On the Use of a Hybrid Linear-ERB Frequency Scale for Lightweight Simulation of HRTFs

[1st] Maurício do Vale Madeira da Costa
MTDML, IMM
University of Osnabrück
Osnabrück, Germany
madovalemade@uni-osnabrueck.de

[2nd] Luiz Wagner Pereira Biscainho
DEL/Poli & PEE/COPPE
Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
wagner@smt.ufrj.br

[3rd] Michael Oehler
MTDML, IMM
University of Osnabrück
Osnabrück, Germany
michael.oehler@uni-osnabrueck.de

***Abstract*—This paper explores a lightweight method for numerical simulation of HRTFs based on a hybrid linear-ERB frequency scale applied to the boundary element method. The HRTFs simulated according to our modified approach were evaluated by comparison with HRTFs acoustically measured and simulated using the standard linear scale. The HUTUBS dataset was used to assess differences in magnitude, interaural phase, and interaural time. We explored two approaches for phase at high frequencies: interpolation and extrapolation. The results suggest that extrapolation provides better overall agreement of the phase response in comparison to acoustic measurements and that using at least two bins per auditory band is sufficient to provide seamless approximations of the HRTFs simulated with the standard approach, with computing time savings of around 86%.**

***Index Terms*—HRTF, BEM, numerical simulation, HUTUBS**

I. Introduction

In recent years, in the field of the Internet of Sounds (IoS) [1], significant attention has been directed towards the development of networked devices for virtual and augmented realities, where overall audio quality is crucial for ensuring an immersive and realistic experience. More specifically in the field of the Internet of Musical Things (IoMT) [2], interfaces for remote music expression have been shown to benefit from personalized binaural audio reproduction [2]–[6].

Techniques for measuring the effect of the person's body geometry over acoustic waves, which can be mathematically modeled by head-related transfer functions (HRTFs), have driven extensive research. Alternatives to classic acoustic measurement procedures to make the process faster and more accessible have been explored, including individualization based on anthropometric data, numerical simulation, and indirect individualization based on perceptual feedback (see reviews [7], [8]). A current trend focuses on hybrid approaches based on machine learning techniques that produce HRTFs by taking as input 3D head/torso scans, ear images, and anthropometric measurements [7], [9].

HRTF individualization methods based on numerical simulations are still an attractive option, fostered by the popularization of 3D scanning devices that, paired with a computer and some pieces of software to process the acquired data, can be used to produce personal HRTFs [10], [11]. The most relevant and widely adopted method for numerical simulations of HRTFs is the so-called boundary element method (BEM) [12], in which the acoustic field is estimated around a surface (in this case, the subject's geometry) in the frequency domain. For each desired angular position around the subject, the transfer function obtained by solving this problem for regularly spaced frequencies consists in the HRTF, which can, if necessary, be transformed by the inverse Fourier transform into its time-domain counterpart, the head-related impulse response (HRIR). One of the most critical shortcomings of those methods is still the high computational effort required to estimate HRTFs within the full audible range, with high frequencies accounting for a very high proportion of the computing time spent [10], [11], [13], [14].

Recently, as an alternative to this solution, an approximation method was proposed in [15] that simulates HRTFs using the BEM to compute frequency bins distributed on a hybrid nonlinear scale, which starts linear up to a point f_c and then switches to a logarithmic distribution, spacing the frequency bins progressively. Considering that the computing time required to estimate the acoustic field grows exponentially with frequency [15], this approach exploits the roughly logarithmic frequency resolution of the auditory system (see [16], [17]) at high frequencies to save computing time by coarsening the resolution of transfer functions at high frequencies.

In the method described in [15], a linear interpolation is performed to transform the magnitude samples back to the regular frequency scale. As for the phase, since it evolves within a periodic domain, the progressively large frequency interval between the samples to be simulated is likely to produce aliasing during the unwrapping procedure.[1] A solution given in [15] to solve this problem consists in using the originally simulated phase values only for the regularly spaced portion of the spectrum, and then using the group delay calculated for these samples to linearly extrapolate the phase for samples above f_c. However, this method involves a compromise in the selection of f_c: low values of f_c result in higher savings in

This work was funded by Volkswagen Foundation (Volkswagen Stiftung) Germany, under Grant no. 96 881, and the National Council for Scientific and Technological Development (CNPq), under Grant 311146/2021-0.

[1] The aliasing effect happens if there is a phase difference larger than π between two consecutive samples simulated in the frequency domain.

computing time at the cost of worsening (or oversimplifying) the phase response.

In the present paper, we explore some modifications to this method. Firstly, we propose using the equivalent rectangular bandwidth (ERB) frequency scale [17], which is based on auditory filters, instead of the logarithmic scale; this allows potential users to define their desired frequency scale directly in perceptual terms, in bins/ERB. In addition, we decouple the crossover frequency f_c from the point in frequency at which the phase extrapolation begins and study the impact of the aforementioned aliasing effect on the phase response. Besides, in this paper, we propose to set f_c automatically at the frequency where the linear and ERB resolutions meet, thus ensuring a smooth transition in resolution between the different frequency scales and a minimum computing cost for the desired resolutions. This also aims to facilitate method configuration, requiring the definition of one less parameter.

To further expand the results found in [15], this paper presents a variety of numerical assessments conducted using the HUTUBS database [18], which provides, among other data, both acoustically measured and simulated HRTFs. This allowed us to measure the relevance of the spectral distortions (in magnitude and phase) caused by the approximation technique explored, also using acoustically measured HRTFs as a reference. In particular, we assessed: (i) computing time; (ii) magnitude spectral difference; (iii) interaural phase difference; and (iv) interaural time difference.

The paper is organized as follows. In Section II, the methodology of this work is introduced, comprising a theoretical background of numerical simulation of HRTFs followed by a description of the non-linear sampling method; then, in Section III, the database HUTUBS, which was used in our evaluations, is described; subsequently, the numerical assessment carried out and the results obtained are presented in Section IV, followed by a discussion in Section V; finally, we conclude this paper in Section VI.

