRTF SOFA files to detect abnormalities in the measurement grid, find common directions between multiple files, and normalize the sampling frequency, HRIR length, and maximum gain. However, this study did not aim at reducing the signature of the measurement setup.

DeRus et al. [21] evaluated the impact of some preprocessing techniques on HRTF classification into elevation sectors (median plane only), using a collection of eleven databases. The preprocessing steps included gain normalization, different restrictions of the frequency range, mel warping (selected frequency bins equally spaced on a mel scale), linear or logarithmic amplitude scaling. Results indicated that these preprocessing steps had no significant effect on the classification results and that the classification into elevation sectors was better when the model was trained and tested within a subset of three databases measured in the same facility, thus revealing the influence of the measurement system's signature.

Wen et al. [22] attempted to normalize eight databases by removing the per-direction systematic response of the measurement system. This response was estimated for each database by averaging all HRTFs measured at a single direction (all subjects, left and mirrored right ear) for two directions common to all databases (front and side). They showed that the systematic response of the measurement system varied with the direction and that the database classification accuracy was reduced by this normalization procedure.

Pauwels and Picinali [18] tested the impact of some data processing on the ability of an ML classifier to distinguish between ten databases (eight acoustically measured, two simulated). They used the free-field equalized versions of the HRIRs provided by the original laboratories. A common processing methodology was applied, including frequency resampling, gain normalization, HRIR truncation to 5.33 ms, DC removal, low-pass filtering at 18 kHz, and minimum-phase representation. Results showed that, despite the preprocessing, the ML classifier could identify the data set origin. Contributors to the data set signature include the magnitude at low and high frequencies and room reflections. The variability in the equalization methods applied by the providers of the databases may be a source of the remaining discrepancy.

Unfortunately, all these methods are limited in removing the database signature. First, the evaluation methods were limited, considering the perceptual perspective. For example, the classification results highly depend on the distance metric, and the metrics used did not consider all aspects of the actual auditory processing in humans. Also, the comparison was limited to a few directions on the horizontal plane. Further, the evaluations provide little insight on the potential degradation of spatial information after having the normalization applied. Second, these methods did not account for other important sources of discrepancies, such as the low-frequency extension, the distance between the source and listener, or the equalization method.

2 THE NORMALIZATION PROCEDURE

The authors propose a set of postprocessing steps aiming at removing the signature of the measurement setup, hence normalizing the HRTF database. The removal of the signature of the measurement system may impact the directional cues and therefore affect perceived sound localization. For instance, the removal of the spectral characteristics of the transducers may impact the HRTF magnitudes of the two ears independently, thus affecting the ILD (e.g., by compensating the left-right gain imbalances). Hence, the goal of this work is to define a normalization procedure that fulfills the following criteria:

- The HRTFs of the same head measured for the same spatial position but in different laboratories yield an equivalent listening experience.
- The perceptually relevant spatial information is preserved, as evidenced by negligible differences in sound localization as compared with the original HRTFs.

To provide a perceptual equivalence, the various perceptual attributes of HRTFs, as was highlighted in Simon et al. [23], shall be considered: coloration, elevation, immersion, position in terms of front-back confusions, lateral position, realism, and relief. Note that, although not retained in [23], loudness is also an important attribute.

In the following, the authors describe the proposed normalization procedure consisting of six successive processing steps as illustrated in Fig. 1: low-pass filtering to harmonize the high-frequency support, temporal alignment of the HRIRs, temporal windowing to harmonize the HRIR's duration and envelope, diffuse-field equalization to remove system-specific colorations, low-frequency extrapolation to harmonize the low-frequency support, and far-field correction to harmonize across measured source-to-listener distances. Note that the original data sampling rate is preserved because these processing steps do not need to normalize the sampling rate. The normalization procedure is implemented

in the SOFA Toolbox version 2.4.0[2] and can be applied independently to any existing HRTF data set.

2.1 Low-Pass Filtering

Above 18 kHz, the measurement gain is usually low, and the measurements are not reliable due to the insufficient energy response of the transducers. Thus, the first processing step consists in applying a low-pass filter at 18 kHz on all HRIRs [6]. The low-pass filter is implemented as a linear-phase filter with an impulse response of 5.3 ms length. It was designed using least-squares error minimization such that its spectral amplitude drops from one at 18 kHz to zero at 19 kHz, not affecting other frequencies below 18 kHz.

2.2 Temporal Alignment

The HRIRs' time of arrival (or onset) varies with the source position and ear, and the minimum time of arrival in an HRIR set depends on the original database. In some applications (e.g., when combining parametric reverberation with HRIR sets from different databases), it is useful to time-align the HRIRs to the same reference time mark. Additionally, time alignment facilitates temporal windowing.

For these reasons, the second step in the normalization procedure consists in aligning the onset of the HRIR at the frontal direction to a reference time mark, here chosen to be 1 ms. More specifically, the earlier onset between the left and right HRIRs at the frontal direction is aligned to 1 ms. Note that the frontal direction (azimuth and elevation angles of 0°) is usually present in all HRTF data sets.

Although there exist various onset detection methods (e.g., see [24] for room impulse responses), HRIR onsets are usually estimated through leading-edge detection (i.e., by detecting the time instants at which the HRIRs reach a certain percentage of the maximum peak amplitude [25]). This percentage, or threshold, varies greatly between studies. Following International Organization for Standardization (ISO) recommendations for room impulse responses (ISO 3382-1), the HRIR onset is defined as the last point before and below 20 dB of the HRIR peak. A reference time mark of 1 ms was chosen to preserve the earliest HRIR onsets[3] (considering that the largest ITD is about 0.8 ms for human heads [25, 26]) while leaving enough time for the HRIR tail. Fig. 2 shows an example of such a temporally aligned HRIR. For each subject of a database, the HRIRs at all directions and both ears are delayed by the same time shift, thus preserving the original ITD information for all measured HRTF directions.

2.3 Temporal Windowing

Windowing the HRIR allows for removal of reflections from the acoustic environment and potential preringing artifacts from mechanical transmission. The choice of the window length is important: a window that is too short may remove some low-frequency information, and a window that is too long may keep undesired noise or room reflections. According to previously published databases, the window is usually an asymmetrical Hann window whose length ranges between 2.7 ms (VIKING [27]) and 11.6 ms (Crossmod, Listen), with a median length of 5.8 ms (i.e., 256 samples at 44.1 kHz). In Vidal et al. [28], a frequency-dependent window was applied with a length of 43 ms for the lower frequencies (<700 Hz) and 7.4 ms for the higher frequencies.

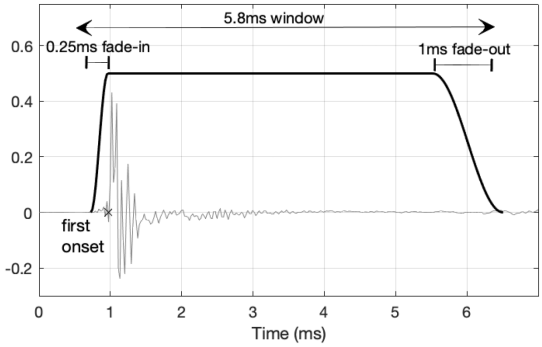


Fig. 2. Temporal window (black) and an example of the temporally aligned frontal HRIR (gray).

Here, an asymmetric Hann window of 5.8 ms is applied starting at the HRIRs' first onset minus a 0.25-ms margin. The window contains a half-Hann fade-in of 0.25 ms and a half-Hann fade-out of 1 ms. An illustration of the window is given in Fig. 2. The HRIRs' first onset is defined by the earliest time onset across all directions and the two ears. The onset detection method is the same as the one presented in SEC. 2.2. The original HRIR lengths are preserved unless they are shorter than 5.8 ms, in which case zero-padding (to 5.8 ms) is applied.

2.4 Diffuse-Field Equalization

HRTF equalization aims at removing the response of the measurement system, including loudspeakers and microphones [29]. Free-field equalization and diffuse-field equalization are two commonly applied methodologies. In the context of the equalization of headphones and dummy heads, free-field equalization conventionally consists in compensating for the HRTF at the frontal direction [29, 30]. In the context of HRTF measurements, it is more common to use a reference free-field measurement made by placing, at the center of the measurement system (with the head absent), either a reference microphone or the in-ear microphones used for measuring the HRTFs. However, this reference measurement is usually not published along with the HRTF database.

