applied sciences

MDPI

Article

Ear Centering for Accurate Synthesis of Near-Field Head-Related Transfer Functions †

Ayrton Urviola [1,*,‡], Shuichi Sakamoto [2,‡] and César D. Salvador [1,‡]

[1] Perception Research, 4000 Lima, Peru
[2] Research Institute of Electrical Communication (RIEC) and Graduate School of Information Sciences (GSIS), Tohoku University, Sendai 980-8577, Japan
* Correspondence: aurviola@perception3d.com
† This paper is an extended version of our paper published in the International Conference on Immersive and 3D Audio (I3DA 2021), Bologna, Italy, 8–10 September 2021.
‡ These authors contributed equally to this work.

Abstract: The head-related transfer function (HRTF) is a major tool in spatial sound technology. The HRTF for a point source is defined as the ratio between the sound pressure at the ear position and the free-field sound pressure at a reference position. The reference is typically placed at the center of the listener's head. When using the spherical Fourier transform (SFT) and distance-varying filters (DVF) to synthesize HRTFs for point sources very close to the head, the spherical symmetry of the model around the head center does not allow for distinguishing between the ear position and the head center. Ear centering is a technique that overcomes this source of inaccuracy by translating the reference position. Hitherto, plane-wave (PW) translation operators have yield effective ear centering when synthesizing far-field HRTFs. We propose spherical-wave (SW) translation operators for ear centering required in the accurate synthesis of near-field HRTFs. We contrasted the performance of PW and SW ear centering. The synthesis errors decreased consistently when applying SW ear centering and the enhancement was observed up to the maximum frequency determined by the spherical grid.

Keywords: head-related transfer functions; ear centering; acoustic centering; translation operator; spherical Fourier transform; distance-varying filter

Citation: Urviola, A.; Sakamoto, S.; Salvador, C.D. Ear Centering for Accurate Synthesis of Near-Field Head-Related Transfer Functions. *Appl. Sci.* **2022**, *12*, 8290. https://doi.org/10.3390/app12168290

Academic Editors: Lamberto Tronchin and Francesca Merli

Received: 25 June 2022
Accepted: 16 August 2022
Published: 19 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The head-related transfer functions (HRTF) are a major tool in spatial sound technology for personal use [1–4]. They are linear filters describing the transmission of sound from a point in space to the eardrums of a listener [5]. The HRTF for a point source is defined as the ratio between the sound pressure at the ear position and the free-field sound pressure at a reference position. The reference is typically placed at the center of the listener's head. The HRTFs are commonly obtained for a sparse set of points at a single distance measured from the center of the head. Such distance is greater than 1 m and lies in a region of space referred to as the far field.

Besides far-field HRTF datasets, there is a growing interest in accurately synthesizing HRTFs for arbitrary points in the near field, that is, in the region of space within 1 m from the center of the head [6–8]. Interests include the development of near-field auditory displays [9] and the consideration of distance in auditory attention experiments [10].

A promising synthesis approach extrapolates near-field HRTFs starting from far-field ones using spherical Fourier transforms (SFT) and distance-varying filters (DVF) [11–13]. When using the SFT to represent spherical HRTF datasets, the default spherical symmetry of the SFT is specified with respect to the head center. The default spherical symmetry, however, does not allow for distinguishing between the reference (the head center) and

the measurement positions (the ears). Such mismatch produces a demand of a high number of basis functions in the SFT representation and, therefore, affects the synthesis accuracy.

Ear centering is the name adopted in this paper to address the mismatch between the default SFT center and the ear position in the framework of a more general technique called acoustic centering [14–18]. Ear centering is performed by means of acoustic operators that translate the center of symmetry of the SFT from the head center to the ears [19–23].

Table 1 summarizes the most relevant research interpreted in terms of translation operators for HRTF synthesis. The first column from the left shows the references. The second column indicates whether the proposed translation operation incorporates distance information or not. The third column highlights the translation point used for the translation process. It is worth noting that only [24] used the ear position provided in the HRTF dataset as the translation point. The fourth column indicates the mathematical domain where the translation is applied, either the unit sphere (the natural domain) or the SFT domain (the transform domain). Finally, the fifth column classifies the translation models.

Table 1. Overview of previous research in translation operators for ear centering in HRTF synthesis.