II. Methodology

A. Numerical Simulation of HRTFs in the Frequency Domain

In this paper, as in [15], we focus on numerical simulations of HRTFs in the frequency domain, particularly using the BEM. Firstly, let the HRTF be defined as set of complex sound pressure bins in the frequency domain. Provided the 3D head model is aligned with the x axis [10], the receivers are approximately on the y axis, and the vertical axis of the head is parallel to the z axis, a (left, right) HRTF pair related to an acoustic path from a location $\mathbf{x}^*$ until the left and right ear canals (or eardrums) is described as

$$\begin{aligned} \mathrm{H_L}[\mathbf{x}^*, f] &= \frac{p_\mathrm{L}[\mathbf{x}^*, f]}{p_0[f]} \quad \text{and} \\ \mathrm{H_R}[\mathbf{x}^*, f] &= \frac{p_\mathrm{R}[\mathbf{x}^*, f]}{p_0[f]}, \end{aligned} \tag{1}$$

where p_L and p_R denote the complex sound pressure at the left and right receiver points, respectively, and f denotes the frequency. The reciprocity principle [10] can be used for saving computations by swapping source and receiver positions during simulation. An acoustic normalization is performed w.r.t. the reference complex sound pressure p_0, measured at the origin of the Cartesian plane *in the absence of the head*; a constant source distance to the origin is assumed since all source positions are distributed on the surface of a sphere that surrounds the subject.

The BEM approximates the free-field sound propagation in the frequency domain around an object of interest using the Helmholtz equation [12], producing a complex pressure value for a given source/receiver pair and a given frequency. In our case, a 3D surface that represents the listener's geometry in the discrete domain is used, namely a '3D mesh', being described as a set of points in space that define triangular faces. The mesh can be acquired using a variety of different equipment, including some modern smartphones. The sound field is then independently approximated for frequencies within the audible spectrum in regularly spaced frequency steps F_s [11], [12].

This way, this solution can produce filters in the discrete Fourier domain, which can be done by mirroring the complex conjugate of the estimates for negative frequencies and including a sample with value 1 (0 dB) at 0 Hz. As in all sampling procedures, the Nyquist theorem indicates the range of possible values for F_s, which is upper bounded to avoid potential aliasing induced by fast variations present in the functions to be sampled (in this case, in the frequency domain). For a time sampling rate of r_s, the set of regularly spaced frequencies f can be formally described as

$$\mathcal{F} = \{f \mid 0 \leq f \leq f_\mathrm{max}, f_{i+1} - f_i = F_\mathrm{s}\}, \tag{2}$$

where $f_\mathrm{max} = r_\mathrm{s}/2$. Finally, a time delay can be added to the entire set of HRTFs to ensure causality [13].

B. Numerical Simulation of HRTFs in Non-linear Frequency Scales

An approximation technique has been proposed in [15] whose main idea is to simulate HRTFs using a hybrid frequency scale that progressively decreases frequency resolution, thus reducing the computational load inherent to this procedure. It takes advantage of the nearly logarithmic nature of the spectral resolution of the human auditory system [16], [17] to avoid the exponential growth in frequency of the computing time required to produce the complex pressure values. Therefore, the high-frequency range, which represents a very high proportion of the total computing time spent to compute the entire spectrum, would be sampled at wider intervals between frequency bins.

The linear-logarithmic frequency scale proposed in [15] consists in using a fixed frequency step F_s up until a crossover frequency f_c, after which a logarithmic frequency spacing, defined by a number B of bins/octave, takes place. This scale can be denoted as $\hat{\mathcal{F}} = \{\mathcal{F}_\mathrm{lin}, \mathcal{F}_\mathrm{log}\}$, where

$$\begin{aligned} &\mathcal{F}_\mathrm{lin} = \{f \mid 0 \leq f \leq f_\mathrm{c}, f_{i+1} - f_i = F_\mathrm{s}\} \quad \text{and} \\ &\mathcal{F}_\mathrm{log} = \{f \mid f = f_\mathrm{max} 2^{-l/B}, 0 \leq l \leq B \log_2 (f_\mathrm{max}/f_\mathrm{c})\}. \end{aligned} \tag{3}$$

In this paper, a variant of this frequency scale is used by replacing the logarithmic part with the ERB frequency scale, widely adopted to model the resolution of the human auditory filters [17], defined as

$$\mathrm{ERB}(f) = 0.108f + 24.7, \tag{4}$$

where $\mathrm{ERB}(f)$ is the bandwidth of the auditory filter around frequency f, both in Hz. As in [15], $f_{\max}$ is the reference frequency from which the remaining frequency bins will be defined in descending order, now following the ERB scale. By indicating the frequency resolution as E bins per auditory frequency band, this modification is intended to provide the potential user with a perceptually related dimension.

The linear-ERB frequency scale, denoted by $\hat{\mathcal{F}} = \{\mathcal{F}_{\mathrm{lin}}, \mathcal{F}_{\mathrm{ERB}}\}$, then indicates the frequencies in which the set of HRTF pairs $\left(\hat{\mathrm{H}}_{\mathrm{L}}[\mathbf{x}^*, \hat{f}], \hat{\mathrm{H}}_{\mathrm{R}}[\mathbf{x}^*, \hat{f}]\right)$ will be simulated. Differently from [15], in this work we assume that the hybrid scale crossover frequency will automatically be set by the frequency where the chosen resolutions of the linear and the ERB scales meet, resulting in a scale with a maximum resolution of F_{s}.

As the frequency "bins" were irregularly distributed, they do not lend themselves to the fast processing via FFT provided by finite-length impulse response (FIR) digital filters. A simple solution given in [15] is to create sets of FIR filters by converting $\hat{\mathrm{H}}[\mathbf{x}^*, \hat{f}]$ into $\hat{\mathrm{H}}[\mathbf{x}^*, f]$ via magnitude interpolation and phase extrapolation. Since FIR filters map to a regular frequency scale in the digital domain, the HRTFs will be henceforth denoted as $\mathrm{H'}[\mathbf{x}^*, k]$, where $k \in \mathcal{K} \triangleq \{0, 1, 2, ..., K-1\}$ is the frequency index in the discrete frequency domain.