Larcher et al. [29] have shown that diffuse-field equalization reduces the perceptual differences between recordings made with different equipment. Additionally, in the context of normalizing HRTF databases, diffuse-field equalization is attractive because it compensates for all the direction-independent components. As such, it may be applied on

[2] https://ecosystem.sonicom.eu/tools/14; https://doi.org/10.60887/q17dygeudms0aksh.

[3] For instance, for a head with a maximum ITD of 0.8 ms and a source on the side, the ipsilateral HRIR's onset to appear around $(1 - \frac{0.8}{2}) = 0.6$ ms after temporal alignment.

unprocessed or equalized HRTF measurements. Also, it compensates for the overall data set gain and possible left-right imbalances (e.g., caused by differences in microphone placement or response).

For these reasons, diffuse-field equalization is preferred. Ideally, the diffuse-field HRTF is measured in a reverberant room, but this measurement is rarely available. Alternatively, it can be obtained by superposing the transformations undergone by plane waves in the free-field coming from all directions [29]. In practice, the power average of measured HRTFs, or common transfer function (CTF), may be used. Because the spatial sampling may not be regular and complete, each HRTF is weighted by the solid angle surrounding the measurement point, estimated using a spherical Voronoi diagram [31].[4] Therefore, the diffuse-field HRTF is estimated using the weighted power average of all available HRTFs (or weighted CTF):

$$\mathrm{CTF}_e = \sqrt{\sum_{i=1}^{M} \frac{1}{M} |H_e(\theta_i, \phi_i, f)|^2 \cdot w_i}, \tag{1}$$

with $i = 1, \ldots M$ the direction index, θ_i and ϕ_i the azimuth and elevation of direction i, M the total number of measured directions, e the left or right ear, and w_i the weight proportional to the Voronoi surface.

Diffuse-field equalization is applied by dividing the complex HRTF at each direction by the minimum-phase CTF. Similar to Middlebrooks [32], the minimum-phase CTF is a combination of the CTF magnitude spectrum [in Eq. (1)] and CTF minimum-phase spectrum (obtained by Hilbert transform of the logarithm of the magnitude spectrum). To avoid the amplification of low-magnitude values at inversion, the CTF magnitude is flattened above 18 kHz, with a transition interval between 16 and 18 kHz, where it is set to be equal to the average CTF magnitude estimated in the range from 200 Hz to 16 kHz.

2.5 Low-Frequency Extrapolation

The information carried by raw HRTF measurements at low frequencies is not relevant for several reasons: (1) the human morphology becomes small compared with the wavelength, and HRTF variations are marginal below approximately 400 Hz [25]; (2) small studio monitors such as typically used for HRTF measurements cannot reproduce frequencies below 100 Hz due to their small size [25]; and (3) the anechoic behavior of the environment is valid up to a lower cut-off frequency (between 75 Hz and 315 Hz, depending on its characteristics) [33]. Consequently, various methods have been employed to correct the HRTF measurements at low frequencies.

Similar to Bernschutz et al. [34], the analytical solution of a rigid sphere can be used. Indeed, the spherical head model (SHM) closely matches typically measured HRTF patterns in the low-frequency range (below 1 kHz) [35]. The solution for an SHM with a standard head radius of 8.7 cm [36] exhibits a low-frequency asymptote below 200 Hz. This low-frequency gain is computed for each HRTF's source direction and distance, and the HRTF magnitude near 250 Hz is linearly interpolated toward the SHM gain at DC.

2.6 Far-Field Correction

The HRTF measurement distance varies between databases and is often driven by practical constraints. Beyond 1 m, HRTF filters are asymptotically independent of source distance [25] and are called far-field HRTFs. In contrast, near-field effects are observed below 1 m, such as an ILD increase for lateral sources [35]. Far-field HRTFs can be inferred from near-field HRTFs (and reciprocally) by applying distance variation functions (DVFs) designed with the SHM [37, 38]. The SHM correctly approximates the spectral changes between the near-field and far-field on human heads, and when combined with the measured HRTF, the method preserves the unique features of the head, torso, and pinna [38]. Similarly, to harmonize the measurement distance of several HRTF data sets, DVFs can be used:

$$\mathrm{DVF}(\theta_i, \phi_i) = \frac{H(r_{ref}, \theta_i, \phi_i)}{H(r_m, \theta_i, \phi_i)}, \tag{2}$$

with r_{ref} the reference far-field distance and r_m the measurement distance. Here, the DVFs are designed with an SHM of radius 8.7 cm and a reference far-field distance of r_{ref} =100 m (approximating infinite distance). After low-frequency extrapolation and far-field correction, the HRTF magnitude converges to 0 dB at DC.

3 EVALUATION

Following the criteria presented in SEC. 2, the normalization process is evaluated in two steps. First, it is verified that the normalization improves the spectral and perceptual consistency between HRTF data sets measured on a same head in different facilities. Second, it is verified that the normalization has a negligible impact on directional cues by comparing the localization performance obtained with HRTFs before and after the normalization.

Perceptual comparison of HRTFs can be performed by means of behavioral listening tests or by simulations. Some perceptually driven models have been proposed and made available in MATLAB or as part of the Auditory Modeling Toolbox (AMT) [39]. For instance, they allow for prediction of the sound localization [40], coloration [41], and loudness [42] of binaural signals. As they have been evaluated with respect to experimental data, they are employed in the following as an efficient simulation of listening tests.

3.1 Data Sets

The evaluation was conducted with 17 HRTF data sets of a dummy head (Neumann KU100) measured at multiple facilities. This dummy head does not have any torso or shoulders, and it is equipped with high-quality internal

[4] In the absence of measurements in a specific portion of the grid (e.g., at low elevations), the weights associated with the adjacent measurement points are higher as they cover larger solid angles. This is equivalent to considering that the magnitude spectrum is constant in this region.

microphones positioned at the ear canal entrance point and features built-in diffuse-field equalization provided by the manufacturer.[5]

The data sets are listed in Table 1. The first twelve sets are HRTF data sets from the Club Fritz database [6] (single dummy head microphone). Data sets 1 to 11 are unprocessed HRTF measurements, whereas for the latter six sets, specific postprocessing was applied: the HRIRs of the Acoustics Research Institute (ARI) were windowed and free-field compensated [43]; the HRIRs of the IRCAM BiLi database [44] were windowed and compensated for the interloudspeaker differences; the two HRIR sets of the Technische Hochschule Köln (THK) [15] were windowed, free-field equalized, extended at low frequencies and normalized in gain; for SADIE II [45], multiple postprocessing steps were applied (which are presented in detail in SEC. 3.2.3 of the paper), including diffuse-field equalization; for SS2 [46], HRIRs were windowed, free-field equalized, and low-pass filtered at 16 kHz.

These data sets all used the built-in dummy head microphones, except for data set 17, which used external in-ear microphones. Table 1 provides more information about the data sets. Note that they share only two spatial directions: at the azimuth angles of 0° (front) and 180° (back) in the horizontal plane. In the following, all data sets were processed using the method described in SEC. 2. The data sets are available through the SONICOM Ecosystem.[6]

3.2 Diffuse-Field Head-Related Transfer Function Estimates

Fig. 3(a) shows the left ear diffuse-field HRTF magnitude spectrum, estimated using Eq. (1), for each of the 17 KU100 data sets. Following the processing steps outlined in SEC. 2, these estimates were obtained after applying low-pass filtering, temporal alignment, and windowing to the HRIRs. The magnitude drop around 18 kHz is the result of the low-pass filter and justifies flattening the diffuse-field equalization filter in this region to avoid amplifying these low-magnitude values during inversion.

The diffuse-field HRTF magnitude spectra clearly exhibit inter–data set differences in overall gain, which can be compensated for by applying diffuse-field equalization. Spectral differences across data sets are also apparent. Notably, despite the diffuse-field equalization applied by the dummy head manufacturer, the diffuse-field HRTF magnitudes estimated from HRTF measurements are not flat.