Reference	Distance	Translation Point	Domain	Translation Model
Richter et al., 2014 [19]	Yes	Optimized point around the ear	SFT domain	Ratio of hankel functions
Zaunschirm et al., 2018 [20]	No	y-axis with 8.5 cm radius	Unit sphere	Plane-wave
Ben-Hur et al., 2019 [21]	No	y-axis with 8.75 cm radius	Unit sphere	Plane-wave
Porschmann et al., 2019 [22]	No	y-axis with 9.19 cm radius	Unit sphere	Plane-wave with rigid sphere
Arend et al., 2021 [23]	No	y-axis with 9.19 cm radius	Unit sphere	Plane-wave with rigid sphere
Urviola et al., 2021 [24]	Yes	Ear position	Unit sphere	Spherical-wave

Richter et al. [19] used a ratio of Hankel functions to compensate for the difference in distance between each source and the ears, and then extrapolated the HRTF to a different distance from the head center. In our context, this operation is equivalent to a translation operator applied in the SFT domain. Zaunschirm et al. [20] calculated the time difference between the arrivals from each source to the ears and the head center, and then applied it as a phase difference to the HRTF dataset before the SFT. This can be interpreted as a plane-wave translation operator applied on the unit sphere domain. Ben-hur et al. [21] applied a ratio of pressures using plane waves arriving to the head center and the ears. This, again, corresponds to a plane-wave translation operator because it is changing the reference point of the HRTF. Porschmann et al. [22] applied a directional equalization to the HRTF using a rigid spherical scatterer, removing direction-dependent spectral components, and reducing the spatial complexity of the HRTF dataset. This whole process can be interpreted as a plane-wave translation operator, but rather than being applied in the free-field, it was applied in the presence of a rigid sphere, which emulated the most simple model of the human head. Urviola et al. [24] extended the work in [20,21] applying a free-field spherical-wave translation operator to incorporate distance information, as well as DVFs to synthesize HRTFs at close distances from the center of the head from both sparse and dense far-field HRTF datasets.

In summary, translation operators can include distance information [19,24] or not [20–23], can be applied in the unit sphere domain [20–24] or in the SFT domain [19], and can consider acoustic propagation in the free field [19–21,24] or include an acoustically rigid scatterer that mimics a simple head [22,23].

Hitherto, ear centering with free-field translation operators based on a plane-wave (PW) model, applied to HRTF datasets on the unit sphere, have yielded optimum use of SFT basis functions and accurate synthesis when distances between the sound source and the ears are large [21]. However, when PW translation operators are used to synthesize near-field HRTFs, the accuracy is affected because the PW model does not consider the distance information. Following this approach, it would be useful to have a translation operator that considers the distance between the sound source and the ears to synthesize near-field HRTFs.

In this paper, we propose to use a free-field translation operator based on a spherical-wave (SW) model for ear centering in near-field HRTF synthesis. Although the basic idea of our SW-based method has been proposed in our previous conference paper [24], there are still some unsolved problems. The main contributions of this paper when contrasted with [24] are as follows:

(a) The new SW translation operators translate the reference position and maintain fixed ear positions, whereas the previous translation operators translate the ear positions. The new operators are therefore consistent with the definition of HRTFs.
(b) The open question on whether PW and SW translation operators are equivalent when encoding HRTF datasets in the far field is addressed in this paper. The results are presented as a novel content in Section 4.
(c) The use of complex-valued SFT basis functions in this paper generalizes previous formulations based on real-valued SFT basis functions. This generalization allows for more precise analyses in terms of magnitude and phase.
(d) The review structured in Table 1 is new.
(e) This paper generalizes the theory, whereas the conference paper was oriented to implementing a particular case.

The reminder of this paper is organized as follows. Section 2 formulates ear centering for near-field HRTFs using translation operators. Section 3 compares PW and SW translation operators with calculated data. Section 4 dives into the effectiveness of using the proposed SW translation operator in the encoding stage of synthesis. Finally, Section 5 states the conclusions.