Due to the periodic nature of the phase, its interpolation between progressively more spaced samples might lead to aliasing, as mentioned above. In [15], the solution explored to avoid such a problem is the upward extension of the average group delay of the simulated spectrum below f_{c}, defined as

$$d(\hat{\mathrm{H}})_{[k_{\mathrm{c}}]} = \frac{1}{k_{\mathrm{c}}} \sum_{0 \leq k < k_{\mathrm{c}}} \angle\hat{\mathrm{H}}[k+1] - \angle\hat{\mathrm{H}}[k], \tag{5}$$

where $k_{\mathrm{c}} = \lfloor f_{\mathrm{c}}/F_{\mathrm{s}} \rfloor$ is the index of the digital frequency related to f_{c}, and $\angle\hat{\mathrm{H}}[k]$ is the unwrapped phase of $\hat{\mathrm{H}}$ at frequency k. Variable $\mathbf{x}^*$ has been omitted to simplify notation and $\lfloor . \rfloor$ denotes the floor operator. The resulting phase $\angle\mathrm{H'}[\mathbf{x}^*, k]$ is then comprised of the combination of the originally simulated phase up until k_{c} followed by a linear extrapolation of $\angle\hat{\mathrm{H}}[k]$ from k_{c} upwards.

In the present work, however, our approach for phase extrapolation has to be slightly different due to the fact that we are forcing the crossover point between the different frequency scales to be as low as possible and also allowing the frequency f_{e} at which we want to start the phase extrapolation to be freely chosen. Whenever $f_{\mathrm{c}} < f_{\mathrm{e}}$, an interpolation will be performed from f_{c} to f_{e}.

III. Dataset

We used the HUTUBS database [18] for all evaluations, which is publicly available and comprises HRTFs of 96 subjects measured using an acoustic full-sphere measurement system. Within this pool of subjects, 93 are from different human subjects, 10 female and 83 male; the mean age of the subjects was 36 years (SD 9 years). Two subjects were measured twice for evaluation purposes and measurements of a custom-made dummy head are also included.

Additionally, the dataset includes corresponding numerically simulated HRTFs for all subjects, enabling comparisons between measurement and simulation techniques. The dataset also contains 25 anthropometric features for each subject and several high-resolution head meshes. Finally, the dataset includes Headphone Transfer Functions (HTRFs) for two headphone models. All HRTFs are stored in SOFA files [19] along with their spatially continuous representation in the form of spherical harmonics (SH) [20], which are stored in Matlab files.

A. Numerical Simulations

First, the meshes obtained were gradually downsampled from 1 mm at the simulated ear to 10 mm at the opposite ear using the method described in [21] (plugin available in Mesh2HRTF [13]), resulting in meshes having around 14,000 to 20,000 elements [18]. Then, numerically simulated HRTFs were produced using Mesh2HRTF for the frequency range between 100 Hz and 22 kHz in steps of 100 Hz. The HRTFs were sampled on a $Q = 1730$ point Lebedev grid [22], on a sphere of radius 1.47 m. The HRTFs were referenced and normalized by dividing the pressure at the ear canal by the pressure at the center of the head in its absence [13]. Afterwards, the transfer functions were obtained following the same procedure presented in the previous section; HRIRs were then obtained via inverse Fourier transform followed by a circular shift of 60 samples to ensure causality for all filters. As a postprocessing procedure, HRIRs were resampled to a sampling rate of 44.1 kHz and truncated to 256 samples using a 10 sample fade-in and a 20 sample fade-out, both applying the squared sine function. Spherical harmonics representations of the simulated HRTFs were also computed and made available via an SH transform of order 35.

B. Acoustic Measurements

HRTFs were also acoustically measured for all subjects. Such measurements were conducted in the anechoic chamber of the Technical University Berlin using 37 speakers mounted in a circular structure with a working distance of 1.47 m from the center of the array to the membrane of the speakers, which resulted in a final resolution of 5° in elevation. Subjects used a pair of custom made in-ear microphones [23] and were sitting steadily on a turning chair equipped with a motor, with their head located at the center of the circular array. The measurements were then taken while the subjects were under a continuous rotation (one full revolution per minute) and normalized least mean squares (NLMS) adaptive filters [24] were used to post-process the data, resulting in quasi-continuous HRIRs in azimuth. To create the final set of HRTFs, the resolution in azimuth was chosen to yield an almost constant great circle distance between neighboring points of the same elevation [18], resulting in a full-spherical sampling grid with $Q = 440$ points.

To compensate for the poor low-frequency response of the loudspeakers, the numerically simulated HRTFs were used to extrapolate the frequency response (both in magnitude and phase) using 4th order Linkwitz-Riley cross-over filters with a –6 dB cut-off frequency at 300 Hz [18]. Spatially continuous HRTF representations were also made available, using an SH transform of order $N = 16$, in this case. All measurements were made using a sampling rate of 44.1 kHz.

IV. Evaluations

The numerical assessments conducted in this work are designed to measure the impact of using the lin.-ERB frequency scale to simulate HRTFs. By using the HUTUBS, we were able to compare simulated HRTFs using our proposed non-linear scale with both the original simulations (i.e., HRTFs simulated on the regularly sampled frequency scale) and the acoustically measured HRTFs. Such comparisons can provide further insights into the relevance of the distortions caused by the presented approximation technique. As in [15], the method described and all numerical experiments were implemented in Python, by modifying the code of Mesh2HRTF[2] according to our needs.