This may be due to two main reasons. First, inter–data set variations in the diffuse-field HRTF magnitude spectrum suggest the presence of residual measurement system signatures (e.g., loudspeaker responses, environmental reflections, or slight positioning errors) and/or data set–specific processing applied by the original laboratory. In the context of HRTF database normalization, this is precisely the type

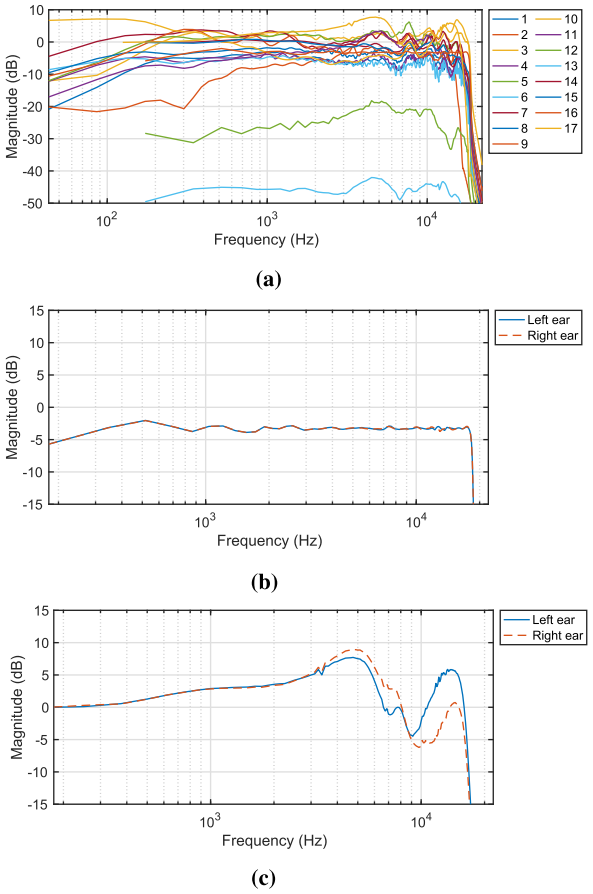


Fig. 3. Diffuse-field HRTF magnitude spectra estimated using Eq. (1). (a) Left ear diffuse-field HRTF magnitude spectra for all data sets. (b) Example of a data set that has already been diffuse-field equalized by the original laboratory (data set 16). (c) Example of a data set showing left-right asymmetries in diffuse-field HRTF magnitude spectra (data set 17).

of variability that is targeted for elimination, thereby justifying the removal of estimated diffuse-field HRTF from all measurements.

Second, differences may arise from the specific method used to derive the diffuse-field response. Here, the diffuse-field transfer function is estimated via a weighted average of HRTF measurements across all available directions [see Eq. (1)]. When the directional sampling is sufficiently dense and uniform and covers the full sphere, this approach should provide an accurate estimate. In practice, however, measurement grids rarely achieve full and uniform spherical coverage. Although spatial weighting compensates for these irregularities [see Eq. (1)], the estimated diffuse-field HRTF response may still deviate slightly from the "true" one. Nonetheless, as demonstrated in the next section, dividing the HRTFs by these estimates significantly reduces inconsistencies across data sets.

Fig. 3(b) highlights the specific case of data set 16, which was already diffuse-field equalized by the original laboratory [45]. Its diffuse-field HRTF shows an overall gain offset of approximately −3 dB and is otherwise almost flat. The small residual fluctuations are consistent with Fig. 6 of the original paper [45]. Apart from correcting the overall gain, further diffuse-field equalization has little effect on

[5] See "Operating instructions KU100, Engl./Germ., 05/2020" available at www.neumann.com.

[6] https://ecosystem.sonicom.eu/databases/25; https://doi.org/10.60887/j57ffyefbxw69cw2.

Table 1. Information about the 17 data sets of the same head used to evaluate the proposed normalization.

Identification number	Origin	Year	M	L	f_s (kHz)	D (ms)	Elevation range (°)	r (m)
1	IRCAM	2004	2016	1	44.1	23.2	−45 to 90	1.95
2	UMD	2007	823	6	44.1	23.2	−35 to 90	0.9
3	NASA	2007	432	12	96	23.2	−40 to 70	0.9
4	IRCAM	2007	1944	3	44.1	23.2	−40 to 90	2
5	ITA	2009	2016	1	44.1	23.2	−80 to 90	2
6	Helsinki	2009	180	1	48	23.2	−40 to 40	1.5
7	NHK	2009	720	10	48	23.2	0 to 90	1.3
8	NICT	2009	1008	1	48	23.2	−40 to 90	1
9	Nagoya	2009	2017	1	48	23.2	−50 to 90	1
10	RIEC	2010	648	35	48	23.2	−80 to 90	1.5
11	IRCAM	2014	1560	1	96	23.2	−39 to 86	1.95
12	ARI	2014	1550	22	48	23.2	−30 to 80	1.2
13	IRCAM BiLi	2016	1680	4	96	5.3	−50.5 to 85.5	2.06
14	THK-nf	2016	2702	1	48	2.7	−90 to 90	0.5
15	THK-ff	2016	2702	1	48	2.7	−90 to 90	1.5
16	SADIE II	2018	8802	23	48	5.3	−90 to 90	1.2
17	SS2	2023	1625	54	48	8.0	−69 to 90	2

Note: Data set 1–12 are from the Club Fritz database [6]. M = number of spatial positions; L = number of loudspeakers; f_s = sampling rate; D = HRIR duration; r = distance of the sound source. ARI = Acoustics Research Institute; IRCAM = Institut de Recherche et de Coordination Acoustique/Musique; UMD = University of Maryland; NASA = Advanced Controls and Displays Group, NASA Ames Research Center; ITA = Institute of Technical Acoustics, RWTH Aachen; Helsinki = Department of Signal Processing and Acoustics at Helsinki University of Technology; NHK = Science & Technology Research Laboratories department of the Japan Broadcasting Corporation; NICT= Multisensory Cognition and Computation Laboratory at National Institute of Information and Communication Technology; Nagoya = Takeda Laboratory, Nagoya University; RIEC = Research Institute of Electrical Communication, Tohoku University; THK-nf = near-field (nf) data of the Technische Hochschule Köln (THK); THK-ff = far-field (ff) data of the THK; SADIE II = database of the University of York; SS2 = Sound Sphere 2 database of Reality Labs Research.

this data set, consistent with the idempotent nature of the operation.

Finally, diffuse-field HRTF magnitude spectra may also exhibit left-right asymmetries caused by the measurement environment or slight differences in the response or placement of the in-ear microphones. This is clearly observed in Fig. 3(c) for data set 17, where external in-ear microphones were used. These asymmetries are effectively captured in the estimated diffuse-field HRTF magnitude spectra and are therefore addressed by the proposed normalization procedure.

3.3 Spectral Consistency

Fig. 4 shows the left ear HRTF magnitudes at the frontal direction before [Fig. 4(a)] and after the processing steps described in SEC. 2 [Fig. 4(b)] for the 17 data sets. Note that normalized data conserves its original sampling rate and that zero-padding was applied to visually increase the frequency resolution.

In Fig. 4(a), the original HRTF magnitudes exhibit differences in overall gain and various low-frequency and high-frequency behaviors. In particular, data set 2 lacks energy below 1 kHz and above 15 kHz. One notices the presence of oscillations on some magnitudes, which can be due to noise or reflections that interfere with the direct sound. The amount of smoothing also depends on the number of samples.

In Fig. 4(b) (after the normalization), a better spectral consistency in overall gain and low-frequency behavior and a reduction of oscillations is observed. At high frequencies, the consistency is mostly improved by the low-pass filtering at 18 kHz. However, the magnitudes of data sets 2 and 17

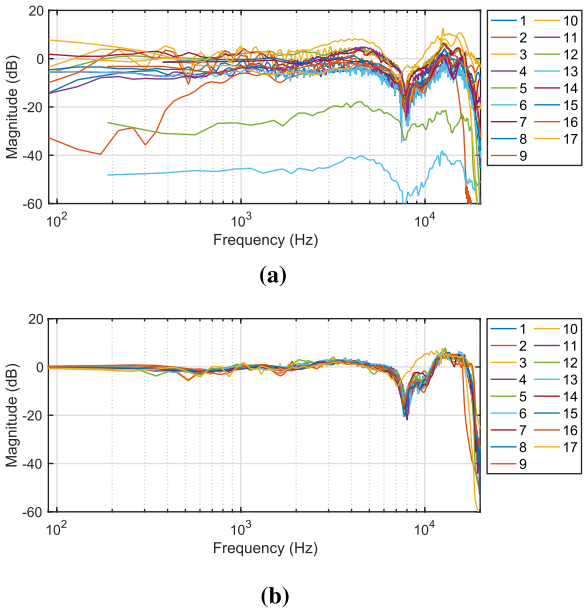
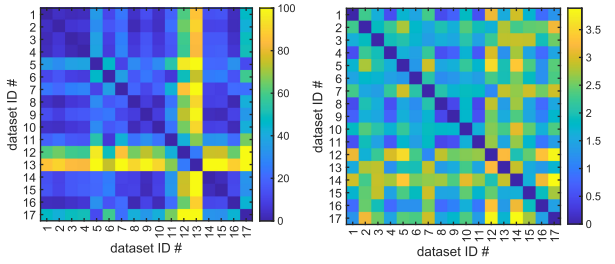


Fig. 4. Spectral magnitudes of the left ear HRTFs at frontal direction from the evaluated data sets before (a) and after (b) normalization.