2. Ear Centering for Near-Field HRTF Synthesis

In the spherical coordinate system shown in Figure 1, a point in space $\mathbf{r} = (r, \theta, \phi)$ is specified by its radial distance r, azimuthal angle $\theta \in [-\pi, \pi]$, and elevation angle $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The azimuthal angle θ is measured from the x-axis, whereas the elevation angle ϕ is measured from the xy-plane. Positions in front of the listener lie along the positive x-axis or the direction $(\theta = 0, \phi = 0)$. All of what follows considers acoustic waves satisfying the Helmholtz equation with time-harmonic dependence e^{jkct}, where k denotes the wave number, c is the speed of sound in air, and j is the imaginary unit. The relation between frequency f and wave number k is $2\pi f = kc$.

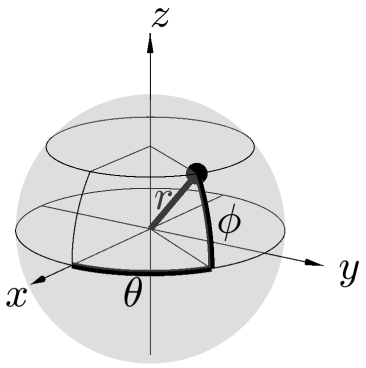


Figure 1. Spherical coordinate system.

Figure 2 shows the top-view geometry for theoretical near-field HRTF synthesis. The center of the head coincides with the origin $\mathbf{0} = (0, 0, 0)$ and the ear position is denoted by $\mathbf{r}_{\text{ear}} = (r_{\text{ear}}, \theta_{\text{ear}}, \phi_{\text{ear}})$. Let $\mathbf{a} = (a, \theta_a, \phi_a)$ be a point in a continuous, spherical distribution at a far distance a. Let $\mathbf{b} = (b, \theta_b, \phi_b)$ be an arbitrary point at a near distance b.

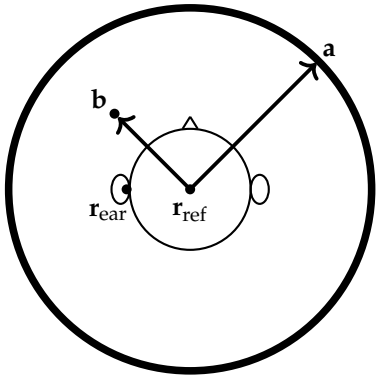


Figure 2. Geometry for near-field HRTF synthesis.

The HRTF is defined as the sound pressure at the ear position due to a point source at the source position, divided by the sound pressure at the reference position when the head is not present [5]. The HRTF, denoted by $\mathcal{H}$, depends on the source position $\mathbf{r}$, the ear positions $\mathbf{r}_{\text{ear}}$, and the reference position $\mathbf{r}_{\text{ref}}$. The HRTF is defined as follows:

$$\mathcal{H}(\mathbf{r}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ref}}) = \frac{\Psi(\mathbf{r}, \mathbf{r}_{\text{ear}})}{\Psi_{FF}(\mathbf{r}, \mathbf{r}_{\text{ref}})}, \tag{1}$$

where $\Psi(\mathbf{r}_{\text{s}}, \mathbf{r}_{\text{r}})$ denotes the pressure emanated from a source position $\mathbf{r}_{\text{s}}$ measured at a receiver position $\mathbf{r}_{\text{r}}$ and the sub-index FF stands for "free-field", which indicates that the head is not present. Throughout this paper, the reference position is the head center $\mathbf{r}_{\text{ref}} = \mathbf{0}$.

Figure 3 overviews the synthesis process with ear centering. The input is a continuous, spherical distribution of HRTFs from $\mathbf{a}$ to $\mathbf{r}_{\text{ear}}$, denoted by $\mathcal{H}(\mathbf{a}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ref}})$, whereas the output is a synthesized HRTF from $\mathbf{b}$ to $\mathbf{r}_{\text{ear}}$, denoted by $\mathcal{H}(\mathbf{b}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ref}})$. For simplicity, only the left ear is considered; however, the formulations below that relate the output to the input hold for both ears.