A. Computing Time

To assess the savings in computing time, we first run numerical simulations using the 3D meshes available to calculate a reference average computing time per frequency bin $\bar{T}[f]$, related to $f \in \mathcal{F}$ (the standard frequency scale). The computing time per frequency $T[\hat{f}]$, in the non-linear scales $\hat{f} \in \hat{\mathcal{F}}$, is then approximated by linearly interpolating the reference $\bar{T}[f]$, as performed in [15]. Finally, the relative computing time up to a predefined maximum frequency F is then calculated as

$$\tau[F] = \frac{\sum_{f=0}^{F} T[\hat{f}]}{\sum_{f=0}^{F} \bar{T}[f]}. \tag{6}$$

Since HRTFs can be simulated for different frequency ranges, typically varying only the top frequency F, this figure of merit is useful to assess the savings in processing time for different choices of F. Such a calculation was carried out for HRTFs computed using a resolution of 100 Hz for the linear portion of the scales, $B = \{3, 6, 12, 18, 24\}$ bins/octave for the log scales, and $E = \{0.5, 1, 2, 3, 4\}$ bins/ERB for the ERB scales.

The estimated average relative computing times $\tau[f_{\max}]$ are shown in Figure 1 for both lin.-log and lin.-ERB frequency scales, for comparison. The HRTFs computed with the standard linear frequency scale are taken as a reference, i.e. 100% for all frequencies. It is worth mentioning that the savings observed differ from results presented in [15] because the simulations in the HUTUBS dataset use a frequency interval of 100 Hz, whereas in [15] this value was 150 Hz. This difference results in greater savings in the evaluations carried out here.

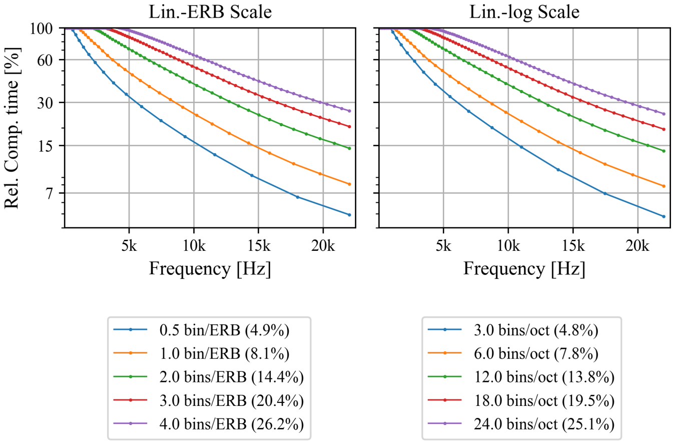

Fig. 1. Relative total computing time (%) for simulations using lin.-log and lin.-ERB resolutions. Maximum resolution of 100 Hz and results reported in legend for $F = f_{\max} = 22$ kHz).

The different frequency scales provide similar savings in processing time.[3] Interestingly, the assessments presented in [15] suggested that $B = 6$ bins/octave would be close to a minimum resolution that could yield imperceptible spectral distortion, which indicates one should use $E > 1$ bin/ERB in order to have HRTFs that are perceptually equivalent to the ones computed using the linear frequency scale. In this case, the relative computing time is $\approx 8\%$ (savings of $\approx 92\%$). If a lower top frequency were adopted for the simulation, e.g. $F = 16$ kHz, the total relative cost in this case would grow to $\approx 14\%$. For a higher resolution, e.g. using $B = 18$ bins/octave and $E = 3$ bins/ERB, the relative processing time is $\approx 20\%$ (savings of $\approx 80\%$), which is already remarkably cheap. The next evaluations will be performed for lin.-ERB frequency scales only, given the verified equivalence in resolution.

B. Magnitude Distortion

The impact of the lin.-ERB frequency scale on magnitude spectral distortion was measured using the log-spectral difference in auditory filters, as conducted in [18]. To this end, the energy of each HRTF $\mathrm{H}[\mathbf{x}_d, k]$ was integrated in bands using the ERB scale, producing $\mathrm{G}[\mathbf{x}_d, k_{\mathrm{ERB}}]$, where k_{ERB} indexes the center frequencies $f \in \mathcal{F}_{\mathrm{ERB}}$ in an ERB scale with resolution 1. The absolute energetic spectral differences between a HRTF$_1$ and a reference HRTF$_2$, in ERB scale, were then computed as

$$D[d, k, \Gamma] = \left| 10 \log_{10} \frac{|\mathrm{G}_1[\mathbf{x}_d, f, \Gamma]|}{|\mathrm{G}_2[\mathbf{x}_d, f, \Gamma]|} \right|, \tag{7}$$

where $|.|$ denotes the absolute value and Γ indexes the subject. By averaging $D[d, k, \Gamma]$ over d and all subjects Γ, the average distortion per frequency $\bar{D}[k]$ is computed. Also as conducted

[2] Available online at https://www.mesh2hrtf.org/

[3] In fact, at high frequencies, an ERB scale of 1 bin/ERB is best approximated by a log scale having 7 bins/octave, hence the slightly lower proportional savings yielded by the lin.-log scales with multiples of 6 bins/octave.

in the HUTUBS publication [18], only directions within and above the horizontal plane were considered. We used the same values for resolutions E bins/ERB mentioned in the computing time assessment.

Results of the first evaluation are shown in Fig. 2, which shows spectral differences between the HRTFs simulated using the proposed lin.-ERB scale and the original simulations. Spectral differences below $0.4\,\text{dB}$ are observed for all frequency bands when using $E \geq 2\,\text{bins/ERB}$. For $E = 1\,\text{bin/ERB}$, the spectral difference reaches a maximum of around $1.5\,\text{dB}$, for the frequency range above $10\,\text{kHz}$.

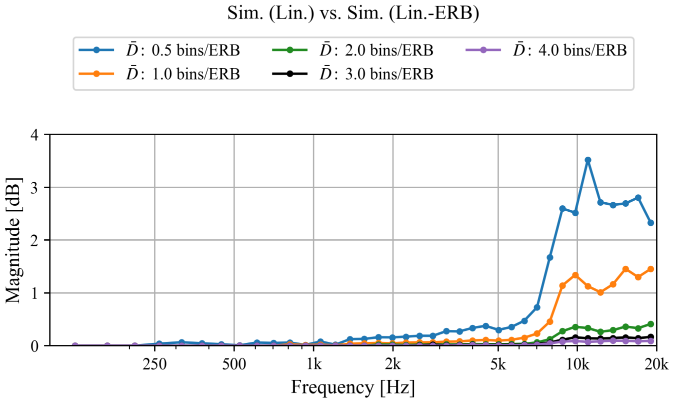

Fig. 2. Spectral differences in auditory bands between the HRTFs simulated with the lin.-ERB scale and the originally simulated HRTFs.