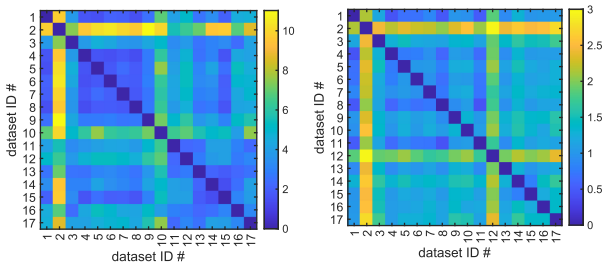
drop at a lower frequency than the other curves (around 16.1 and 16.6 kHz, respectively, instead of 17.5 kHz).

As mentioned earlier, data set 2 lacks energy above 15 kHz, and data set 17 was low-pass filtered at 16 kHz by the original laboratory. As designed, the normalization does not compensate for missing energy in the range between 16 and 18 kHz because the magnitude of the diffuse-field equalization filter is progressively flattened in this range. It also does not resolve the spectral particularity of data set



(a) (b)

Fig. 5. Loudness differences (in sones) between all considered data sets calculated before (a) and after (b) the normalization. Note the different color scales between the two panels.



(a) (b)

Fig. 6. Coloration differences (in sones) between all considered data sets calculated before (a) and after (b) the normalization. Note the different color scales between the two panels.

17 in the range between 7 and 12.5 kHz (note the notch around 7 kHz instead of 8 kHz in the other data sets). This is probably due to the use of external in-ear microphones, whose position in the ears differs from the built-in dummy head microphones.

To quantify the improvement in spectral consistency, the spectral metric of the Baumgartner et al. (2014) model [47] was used. Spectral profiles were computed by applying a filterbank (composed of gammatone filters with one equivalent rectangular bandwidth [ERB]) from 700 Hz to 18 kHz, and the variance of the difference in decibels was calculated for all pairs of HRTFs. Note that this metric is independent from a difference in gain. Results indicate that the average spectral difference at the frontal direction decreases from 14.36 dB to 2.46 dB after the normalization.

3.4 Perceptual Consistency

3.4.1 Loudness

To verify that the proposed normalization procedure reduces perceived differences in loudness between data sets, the Moore-Glasberg method, as defined by ISO 532-2 [42, 48], was used (implemented in MATLAB function `acousticLoudness`). This loudness model calculates the specific loudness over the ERB scale for stationary signals.

A set of 46 directions was considered: 22 directions at elevation 0° (horizontal plane) and 24 directions at elevation 20° (above the horizontal plane) that closely correspond to directions of the 17 data sets, with an azimuth and elevation mismatch of less than 2.5°. The HRTFs at each of the 46 directions were resampled to 48 kHz to remove the effect of the sampling frequency and filtered with a white noise signal of a duration of 100 ms.

Fig. 5 shows the loudness differences between the 17 data sets before [Fig. 5(a)] and after the normalization [Fig. 5(b)], averaged over sound directions. As expected, the difference matrix was zero on the diagonal and symmetric about the diagonal. The normalization reduced differences in loudness: the average loudness difference was 29.7 sones for original HRTFs, 2.1 sones for diffuse-field equalized HRTFs, and 1.7 sones for normalized HRTFs. By compensating for overall gain mismatches, the diffuse-field equalization was mostly responsible for this improvement.

3.4.2 Coloration

To verify that the proposed normalization procedure reduces perceived differences in coloration between data sets, the metric of McKenzie et al. [41] was used (implemented as `mckenzie2022` in the AMT). This metric, referred to as predicted binaural coloration, is based on loudness calculation and predicts the difference in perceived coloration between two sets of binaural signals. It accounts for both loudness and frequency weighting (lesser importance of quieter sounds and less sensitive frequencies). This metric has been shown to strongly correlate with subjective spectral similarity ratings and to outperform other widely used metrics.

To allow pairwise comparisons, both original and normalized HRIRs were resampled to 48 kHz and zero-padded. Also, to ignore overall gain differences (already studied in the previous section), original HRTFs were normalized in gain (by the overall maximum evaluated within a data set). Finally, the metric was averaged over frequency bins (weighted with respect to ERB bandwidth), 46 directions, and two ears to obtain a single distance for each HRTF pair. The full audible range between 20 Hz and 20 kHz was considered.

Fig. 6 shows the predicted differences in perceived coloration between the 17 data sets before [Fig. 6(a)] and after the normalization [Fig. 6(b)]. The normalization significantly reduced differences in the perceived coloration: the average difference was 3.8 sones for original HRTFs, 1.5 sones for diffuse-field equalized HRTFs, and 1.2 sones for normalized HRTFs. As expected, diffuse-field equalization helped to remove system-specific colorations. In Fig. 6(b), higher differences are observed in perceived coloration for data set 2. The reason is that the amplitude below 1 kHz for this data set is very low and varies with direction. As such, the normalization procedure failed to compensate for the missing low-frequency energy in some directions.

3.4.3 Localization

To verify that the proposed normalization procedure reduces perceived differences in localization between data sets, the Bayesian localization model from Barumerli et al. [40] (implemented as `barumerli2023` in the AMT) was used to predict localization performance in both lateral and polar dimensions.

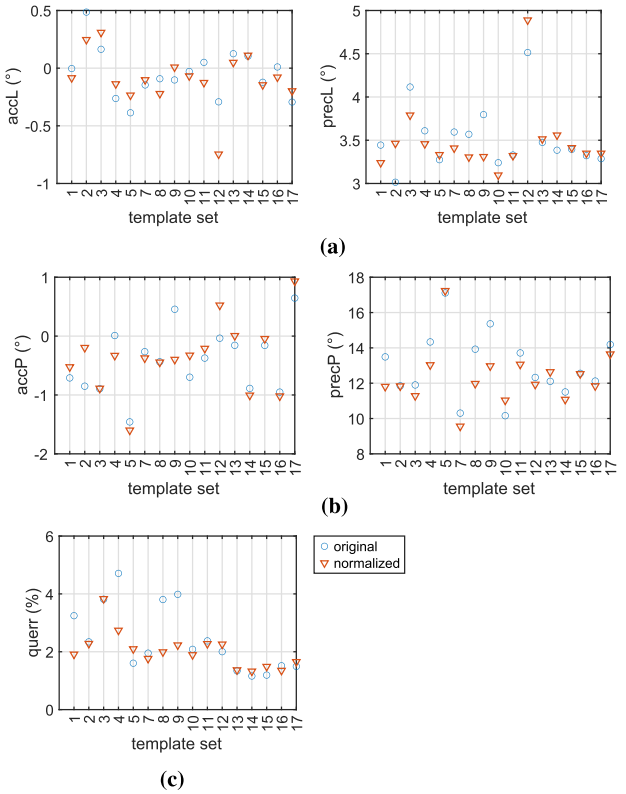


Fig. 7. Localization errors simulated for the tested data sets before (original, blue circles) and after (normalized, red triangles) the normalization. (a) *accL* and *precL*. (b) *accP* and *precP*. (c) *querr*. The simulations were carried out with a model for directional sound localization (Barumerli et al. [40]).

Templates and targets were defined using either the original or normalized HRTFs, and model parameters were set according to Daugintis et al. [49] (including 300 repetitions). Predictions were conducted separately for each data set, with the target and templates corresponding to the same HRTF set. Data set 6 was excluded because it did not meet the model criterion for the number of points.[7] The 46 directions as described in SEC. 3.4.1 (full azimuthal range at elevations 0° and 20°) were considered for the targets.