Direct ear centering is performed by an operator $\mathcal{T}$ that translates the reference of $\mathcal{H}$ from $\mathbf{r}_{\text{ref}}$ to $\mathbf{r}_{\text{ear}}$ as follows:

$$\mathcal{H}(\mathbf{a}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) = \mathcal{T}(\mathbf{a}, \mathbf{r}_{\text{ref}} \mapsto \mathbf{r}_{\text{ear}}) \mathcal{H}(\mathbf{a}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ref}}). \tag{2}$$

The existing PW translation operator proposed in [21] can be formulated as

$$\mathcal{T}_{\text{PW}}(\mathbf{a}, \mathbf{r}_{\text{ref}} \mapsto \mathbf{r}_{\text{ear}}) = e^{jk(r_{\text{ref}} \cos \Theta_{\mathbf{a}, \mathbf{r}_{\text{ref}}} - r_{\text{ear}} \cos \Theta_{\mathbf{a}, \mathbf{r}_{\text{ear}}})}, \tag{3}$$

where $\Theta_{\mathbf{a}, \mathbf{r}_{\text{ear}}}$ denotes the angle between $\mathbf{a}$ and $\mathbf{r}_{\text{ear}}$ and $\Theta_{\mathbf{a}, \mathbf{r}_{\text{ref}}}$ denotes the angle between $\mathbf{a}$ and $\mathbf{r}_{\text{ref}}$. Considering a PW emanating from $\mathbf{a}$, (3) stems from the ratio of PW observations at $\mathbf{r}_{\text{ref}}$ and $\mathbf{r}_{\text{ear}}$.

To include the distance information, we propose to use the following SW translation operator:

$$\mathcal{T}_{\text{SW}}(\mathbf{a}, \mathbf{r}_{\text{ref}} \mapsto \mathbf{r}_{\text{ear}}) = \frac{\|\mathbf{a} - \mathbf{r}_{\text{ear}}\|}{\|\mathbf{a} - \mathbf{r}_{\text{ref}}\|} e^{-jk(\|\mathbf{a} - \mathbf{r}_{\text{ref}}\| - \|\mathbf{a} - \mathbf{r}_{\text{ear}}\|)}, \tag{4}$$

where $\|\cdot\|$ denotes Euclidean norm. Considering an SW emanating from $\mathbf{a}$, (4) stems from the ratio of SW observations at $\mathbf{r}_{\text{ref}}$ and $\mathbf{r}_{\text{ear}}$.

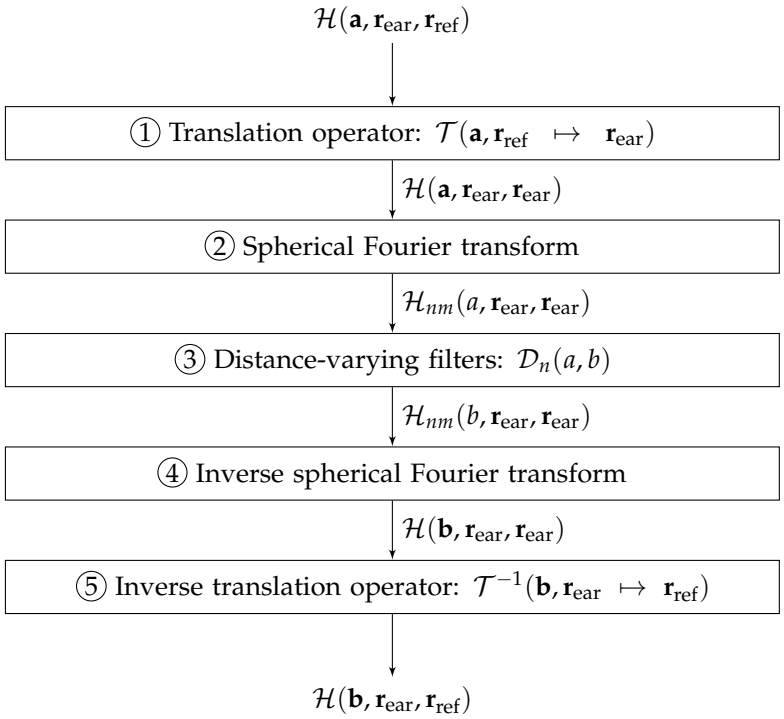


Figure 3. Near-field HRTF synthesis with ear centering based on translation operators.