Spectral differences between the acoustically measured HRTFs and their simulated counterparts (both using the original linear and the hybrid lin.-ERB frequency scales) were also computed, and are presented in Fig. 3. This result closely reproduces the spectral differences reported in [18] for acoustically measured HRTFs vs. simulated HRTFs. In comparison to the acoustic measurements, which are usually considered more accurate [7], all HRTFs simulated using $E \geq 1\,\text{bin/ERB}$ presented nearly identical spectral differences to the HRTFs simulated using the standard linear scale, suggesting that the spectral differences that can be observed between simulations within this range of resolutions are irrelevant to their ability to approximate the acoustically measured HRTFs.

C. Interaural Phase Difference

In the next evaluation, we assessed differences between the HRTFs simulated using the linear frequency scale and: (i) HRTFs simulated using the lin.-ERB frequency scale whose phase was reconstructed by *linear interpolation* for the entire frequency spectrum; and also (ii) HRTFs whose phase for frequencies above $5\,\text{kHz}$ was *linearly extrapolated* according to the average group delay below this frequency, as done in [15]. An example of the different approaches for phase mentioned is illustrated in Fig. 4, including, as a reference, the acoustically measured HRTF.

In this plot, the problem with performing linear interpolation of the unwrapped phase becomes visible when comparing the phase calculated with $E = 0.5\,\text{bin/ERB}$ with the original simulation ('Sim.'). After $15\,\text{kHz}$, the unwrapped phase deviates from the reference due to aliasing, i.e. the absolute phase difference between consecutive samples was larger than π. In this particular example, the phase values calculated using $E \geq 1\,\text{bin/ERB}$ did not show large deviations of this kind, although it may happen for other HRTFs. It can be noticed that the phase response obtained by the acoustic measurement is considerably different from the numerically simulated one. Finally, the straight line produced for frequencies above $5\,\text{kHz}$ in the extrapolated phase has a slope reasonably similar to the global slope of the interpolated phases, showing some agreement between the group delay of the frequency range below $5\,\text{kHz}$ and the global group delay.

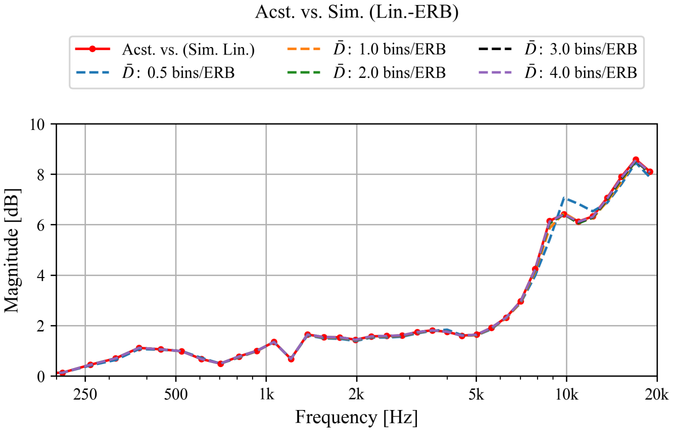

Fig. 3. Spectral differences in auditory bands between the HRTFs simulated with the lin.-ERB scale and the acoustically measured HRTFs (dashed lines); originally simulated HRTFs are used as a benchmark (solid line).

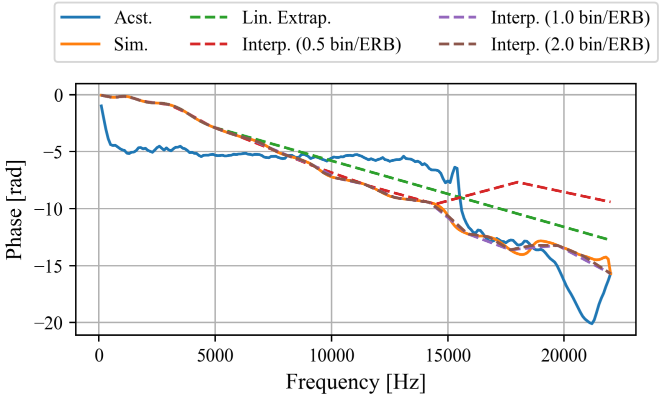

Fig. 4. Unwrapped phase responses for an HRTF: acoustically measured ('Acst.'); standard simulation ('Sim.'); simulations using the lin.-ERB frequency scale, which has interpolated phase ('Interp. (E bins/ERB)'); and standard simulation with extrapolation above $5\,\text{kHz}$ [15].

While large deviations in phase response might affect timbre perception, localization is mainly affected by interaural phase differences (IPD). The $|\Delta\text{IPD}|$ between a given HRTF_1 and a reference HRTF$_2$ can be calculated as

$$|\Delta\mathrm{IPD}[k]| = |\mathrm{IPD}(\mathrm{H}_1[k]) - \mathrm{IPD}(\mathrm{H}_2[k])|\,, \text{where} \quad (8)$$

$$\mathrm{IPD}(\mathrm{H}[k]) = \angle\mathrm{H}_\mathrm{L}[k] - \angle\mathrm{H}_\mathrm{R}[k]. \quad (9)$$

We first compared the $|\Delta\mathrm{IPD}|$ averaged across all HRTFs above and within the horizontal plane, for all subjects, having the HRTF simulated with the standard linear scale as a reference. Fig. 5 presents the resulting $|\Delta\mathrm{IPD}|$ for the linearly extrapolated phase (‘Sim. lin. extrp.’) and for the interpolated unwrapped phase values (‘Sim. ERB’). The results show a clear advantage in extrapolating the phase for the HRTFs simulated with $E \leq 1$ bin/ERB. When using $E \geq 2$ bins/ERB, on the other hand, the results are favorable to the use of interpolation, suggesting that the aforementioned aliasing effect is occurring much less frequently, and possibly the remaining errors observed are caused by the linear interpolation alone.