Fig. 7 presents the predictions of five localization metrics. The localization errors were calculated by subtracting the target angles from the predicted response angles, following Majdak et al. [43]. Lateral and polar accuracy errors (*accL* and *accP*) are measures of the signed average errors in the lateral and polar dimensions, and lateral and polar precision errors (*precL* and *precP*) are measures of the variance in the predicted responses, which can be interpreted as a spatial blur in the lateral and polar dimensions, respectively. Quadrant errors (*querr*) correspond to the percentage of weighted polar errors exceeding 45° within lateral angles of ±60°. These weights compensate for the compression of the polar angle at large lateral angles (see [43]). Note that

[7] The model uses spherical harmonics decomposition up to order $N_{SH} = 15$ to interpolate the model features on a regular grid. For this purpose, it requires that the number of points M of the input data meet the condition $M > (N_{SH} + 1)^2$.

Table 2. Averages $\overline{X}_{KU100}$ and SDs σ_{KU100} of localization errors calculated over all considered data sets before (original) and after (normalized) the normalization.

	$\overline{X}_{KU100} \pm \sigma_{KU100}$ (original)	$\overline{X}_{KU100} \pm \sigma_{KU100}$ (normalized)
accL (°)	−0.05 ± 0.2	−0.09 ± 0.2
precL (°)	3.52 ± 0.4	3.49 ± 0.4
accP (°)	−0.42 ± 0.6	−0.37 ± 0.6
precP (°)	12.94 ± 1.8	12.34 ± 1.6
querr (%)	2.41 ± 1.1	2.03 ± 0.6

Note: The simulations were carried out with a model for directional sound localization (Barumerli et al.) [40].

accP and *precP* were calculated with quadrant errors removed.

In addition to Fig. 7, Table 2 gives the average $\overline{X}_{KU100}$ and SD σ_{KU100} of each metric calculated across the 16 KU100 data sets before and after the normalization. Note that, because the model is probabilistic, small variations of the metrics are expected when run over multiple iterations.

The focus of this section is on the inter–data set consistency in terms of localization, which is reflected in the SD σ_{KU100}. The results shown in Table 2 indicate, if at all, a small change in the SDs after the normalization. Only a small decrease in *precP* and *querr* is observed, which suggests a slight improvement in inter–data set consistency for these two metrics. Changes in localization errors caused by the normalization within each data set are analyzed in the next section.

3.5 Preserving Spatial Information

3.5.1 Localization

To verify that the relevant spatial information has been preserved in terms of sound localization, the localization performance, as predicted by the Barumerli et al. model [40], was compared for HRTF data sets before and after the normalization.

Fig. 7(a) shows the lateral accuracy error (left) and lateral precision error (right) before and after the normalization for each of the 16 KU100 data sets (data set 6 removed). The model predicted lateral accuracy errors close to zero and lateral precision errors close to $\overline{X} = 3.5°$. A good consistency was observed between the original and normalized cases, which was confirmed by the averages in Table 2. Data set 12 exhibited a slightly distinct behavior, with a negative lateral bias that decreased to −0.75° after the normalization and a lateral variance of about 4.5° that increased to 4.9° after the normalization. This distinct result might be explained by a misalignment between the dummy head's orientation and the actual sound direction during the measurements.

Fig. 7(b) shows the polar accuracy error (left) and polar precision error (right) after quadrant errors were removed. For original HRTFs, the polar accuracy error was slightly negative $\overline{X} = -0.4°$, and the polar precision was around $\overline{X} = 13°$. Although the normalization had a small impact on polar accuracy errors, Fig. 7(b) and 7(c) show that it

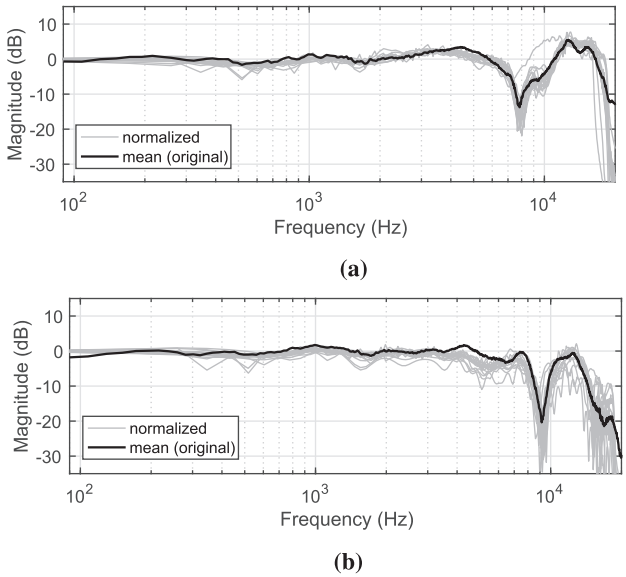


Fig. 8. Spectral magnitudes of the left ear HRTFs from 17 data sets of the same dummy head for the frontal direction (a), azimuth and elevation angles of 0°, and rear direction (b), azimuth angle of 180° and elevation angle of 0°, at eye level. Gray: individual spectra after the normalization. Black: average of the original spectra.

tended to reduce polar precision errors and quadrant errors, especially for data sets 1 to 11. The fact that no processing was applied to these data sets probably induced a larger spatial blur in the polar dimension and larger inter–data set variations, which were both reduced by the normalization.

According to the above observations, the normalization had a negligible effect on the simulated localization performance. It tended to reduce the variance of polar errors and percentage of quadrant errors for data sets containing a strong signature of the measurement system.

3.5.2 Spectral Inspection

To remove the signature of the measurement laboratory from HRTF measurements, an alternative approach to the normalization procedure proposed in SEC. 2 is to average multiple data sets of the same head obtained in different laboratories. The component specific to the measurement system vanishes through the averaging, and the component common to all data sets (i.e., specific to the head) remains. However, this is limited to common directions. This approach was used to validate the results of the normalization for the two directions that are common to the 17 data sets. The black lines in Figs. 8(a) and 8(b) represent the HRTF magnitude averaged over the 17 *original* HRTFs for directions 0° and 180°, respectively.

Overall, a good fit is observed between the average curve (in black) and the 17 normalized HRTF magnitudes (in gray). This gives confidence that no information has been lost through the normalization process. Small mismatches can still be noticed, around 100 Hz because the normalization process has forced the HRTFs to converge to 0 dB, around 4.5 kHz because diffuse-field equalization tends to reduce the pinna resonance, and above 16 kHz due to the effects of diffuse-field equalization and low-pass filtering.

Table 3. Averages $\overline{X}_{BiLi}$ and SDs σ_{BiLi} of localization errors calculated over 54 data sets of the BiLi database before (original) and after (normalized) the normalization.

	$\overline{X}_{BiLi} \pm \sigma_{BiLi}$ (original)	$\overline{X}_{BiLi} \pm \sigma_{BiLi}$ (normalized)	X_{SHM}
accL (°)	0.05 ± 0.2	0.08 ± 0.2	0.02
precL (°)	3.64 ± 0.3	3.74 ± 0.3	4.10
accP (°)	0.51 ± 0.6	0.28 ± 0.7	−4.41
precP (°)	10.12 ± 0.8	10.38 ± 0.8	22.61
querr (%)	0.82 ± 0.6	1.28 ± 0.9	47.86

Note: The rightmost column provides references obtained for an SHM. The simulations were carried out with a model for directional sound localization (Barumerli et al. [40]).

4 APPLICATION TO HUMAN HEAD-RELATED TRANSFER FUNCTIONS

The normalization procedure could conceivably lead to smoothing out both the intrasubject and intersubject spatial cues. The risks could be to deteriorate the individual spatial selectivity and to reduce the interindividual differences. In SEC. 3.5, it was verified that the normalization did not deteriorate the intrasubject spatial cues for a dummy head. In this section, it will be verified that the same conclusions can be drawn for human subjects and that the normalization does not reduce the interindividual differences. To do so, the HRTF data sets of 54 human subjects from the BiLi database [44] were used. Additionally, the SHM was used as a reference, simulating a human without a pinna, thus illustrating the case where the spatial cues would have been completely smoothed out.

The normalization was applied and then a similar evaluation was performed as in SEC. 3.5. Table 3 provides the average $\overline{X}_{BiLi}$ and SD σ_{BiLi} across the 54 data sets. The last column of Table 3 gives the simulation results for the analytical solution of an SHM of radius 8.7 cm.