The SFT of the ear-centered $\mathcal{H}(\mathbf{a}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}})$ is defined by

$$\mathcal{H}_{nm}(a, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) = \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathcal{H}(\mathbf{a}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) Y_n^m(\theta_a, \phi_a) \cos(\phi_a) d\phi_a d\theta_a. \tag{5}$$

Here, the SFT basis functions are complex-valued spherical harmonic functions Y_n^m of order n and degree m, defined as

$$Y_n^m(\theta, \phi) = N_{nm} P_n^{|m|}(\sin \phi) e^{jm\theta}, \tag{6}$$

where P_n^m is the non-normalized associated Legendre polynomial [25], the symbol $|\cdot|$ denotes absolute value, and N_{nm} is the following normalization factor:

$$N_{nm} = (-1)^{|m|} \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}}. \tag{7}$$

Distance variation from a to b is performed in the SFT domain according to the following expression:

$$\mathcal{H}_{nm}(b, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) = \mathcal{D}_n(a, b) \mathcal{H}_{nm}(a, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}). \tag{8}$$

Here, $\mathcal{D}_n$ denotes the spherical DVF of order n defined as

$$\mathcal{D}_n(a, b) = \frac{h_n^{(1)}(kb)}{h_n^{(1)}(ka)}, \tag{9}$$

where $h_n^{(1)}$ is the spherical Hankel function of the first kind and order n [26].

The inverse spherical Fourier transform (ISFT) extracts HRTFs for arbitrary directions using the following expression:

$$\mathcal{H}(\mathbf{b}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathcal{H}_{nm}(b, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}) Y_n^m(\theta_b, \phi_b). \tag{10}$$

Finally, inverse ear centering is performed with the inverse operator $\mathcal{T}^{-1}$ that translates the reference from $\mathbf{r}_{\text{ear}}$ to $\mathbf{r}_{\text{ref}}$:

$$\mathcal{H}(\mathbf{b}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ref}}) = \mathcal{T}^{-1}(\mathbf{b}, \mathbf{r}_{\text{ear}} \mapsto \mathbf{r}_{\text{ref}})\mathcal{H}(\mathbf{b}, \mathbf{r}_{\text{ear}}, \mathbf{r}_{\text{ear}}). \tag{11}$$

The inverse PW translation operator [21] can be expressed as

$$\mathcal{T}^{-1}_{\text{PW}}(\mathbf{b}, \mathbf{r}_{\text{ear}} \mapsto \mathbf{r}_{\text{ref}}) = e^{jk(r_{\text{ear}} \cos \Theta_{\mathbf{b},\mathbf{r}_{\text{ear}}} - r_{\text{ref}} \cos \Theta_{\mathbf{b},\mathbf{r}_{\text{ref}}})}. \tag{12}$$

Considering a PW emanating from $\mathbf{b}$, (12) stems from the ratio of PW observations at $\mathbf{r}_{\text{ear}}$ and $\mathbf{r}_{\text{ref}}$.

The proposed inverse SW translation operator takes the form

$$\mathcal{T}^{-1}_{\text{SW}}(\mathbf{b}, \mathbf{r}_{\text{ear}} \mapsto \mathbf{r}_{\text{ref}}) = \frac{\|\mathbf{b} - \mathbf{r}_{\text{ref}}\|}{\|\mathbf{b} - \mathbf{r}_{\text{ear}}\|} e^{-jk(\|\mathbf{b}-\mathbf{r}_{\text{ear}}\| - \|\mathbf{b}-\mathbf{r}_{\text{ref}}\|)}. \tag{13}$$

Considering a SW emanating from $\mathbf{b}$, (13) stems from the ratio of SW observations at $\mathbf{r}_{\text{ear}}$ and $\mathbf{r}_{\text{ref}}$.

In summary, in Figure 3, blocks ① and ② compose the encoding stage, block ③ performs distance variation, and blocks ④ and ⑤ compose the decoding stage.

3. Evaluation with Calculated HRTF Datasets

This section compares the performance of the PW and SW translation operators formulated in Section 2. The continuous, spherical distributions that underlie the formulations are sampled using spherical grids. The SFT and DVF algorithms described in [13,27], respectively, were adapted to the purposes of our evaluations.

3.1. Conditions

In practice, HRTF datasets are obtained for a finite number of point sources distributed in spherical grids. The infinite sum in (10) representing the inverse SFT will therefore be truncated up to a maximum order N, which depends on the spherical sampling scheme.