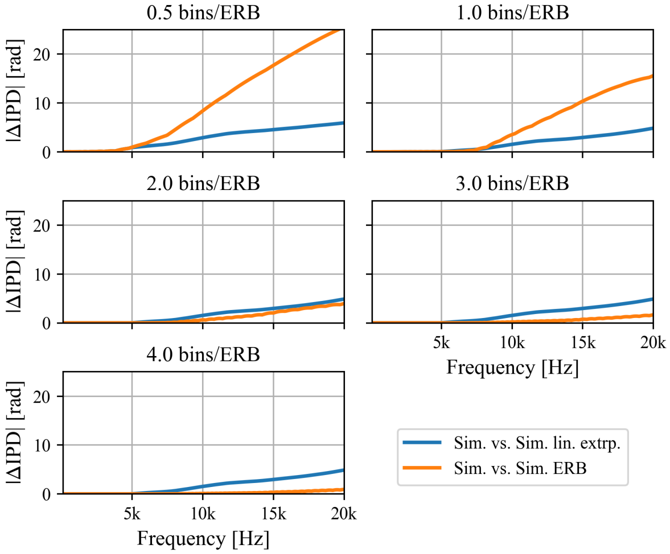

Fig. 5. Average ΔIPD for HRTFs simulated using the hybrid lin.-ERB frequency scale, for extrapolated and interpolated phase responses; HRTFs simulated with linear frequency scale are used as benchmarks.

The same procedure was followed to assess the average $|\Delta\mathrm{IPD}|$ using the acoustically measured HRTFs as references, and the results are shown in Fig. 6. Interestingly, the results suggest that the IPD obtained by linearly extrapolating the phase simulated up to 5 kHz presents better agreement with the IPD of the acoustically measured HRTFs in all scenarios, even having a slight advantage over the HRTFs simulated with the linear frequency scale, above 15 kHz. The results for the HRTFs with extrapolated phase remain practically unchanged for $E \geq 1$ bin/ERB. As for the HRTFs using interpolated phase, their IPD get progressively better with E, and get nearly identical to the HRTFs simulated with linear frequency scale up to 15 kHz, for $E \geq 2$ bins/ERB.

D. *Interaural Time Difference*

Finally, temporal differences between the HRTFs were measured as the broadband interaural time difference (ITD), for HRTFs within the horizontal plane. The ITDs were indicated by the onsets of each HRIR, following the same procedure as in [18]. Low-pass versions of the impulse responses were obtained using an 8th order Butterworth filter to only take into account frequencies below 3 kHz. The onset times were then defined as the instant at which a threshold of -20 dB (measured from the absolute maximum value of the specific HRIR) is exceeded. To increase temporal resolution, an up-sampling of 10 times was performed. The values of $|\Delta\mathrm{ITD}|$ were then computed as the absolute differences between the ITDs of different pairs of HRTFs.

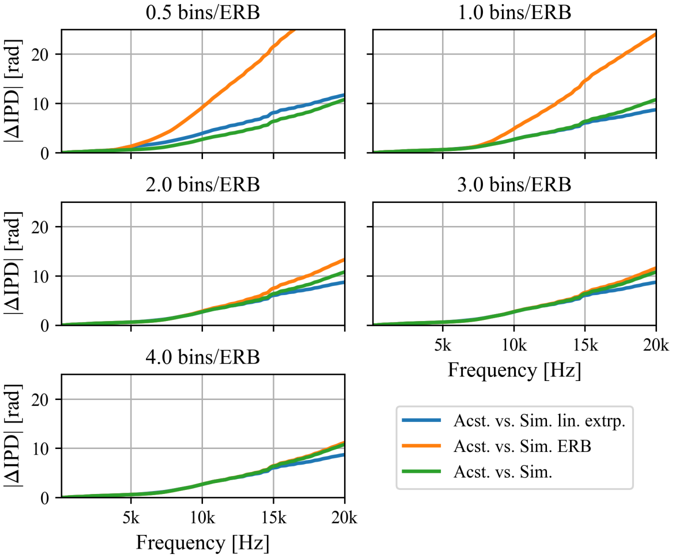

Fig. 6. Average interaural phase difference for HRTFs simulated with hybrid lin.-ERB frequency scale, having linearly extrapolated and linearly interpolated phase responses, and for HRTFs simulated with linear frequency scale; acoustically measured HRTFs are used as references.

The results are presented in Fig. 7 for the HRTFs simulated with the hybrid frequency scale having the HRTFs simulated using the linear frequency scale as a reference, with the just noticeable difference (JND) [25] being indicated in dashed lines. Since the discrepancies in IPD only occur at frequencies above 3 kHz, with an exception for the simulations using $E \geq 0.5$ bin/ERB, virtually zero $|\Delta\mathrm{ITD}|$ is observed for all simulations using $E \geq 1$ bin/ERB. For the HRTFs simulated using $E = 0.5$ bin/ERB, some occasional deviations surpassed the JND threshold, mainly at lateral angles.

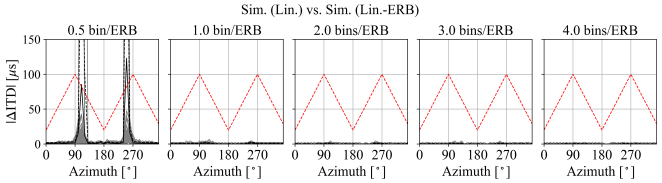

Fig. 7. Comparisons of interaural time differences in the horizontal plane for simulated HRTFs: linear vs. lin.-ERB frequency scales. Individual results (gray solid); average (black, solid) and std. (black, dashed); JND (red, dashed).

Fig. 8 depicts the results for ITDs of the HRTFs simulated with the hybrid frequency scale and having the phase linearly extrapolated compared to the ITDs of the HRTFs simulated using the linear frequency scale. As should be expected, the results are virtually the same as the ones presented above, as the phase is only extrapolated from 5 kHz upwards and the impulse responses are low-passed at 3 kHz.