Similar to previous observations on the KU100, the normalization had a tiny impact on the lateral accuracy and precision (*accL*, *precL*) and the polar accuracy and precision (*accP*, *precP*). However, contrary to previous observations, the percentage of quadrant errors increased slightly from $\overline{X}_{BiLi} = 0.82°$ to $\overline{X}_{BiLi} = 1.28°$. Here, the SD σ_{BiLi} is an indicator of interindividual variability. It was observed that the SDs associated with these metrics changed insignificantly, indicating that interindividual characteristics have been preserved.

Interesting observations can be made when comparing the predicted localization performance between the KU100, human subjects, and an SHM. When comparing Tables 2 and 3, similar errors were observed for *accL*, *precL*, and *accP* between the KU100 and human subjects. However, polar precision errors were somewhat smaller for human subjects, with an average of $\overline{X}_{BiLi} = 10°$ compared with $\overline{X}_{KU100} = 13°$. Quadrant errors were also slightly reduced for human subjects, with an average of $\overline{X}_{BiLi} = 1°$ com-

pared with $\overline{X}_{KU100} = 2^\circ$. This could be interpreted as a better spatial selectivity in the polar dimension in the case of human subjects due to the presence of the torso and finer details in the pinna morphology. Indeed, the KU100 dummy head does not have a torso and has a simplified pinna morphology. The localization performance predicted for the SHM supports this interpretation. Without the pinna, the sphere lacks acoustic cues for sound localization along the polar dimension, leading to significantly larger polar precision errors and quadrant errors.

In view of these observations, polar precision errors and quadrant errors can be interpreted as indicators of the individual spatial selectivity in the polar dimension. The preservation of these errors demonstrates the preservation of the individual spatial selectivity.

5 DISCUSSION

The normalization procedure proposed in this paper was applied and evaluated on 17 HRTF data sets of a dummy head measured in eleven different laboratories. By using models of auditory perception, it was shown that the normalization improved the inter–data set spectral consistency and consistency in perceived loudness, coloration, and localization. These results suggest that the normalization successfully reduced the signature of the measurement laboratory where those different HRTF data sets were acquired and processed. The impact of the normalization on the HRTF spatial cues was also evaluated using a localization model. For this purpose, HRTFs of human listeners from the BiLi database were used in addition to data sets of the dummy head. The results indicate that the normalization has a negligible impact on localization errors.

To validate this model's predictions, listening tests could be conducted, addressing other perceptual attributes and confirming whether the HRTF data sets of the dummy head are perceptually equivalent after the normalization. Although the evaluation in this study mainly focused on the perceptual aspects, it would also be interesting to evaluate the normalization from the point of view of a computational classifier.

Among the normalization's processing steps, diffuse-field equalization was a key step toward the data sets' consistency. A few shortcomings should, however, be noted here. First, diffuse-field equalization was applied as if the measurements were done with only one loudspeaker, which, according to Table 1, is not true for most data sets. The reason is that the information relative to the loudspeaker used for measuring each direction was missing. Apart from data sets 12, 13, and 17, it is not clear whether interloudspeaker differences were compensated beforehand. Second, diffuse-field equalization does not compensate for the nonlinear phase characteristics of the measurement system (e.g., of the loudspeakers). For these two reasons, it is of particular importance that the effect of the equipment be removed by the original laboratory before publishing HRTF measurements. Finally, given Eq. (1), the estimation of the diffuse-field filter depends on the number and distribution of measurement points available. Future work includes the identification of some conditions on the sampling grid that would guarantee the validity of the diffuse-field HRTF estimate.

The normalization procedure proposed here consists of six processing steps. However, additional steps might help to further improve the consistency of, for example, the spatial grid. A common spatial grid would benefit the following: (1) diffuse-field equalization, by improving the consistency of the grid on which the CTF is estimated; (2) evaluations, by increasing the number of common directions between data sets; and (3) spatial audio applications. Representing the HRTFs in a direction-continuous scheme by, for example, using spherical harmonics, might be a useful alternative to the currently common representation in a grid of discrete spatial positions [50]. Other alternative representations could be offered as options in the normalization procedure, such as the minimum-phase plus ITD approximation. Finally, note that a harmonization of the sampling rate was not included in the normalization here. A common sampling rate across all databases is not a trivial choice because upsampling unnecessarily blows up the data set size and downsampling implies a loss of information.

The next step seems to be the application of the normalization to multiple HRTF databases and creation of a large normalized HRTF repository. In this context, some criteria should be defined to automatically verify the validity of the data before normalization (such as sufficient signal-to-noise ratio).

This work does not aim at replacing the original version of the HRTF measurements but rather at offering an additional, normalized version that guarantees a better consistency of HRTF data sets across many databases. For this purpose, a label indicating that a SOFA file followed the proposed normalization process could help in easy identification of a normalized data set. Finally, in an effort to improve the consistency of future HRTF measurements, it would be beneficial to define guidelines on the measurement procedure and postprocessing methodology.

6 CONCLUSION

Here, the authors proposed a normalization procedure for HRTF databases aiming at reducing the differences between them when measured for the same subject and removing the signature of the measurement laboratory. The normalization consists of six steps: low-pass filtering to harmonize the high-frequency support, temporal alignment of the HRIRs, temporal windowing to harmonize the HRIRs' duration and envelope, diffuse-field equalization to remove system-specific colorations, low-frequency extrapolation to harmonize the low-frequency support, and far-field correction to harmonize the distances between the sound sources and listeners. The normalization procedure is implemented in the SOFA Toolbox version 2.4.0 and can be easily applied to harmonize any existing HRTF data set.

The proposed normalization procedure was evaluated on 17 HRTF data sets of the same dummy head measured in different laboratories in terms of its ability to improve

the consistency between data sets and to preserve directional cues. To confirm the directional cue preservation, the normalization was then applied to HRTFs of 54 human subjects measured in the same laboratory, and simulated localization errors were analyzed. The results show that the proposed normalization procedure improved the consistency between HRTF data sets measured on the same head in different laboratories, significantly reduced the signature of the measurement laboratory, and preserved the spatial information relevant for auditory localization.

The proposed normalization procedure can be applied to existing HRTF databases to construct an aggregated HRTF repository in which the HRTFs data sets of common subjects are perceptually similar. This will facilitate the HRTF comparison across databases and the development of future individualization methods, such as the selection of an optimal personal HRTF data set. The normalization procedure in this study can also benefit applications involving binaural synthesis by facilitating the joint use of HRTFs from different databases, improving their perceptual consistency and audio rendering quality.

This study indicates opportunities toward the definition of a common guideline for the measurement and post-processing of HRTFs. However, there are other data set properties, such as the spatial sampling grid or the type of HRTF acquisition (numerical simulations, acoustic measurements), that will require further consideration in the comparison and aggregation of various HRTF databases.

7 ACKNOWLEDGMENT

This work has been supported by the French National Research Agency through partial funding from the HAIKUS project ANR-19-CE23-0023. The work of P. M. and M. M. has been supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement 101017743 (SONICOM project).

8 REFERENCES

[1] C. Frankenbach, C. Herzog, J. Seitz, and J. Fels, "Evaluating Localization Accuracy of Dynamic Binaural Synthesis: Differences Between Individual, Estimated, and Generic HRTFs," *J. Audio Eng. Soc.*, vol. 73, no. 1/2, pp. 5–23 (2025 Jan./Feb.). https://doi.org/10.17743/jaes.2022.0190.

[2] J. C. Middlebrooks, "Virtual Localization Improved by Scaling Nonindividualized External-Ear Transfer Functions in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1493–1510 (1999 Sep.). https://doi.org/10.1121/1.427147.

[3] ITU-T, "Serie P: Telephone Transmission Quality. Objective Measuring Apparatus: Artificial Ears," Tech Rep. ITU-T Recommendation P.57 (2021 Jun.).

[4] D. Fantini, M. Geronazzo, F. Avanzini, and S. Ntalampiras, "A Survey on Machine Learning Techniques for Head-Related Transfer Function Individualization," *IEEE Open J. Signal Process.*, vol. 6, pp. 30–56 (2025 Jan.). https://doi.org/10.1109/OJSP.2025.3528330.

[5] C. Guezenoc and R. Seguier, "HRTF Individualization: A Survey," presented at the *145th Conv. of the Audio Engineering Society* (2018 Oct.), paper 10129. https://doi.org/10.17743/aesconv.2018.978-1-942220-25-1.

[6] A. Andreopoulou, D. R. Begault, and B. F. G. Katz, "Inter-Laboratory Round Robin HRTF Measurement Comparison," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 5, pp. 895–906 (2015 Aug.). https://doi.org/10.1109/jstsp.2015.2400417.

[7] P. Majdak, P. Balazs, and B. Laback, "Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions," *J. Audio Eng. Soc.*, vol. 55, no. 7/8, pp. 623–637 (2007 Jul./Aug.).

[8] S. Li and J. Peissig, "Measurement of Head-Related Transfer Functions: A Review," *Appl. Sci.*, vol. 10, no. 14, p. 5014 (2020 Jul.). https://doi.org/10.3390/app10145014.

[9] A. Kulkarni, S. Isabelle, and H. Colburn, "On the Minimum-Phase Approximation of Head-Related Transfer Functions," in Proc. of *1995 Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 84–87 (New Paltz, NY) (1995 Oct.). https://doi.org/10.1109/ASPAA.1995.482964.

[10] B. F. G. Katz and M. Noisternig, "A Comparative Study of Interaural Time Delay Estimation Methods," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. 3530–3540 (2014 Jun.). https://doi.org/10.1121/1.4875714.

[11] P. Majdak, Y. Iwaya, T. Carpentier, et al., "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," presented at the *134th Conv. of the Audio Engineering Society* (2013 May), paper. 8880.

[12] AES, "AES Standard for File Exchange – Spatial Acoustic Data File Format," AES Standard AES69-2015 (2015 Mar.).

[13] AES, "AES Standard for File Exchange – Spatial Acoustic Data File Format," AES Standard AES69-2022 (2022 Sep.).

[14] A. Andreopoulou and A. Roginska, "Towards the Creation of a Standardized HRTF Repository," presented at the *131st Conv. of the Audio Engineering Society* (2011 Oct.), paper 8571.

[15] J. M. Arend, A. Neidhardt, and C. Pörschmann, "Measurement and Perceptual Evaluation of a Spherical Near-Field HRTF Set," in Proc. of the *29th Tonmeistertagung - VDT Intl. Conv.*, pp. 356–363 (Cologne, Germany) (2016 Nov.).

[16] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, et al., "3D Tune-In Toolkit: An Open-Source Library for Real-Time Binaural Spatialisation," *PLoS One*, vol. 14, no. 3, p. e0211899 (2019 Mar.). https://doi.org/10.1371/journal.pone.0211899.

[17] X. Lu, Y. Wang, S. Jinqiu, and C. Zheng, "BiCG: Binaural Cue Generation from Unified HRTF Datasets," in Proc. of the *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 1–5, (Hyderabad, India) (2025 Apr.). https://doi.org/10.1109/ICASSP49660.2025.10890756.

[18] J. Pauwels and L. Picinali, "On the Relevance of the Differences Between HRTF Measurement Setups for Machine Learning," in Proc. of the *IEEE*

Intl. Conf. on Acoustics, Speech and Signal Processing, pp. 1–5 (Rhodes Island, Greece) (2023 Jun.). https://doi.org/10.1109/ICASSP49357.2023.10096689.

[19] A. Andreopoulou and B. F. G. Katz, "Identification of Perceptually Relevant Methods of Inter-Aural Time Difference Estimation," *J. Acoust. Soc. Am.*, vol. 142, no. 2, pp. 588–598 (2017 Aug.). https://doi.org/10.1121/1.4996457.

[20] B. Tsui and G. Kearney, "A Head-Related Transfer Function Database Consolidation Tool for High Variance Machine Learning Algorithms," presented at the *145th Conv. of the Audio Engineering Society*, 451 (2018 Oct.), paper 451.

[21] J. De Rus, J. Lopez-Ballester, M. Montagud, F. Ferri, J. López, and M. Cobos, "Impact of Input Preprocessing for HRTF Elevation Classification Over Multiple Datasets," in Proc. of the *10th Conv. of the European Acoustics Association Forum Acusticum*, pp. 2161–2168 (Torino, Italy) (2022 Sep.). https://doi.org/10.61782/fa.2023.0984.

[22] Y. Wen, Y. Zhang, and Z. Duan, "Mitigating Cross-Database Differences for Learning Unified HRTF Representation," in Proc. of the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–5 (New Paltz, NY) (2023 Oct.). https://doi.org/10.1109/waspaa58266.2023.10248178.

[23] L. S. R. Simon, N. Zacharov, and B. F. G. Katz, "Perceptual Attributes for the Comparison of Head-Related Transfer Functions," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3623–3632 (2016 Nov.). https://doi.org/10.1121/1.4966115.

[24] G. Defrance, L. Daudet, and J.-D. Polack, "Finding the Onset of a Room Impulse Response: Straightforward?" *J. Acoust. Soc. Am.*, vol. 124, no. 4, pp. EL248–EL254 (2008 Oct.). https://doi.org/10.1121/1.2960935.

[25] B. Xie, R. Dr. Ning Xiang, and J. Blauert, *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL, 2013), 2nd ed.

[26] V. Benichoux, M. Rébillat, and R. Brette, "On the Variation of Interaural Time Differences with Frequency," *J. Acoust. Soc. Am.*, vol. 139, no. 4, pp. 1810–1821 (2016 Apr.). https://doi.org/10.1121/1.4944638.

[27] S. Spagnol, K. Purkhús, S. Bjornsson, and R. Unnthorsson, "The Viking HRTF Dataset," in Proc. of the *16th Sound & Music Computing Conf.*, pp. 55–60 (Malaga, Spain) (2019 May). https://doi.org/10.5281/zenodo.3249252.

[28] A. Vidal, P. Herzog, C. Lambourg, and J. Chatron, "HRTF Measurements of Five Dummy Heads at Two Distances," in Proc. of the *Immersive and 3D Audio: From Architecture to Automotive*, pp. 1–8 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610914.

[29] V. Larcher, J.-M. Jot, and G. Vandernoot, "Equalization Methods in Binaural Technology," presented at the *105th Conv. of the Audio Engineering Society* (1998 Sep.), paper 4858.

[30] G. Theile, "On the Standardization of the Frequency Response of High-Quality Studio Headphones," *J. Audio Eng. Soc.*, vol. 34, no. 12, pp. 956–969 (1986 Dec.).

[31] H.-S. Na, C.-N. Lee, and O. Cheong, "Voronoi Diagrams on the Sphere," *J. Comput. Geom.*, vol. 23, no. 2, pp. 183–194 (2002 Sep.). https://doi.org/10.1016/S0925-7721(02)00077-9.

[32] J. C. Middlebrooks, "Individual Differences in External-Ear Transfer Functions Reduced by Scaling in Frequency," *J. Acoust. Soc. Am.*, vol. 106, no. 3, pp. 1480–1492 (1999 Sep.). https://doi.org/10.1121/1.427176.

[33] S. Singh, M. Garg, and N. Subramanian, "Estimation of the Lower Cut-Off Frequency of an Anechoic Chamber: An Empirical Approach," *Int. J. Aeroacoust.*, vol. 19, no. 1-2, pp. 57–72 (2020 Mar.). https://doi.org/10.1177/1475472X20905070.

[34] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in Proc. of the *39th DAGA*, pp. 592–595 (Merano, Italy) (2013 Mar.). https://doi.org/10.5281/zenodo.3928296.

[35] D. S. Brungart, "Auditory Parallax Effects in the HRTF for Nearby Sources," in Proc. of the *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 171–174 (New Paltz, NY) (1999 Oct.). https://doi.org/10.1109/ASPAA.1999.810877.

[36] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a Spherical-Head Model from Anthropometry," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479 (2001 Jun.).

[37] D. Romblom and B. Cook, "Near-Field Compensation for HRTF Processing," presented at the *125th Conv. of the Audio Engineering Society* (2008 Oct.), paper 7611.

[38] A. Kan, C. Jin, and A. van Schaik, "A Psychophysical Evaluation of Near-Field Head-Related Transfer Functions Synthesized Using a Distance Variation Function," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2233–2242 (2009 Apr.). https://doi.org/10.1121/1.3081395.