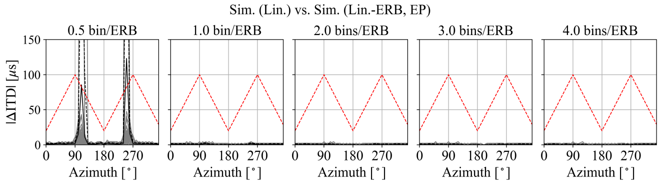

Fig. 8. Comparisons of interaural time differences in the horizontal plane for simulated HRTFs: linear vs. lin.-ERB frequency scales, the latter having the phase extrapolated from 5 kHz upwards. Individual results (gray solid); average (black, solid) and std. (black, dashed); JND (red, dashed).

The same assessment was repeated having the acoustic measurements as references; the results are illustrated in Figs. 9 and 10. The benchmark result is placed on the left side (Ref.), and corresponds to $|\Delta\text{ITD}|$ for the standard simulation, replicating the results presented in [18]. As expected from the previous evaluations, aside from the HRTFs simulated with $E = 0.5\,\text{bin/ERB}$, the $|\Delta\text{ITD}|$ measured are identical to the ones found for the reference HRTFs, with only a few HRTFs surpassing the JND limits, for the same lateral directions mentioned above.

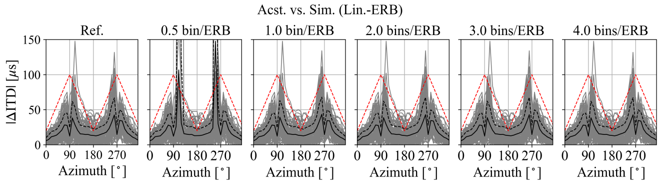

Fig. 9. Comparisons of interaural time differences in the horizontal plane. Acoustic HRTFs vs. simulated HRTF with linear frequency scale (Ref.), and simulated HRTFs with lin.-ERB. frequency scales. Individual results (gray solid); average (black, solid) and std. (black, dashed); JND (red, dashed).

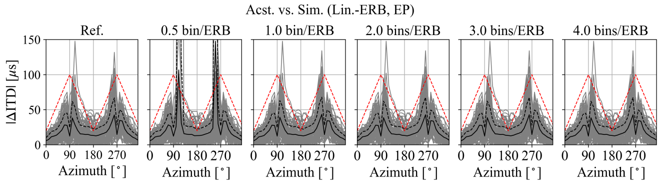

Fig. 10. Comparisons of interaural time differences in the horizontal plane. Acoustic HRTFs vs. simulated HRTF with linear frequency scale, and Acoustic HRTFs vs. HRTFs simulated with lin.-ERB frequency scale and phase extrapolated from 5 kHz upwards. Individual results (gray solid); average (black, solid) and std. (black, dashed); JND (red, dashed).

V. Discussion

We assessed the quality of the HRTFs simulated using the lin.-ERB frequency scale by performing numerical experiments that compared them with both HRTFs simulated using the standard linear frequency scale and acoustic measurements. When compared in terms of magnitude distortion to the standard simulations, the HRTFs produced with $E \geq 1\,\text{bin/ERB}$ provided spectral distortions below $1.5\,\text{dB}$ throughout the entire frequency range, being primarily concentrated at high frequencies (above $\approx 9\,\text{kHz}$). These results are, however, irrelevant in light of the comparisons performed with acoustically measured HRTFs: only when the lowest resolution tested ($E = 0.5\,\text{bin/ERB}$) was used, HRTFs simulated with the proposed hybrid scale differed from the HRTFs simulated with the standard linear one, suggesting one should use at least $E \geq 1\,\text{bin/ERB}$ to achieve comparable results.

As for the phase response, when taking the acoustically measured HRTFs as references, linearly extrapolating the phase at high frequencies using the group delay and $E \geq 1\,\text{bin/ERB}$ has surprisingly resulted in lower $|\Delta\text{IPD}|$ than interpolating the phase values within the lin.-ERB scales or even using the phase simulated using the full linear scale. Such an advantage is, however, only observed above 15 kHz, and the results below this frequency are equivalent for $E \geq 2\,\text{bin/ERB}$.

Furthermore, although the phase response of HRTFs is critical for localization only up to around 3 kHz [26]–[28], it is not clear what perceptual impact the abovementioned deviations caused by aliasing would have. As a general conclusion, for HRTFs simulated using $E \leq 2\,\text{bin/ERB}$, it seems preferable to use the extrapolated phase, whereas the results are equivalent for $E \geq 3\,\text{bin/ERB}$.

Next, assessments conducted regarding $|\Delta\text{ITD}|$, which only accounts for the overall delay of the lower portion of the frequency spectrum, also suggested equivalence between the HRTFs simulated with $E \geq 1\,\text{bin/ERB}$ and those simulated with the linear frequency scale.

Overall, the results indicate that simulating HRTFs using at least $E = 1\,\text{bin/ERB}$, provided an extrapolated phase response above 5 kHz is used, might yield sufficiently similar results to the standard approach for simulation of HRTFs using the BEM in terms of similarity to the acoustically measured HRTFs. This configuration provides savings in computing time of around 92%, for the HUTUBS database. Nevertheless, in order to achieve perceptual similarity to the HRTFs simulated using the linear frequency scale, using $E \geq 2\,\text{bins/ERB}$ seems to be closer to a minimum requirement; at this point, the approach to estimate the phase at high frequencies seems irrelevant. Such a resolution still saves around 86% of processing time. The more conservative resolution range of $E \geq 3\,\text{bins/ERB}$ should provide indistinguishable results to the standard simulations, still offering high savings in computing time (≈ 80%).