[39] Majdak, Piotr, Hollomey, Clara, and Baumgartner, Robert, "AMT 1.x: A Toolbox for Reproducible Research in Auditory Modeling," *Acta Acust.*, vol. 6, p. 19 (2022 May). https://doi.org/10.1051/aacus/2022011.

[40] Barumerli, Roberto, Majdak, Piotr, Geronazzo, Michele, Meijer, David, Avanzini, Federico, and Baumgartner, Robert, "A Bayesian Model for Human Directional Localization of Broadband Static Sound Sources," *Acta Acust.*, vol. 7, p. 12 (2023 May). https://doi.org/10.1051/aacus/2023006.

[41] T. McKenzie, C. Armstrong, L. Ward, D. T. Murphy, and G. Kearney, "Predicting the Colouration Between Binaural Signals," *Appl. Sci.*, vol. 12, no. 5, p. 2441 (2022 Mar.). https://doi.org/10.3390/app12052441.

[42] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness," *J. Audio Eng. Soc.*, vol. 45, no. 4, pp. 224–240 (1997 Apr.).

[43] P. Majdak, M. J. Goupell, and B. Laback, "3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training," *Atten. Percept. Psychophys.*, vol. 72, no. 2, pp. 454–469 (2010 Feb.). https://doi.org/10.3758/APP.72.2.454.

[44] T. Carpentier, H. Bahu, M. Noisternig, and O. Warusfel, "Measurement of a Head-Related Transfer Function Database With High Spatial Resolution," in Proc. of the *7th Forum Acusticum (EAA)* (Krakow, Poland) (2014 Sep.).

[45] C. Armstrong, L. Thresh, D. Murphy, and G. Kearney, "A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database," *Appl. Sci.*, vol. 8, no. 11, p. 2029 (2018 Oct.). https://doi.org/10.3390/app8112029.

[46] M. Warnecke, S. Clapp, Z. Ben-Hur, D. Lou Alon, S. V. Amengual Gari, and C. P., "Sound Sphere 2: A High-resolution HRTF Database," in Proc. of the *5th AES Intl. Conf. on Audio for Virtual and Augmented Reality* (Redmond, WA) (2024 Aug.), paper 31.

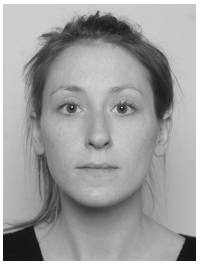
[47] R. Baumgartner, P. Majdak, and B. Laback, "Modeling Sound-Source Localization in Sagittal Planes for Human Listeners," *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802 (2014 Aug.). https://doi.org/10.1121/1.4887447.

[48] ISO, "Acoustics – Methods for Calculating Loudness – Part 2: Moore-Glasberg Method." Tech. Rep. 532-2:2017 (2017 Jun.).

[49] R. Daugintis, R. Barumerli, L. Picinali, and M. Geronazzo, "Classifying Non-Individual Head-Related Transfer Functions With A Computational Auditory Model: Calibration And Metrics," in Proc. of the *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, pp. 1–5 (Rhodes Island, Greece) (2023 Jun.). https://doi.org/10.1109/icassp49357.2023.10095152.

[50] P. Majdak, F, Zotter, F. Brinkmann, et al. "Spatially Oriented Format for Acoustics 2.1: Introduction and Recent Advances, J. Audio Eng. Soc., vol. 70, no. 7/8, pp. 565–84 (2022 Jul./Aug.). https://doi.org/10.61782/fa.2023.0729.

THE AUTHORS


Hélène Bahu


Jean-Marc Jot


Thibaut Carpentier


Markus Noisternig


Michael Mihocic


Piotr Majdak


Olivier Warusfel

Hélène Bahu completed her master's degree in engineering sciences at the Université Pierre et Marie Curie in 2013. She then conducted her Ph.D. at the Institut de Recherche et de Coordination Acoustique/Musique (IRCAM, Paris) on auditory localization in the context of nonindividual binaural synthesis and obtained her diploma from Sorbonne University (Paris) in 2016. During this time she measured the BiLi head-related transfer function (HRTF) database and conducted several auditory localization experiments. She then worked as a researcher at Dysonics and then Google (California), where she specialized in the design of spatial audio rendering algorithms for power-limited devices. She now leads the HRTF normalization project at IRCAM in collaboration with experts from international universities and industries.

•

Jean-Marc Jot is founder and principal of Virtuel Works LLC in Aptos, California. He received his Ph.D. from Telecom Paris in 1992. From 1992 to 1998, he was an audio research scientist at Institut de Recherche et de Coordination Acoustique/Musique (IRCAM) in Paris. After relocating to the United States in 1998, he led the research and development of sound processing technologies, platforms and standards for virtual and augmented reality, gaming, broadcast, cinema, and music creation (with Creative Labs, DTS, Magic Leap, and iZotope). He is a fellow of the Audio Engineering Society (AES), cochair of the AES Technical Committee on Spatial Audio, and member of the AES Technical Committee on Machine Learning and Artificial Intelligence. His research interests include audio effects, room acoustics, and spatial audio signal processing and coding.

•

Thibaut Carpentier is an R&D engineer at IRCAM (Institut de Recherche et de Coordination Acoustique/Musique), Paris. After studying acoustics at the École Centrale and signal processing at the École nationale supérieure des télécommunications (ENST) Telecom Paris, he joined the French National Centre for Scientific Research (CNRS) and the Acoustics and Cognition Research Group in 2008. His research is focused on spatial audio.

•

Markus Noisternig studied electrical and audio engineering at the Graz University of Technology (TUG) and computer music composition at the University of Music and Performing Arts Graz (KUG). He received his Ph.D. in acoustics, audio signal processing, and psychoacoustics from the KUG. Since 2008, he has been a researcher at the Science and Technology of Music and Sounds (STMS) Laboratory (UMR 9912), a joint research unit of the Institute for Research and Coordination in Acoustics/Music (IRCAM), the French National Centre for Scientific Research (CNRS), Sorbonne University (SU), and the French Ministry of Culture. In 2025, he was appointed deputy director of the STMS Laboratory. He also serves as a council member of the Technical Committee on Audio Signal Processing of the European Acoustics Association (EAA). His research focuses on acoustics and audio signal processing, immersive audio for virtual and augmented reality, and room acoustics.

•

Michael Mihocic holds degrees in "Electronics" and "Innovation and Technology Management." Since 2005, he has been a researcher at the Acoustics Research Institute (ARI) of the Austrian Academy of Sciences (OeAW) focusing on spatial audio and sound localization within the Hearing group. His work involves developing specialized software, standards, and toolboxes, notably the spatially oriented format for acoustics (SOFA) and Auditory Modeling Toolbox (AMT). He designs and conducts psychoacoustical experiments and performs acoustical measurements in several fields (humans, loudspeakers, rooms, traffic). Michael also maintains the ARI's laboratory, ensuring optimal research infrastructure. He is an active member of the audio engineering community, serving on the board of the Austrian section of the Audio Engineering Society (AES).

•

Piotr Majdak studied electrical and audio engineering at the University of Technology and the University of Music and Performing Arts (KUG), in Graz, Austria. In 2008, he received his Ph.D. degree in "psychoacoustics and signal processing." His habilitation followed in 2015 in the field of "acoustics and audio engineering" at the KUG, where he regularly teaches. Since 2002, he has worked at the Acoustics Research Institute (ARI) of the Austrian Academy of Sciences (OeAW), where he leads the research group "Binaural Audio and Auditory Modeling." Piotr works at a better understanding of the mechanisms underlying spatial hearing and at applying that knowledge to improve acoustic and audio systems. He has published over 70 peer-reviewed journal articles, proceedings, and book chapters. He is the chair of the technical committee (TC) for Psychological and Physiological Acoustics of the European Acoustics Association and the president of the Austrian section of the Audio Engineering Society.

•

Olivier Warusfel studied electronic engineering at Ecole Nationale Supérieure d'Electronique et Applications (France) and acoustics at Université Paris 6, where he received his Ph.D. degree in 1985. He has been lecturing in musical acoustics at University Paris 8 and room acoustics at University Paris 6. Since 1992, he has been leading the Room Acoustics group (now called "Espaces Acoustiques et Cognitifs") at Institut de Recherche et de Coordination Acoustique/Musique (IRCAM). His main interests focus on audio signal processing, sound spatialization, room acoustics, auditory virtual reality, and spatial auditory perception and cognition.