VI. Conclusions

In this paper, we explored employing a hybrid linear-ERB frequency scale for numerical simulations of HRTFs using the boundary element method. Also, two approaches for phase calculation at high frequencies were tested: interpolation from ERB scale to linear scale and extrapolation based on the average group delay at low frequencies. We conducted numerical experiments using the HUTUBS database to assess the quality of the HRTFs produced by this approach by comparing them with standard simulations of HRTFs and acoustic measurements regarding distortions in magnitude, phase, and interaural time differences. The results obtained suggested that HRTFs simulated using our approach with a minimum resolution of $E = 1\,\text{bin/ERB}$ did not differ from the standard simulated HRTFs, when compared to the acoustic measurements, roughly costing 8% of the original computing time that standard simulated HRTFs would require. Having those standard simulations as a reference, our assessments showed that using at least $2\,\text{bins/ERB}$ should potentially provide seamless results, only taking around 14% of the original processing time. In future studies, perceptual tests should be conducted to validate the proposed approach in realistic scenarios. For instance, the transparency of the approximation procedure could be assessed in double-blind experiments involving 3D simulations using both the standard and the proposed approaches.

REFERENCES

[1] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The internet of sounds: Convergent trends, insights, and future directions," *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11 264–11 292, 2023.

[2] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[3] M. Oehler, M. do V. M. da Costa, M. Regener, and T. M. Voong, "Relevance of individual numerically simulated head-related transfer functions for different scenarios in virtual environments," in *Proceedings of the International Conference on Audio for Virtual and Augmented Reality 2022*, Redmond, USA, August 2022.

[4] M. Oehler, T. M. Voong, M. Regener, and M. do V. M. da Costa, "Effect of using individually simulated HRTFs on the outcome of tournament selection procedures in a virtual environment," in *Proceedings of Sound and Music Computing 2022 (SMC-22)*, St. Etienne, France, June 2022.

[5] M. Oehler, T. M. Voong, M. Regener, M. do V. M. da Costa, and C. Reuter, "Importance of HRTF personalisation for audio rendering in music-related virtual environments," in *Proceedings of the 10th Convention of the European Acoustics Association (Forum Acusticum)*, Torino, Italy, September 2023.

[6] V. Bauer, D. Soudoplatoff, L. Menon, and A. Pras, "Binaural Headphone Monitoring to Enhance Musicians' Immersion in Performance," in *Advances in Fundamental and Applied Research on Spatial Audio*. IntechOpen, June 2022. [Online]. Available: https://universite-paris-saclay.hal.science/hal-04447847

[7] C. Guezenoc and R. Seguier, "HRTF individualization: A survey," in *Proceedings of the 145th Audio Engineering Society International Convention*, New York, USA, October 2018.

[8] L. Picinali and B. F. Katz, "System-to-user and user-to-system adaptations in binaural audio," in *Sonic Interactions in Virtual Environments*. Springer International Publishing Cham, 2022, pp. 115–143.

[9] J. Zhao, D. Yao, J. Gu, and J. Li, "Efficient prediction of individual head-related transfer functions based on 3d meshes," *Applied Acoustics*, vol. 219, no. 30, p. 109938, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X24000896

[10] K. Pollack, W. Kreuzer, and P. Majdak, "Modern acquisition of personalised head-related transfer functions – an overview," in *Advances in Fundamental and Applied Research on Spatial Audio*, 1st ed., B. Katz and P. Majdak, Eds. London, United Kingdom: IntechOpen, January 2022.

[11] F. Brinkmann, W. Kreuzer, J. Thomsen, S. Dombrovskis, K. Pollack, S. Weinzierl, and P. Majdak, "Recent advances in an open software for numerical HRTF calculation," *Journal of the Audio Engineering Society*, vol. 71, no. 7/8, pp. 502–514, July 2023.

[12] W. Kreuzer, K. Pollack, F. Brinkmann, and P. Majdak, "Numcalc: An open-source bem code for solving acoustic scattering problems," *Engineering Analysis with Boundary Elements*, vol. 161, no. 1, pp. 157–178, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0955799724000171

[13] H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions," in *Proceedings of the 22nd International Congress on Sound and Vibration*, Florence, Italy, July 2015.

[14] H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 208–222, July 2015.

[15] M. do V. M. da Costa, L. W. P. Biscainho, and M. Oehler, "Low-cost numerical approximation of HRTFs: A non-linear frequency sampling approach," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Copenhagen, Denmark, Sep. 2023.

[16] B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 750–753, September 1983.

[17] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1, pp. 103–138, August 1990.

[18] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses," *Journal of the Audio Engineering Society*, vol. 67, no. 9, pp. 705–718, 2019.

[19] A. S. Committee, *AES standard for file exchange - Spatial acoustic data file format (SOFA)*, Audio Engineering Society, Inc. Std., 2015.

[20] B. Rafaely, *Fundamentals of Spherical Array Processing*, 1st ed. Berlin, Heidelberg: Springer, 2015.

[21] T. Palm, S. Koch, F. Brinkmann, and M. Alexa, "Curvature-adaptive mesh grading for numerical approximation of head-related transfer functions," in *In Fortschritte der Akustik - DAGA 2021 : 47. Jahrestagung für Akustik*, Vienna, Austria, August 2021.

[22] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Sofia sound field analysis toolbox," in *Proceedings of the ICSA International Conference on Spatial Audio*, 2011.

[23] A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," *Journal of the Audio Engineering Society*, vol. 60, no. 1, pp. 54–62, Jan 2012.

[24] G. Enzner, C. Antweiler, and S. Spors, "Acquisition and representation of head-related transfer functions," in *The Technology of Binaural Listening*, ser. Modern acoustics and signal processing, J. Blauert, Ed. Heidelberg et al.: Springer, 2013, vol. 1, pp. 57–92.

[25] A. W. Mills, "On the minimum audible angle," *Journal of the Acoustical Society of America*, vol. 30, no. 4, pp. 237–246, apr 1958.

[26] E. A. Macpherson and J. Middlebrooks, "Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 5, no. 111, pp. 2219–2236, May 2002.

[27] E. Rasumow, M. Blau, and M. Hansen, "Smoothing individual head-related transfer functions in the frequency and spatial domains," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 2012–2025, May 2014.

[28] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of Ambisonics signals via magnitude least squares," in *In Fortschritte der Akustik - DAGA 2018 : 44. Jahrestagung für Akustik*, Munich, Germany, March 2021, pp. 339–342.