Among the existing calculated near-field HRTF collections [28,29], we chose the one in [29] because its data are open and its resolution across distance is dense. Left-ear HRTFs for two individual head models without torso were used in evaluations. The spatial features due to the torso are prominent at the lower frequencies and can extend up to 3 kHz, whereas the features due to the head and pinna span the middle and high frequencies [30]. Although our assessment focuses on head and pinna features, the lower frequencies are also covered; hence, the torso absence does not limit our results. The HRIRs have 512 samples along time, were sampled at 48 kHz, and $c = 344$ m/s. The left-ear positions were extracted from the head models.

The sound sources were distributed in spherical grids based on subdivisions of the edges of the icosahedron. Icosahedral samplings were chosen because they achieve bounded integration errors distributed across all orders, whereas other samplings aiming at perfect quadrature at low orders yield large errors concentrated in the high orders [27].

The number of points P in an icosahedral grid, generated with a subdivision factor q, is calculated as

$$P = 10q^2 + 2. \tag{14}$$

For almost regular spherical samplings, such as the icosahedral ones, the maximum SFT order N_{g} achievable with a grid of P points is

$$N_{\text{g}} = \lfloor \sqrt{P} \rfloor - 1. \tag{15}$$

It ensures reliable synthesis up to a maximum frequency

$$f_{\max} = \frac{cN_{\mathrm{g}}}{2\pi r_{\mathrm{bound}}}, \tag{16}$$

where r_{bound} is the radius of a sphere fully containing the head.

Datasets at a distance $a = 100$ cm were used as inputs. Three icosahedral grids with $P = 12, 42, 252$, correspondingly $q = 1, 2, 5$, and $N_{\mathrm{g}} = 2, 5, 14$, were used. The maximum SFT orders to analyze the spherical HRTF datasets were limited by the far-to-near field transitions and the input resolutions as follows:

$$N = \min(\lceil kr_{\mathrm{bound}} \rceil, N_{\mathrm{g}}). \tag{17}$$

Here, r_{bound} is the same used in (16) and the value kr_{bound} indicates the far-to-near field transition.

Because the ideal DVFs in (9) yield excessive values for higher orders and lower frequencies [13], their action needs to be limited according to

$$\hat{\mathcal{H}}_{nm}(b, \mathbf{r}_{\mathrm{ear}}, \mathbf{r}_{\mathrm{ear}}) = \mathcal{W}_n(a, b)\mathcal{H}_{nm}(b, \mathbf{r}_{\mathrm{ear}}, \mathbf{r}_{\mathrm{ear}}) \tag{18}$$

with an order truncation and scaling window defined in [31] as:

$$\mathcal{W}_n(a, b) = \begin{cases} \frac{b}{a} e^{-jk(b-a)}, & n \leq N, \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

In (16), (17), and (19), r_{bound} should ideally be the radius of the smallest sphere containing the head. However, this theoretical limit yielded artifacts due to the discontinuities of the truncation rule $n \leq \lceil kr_{\mathrm{bound}} \rceil$. To reduce these artifacts, we have empirically chosen $r_{\mathrm{bound}} = 16$ cm as a convenient value for the two individual head models in [29].

Datasets at distances b, ranging from 20 to 100 cm with 1 cm spacing, were used as target data. For each distance, $P = 642$ directions on an icosahedral grid with $q = 8$ were considered. Datasets were also synthesized for these distributions of points to evaluate three scenarios:

- No ear centering (NEC);
- Plane-wave ear centering (PEC) using the translation operators in (3) and (12);
- Spherical-wave ear centering (SEC) using the translation operators in (4) and (13).

3.2. Error Metric

Target and synthesized HRTF datasets were respectively organized as $\boldsymbol{H}(b_i, \Omega_j, f_\kappa, s_\ell)$ and $\hat{\boldsymbol{H}}(b_i, \Omega_j, f_\kappa, s_\ell)$. Index $i = 1, 2, \cdots, 81$ indicates radial distances; index $j = 1, 2, \cdots, 642$, directions on the sphere with $\Omega_j = (\theta_j, \phi_j)$; index $\kappa = 1, 2, \cdots, 257$, frequency bins; and index $\ell = 1, 2$, individual subjects. The synthesis error across angles and subjects is defined as

$$E(b_i, f_\kappa) = \underset{s_\ell}{\mathrm{RMS}} \left\{ \frac{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H} - \hat{\boldsymbol{H}}\}}{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H}\}} \right\}, \tag{20}$$

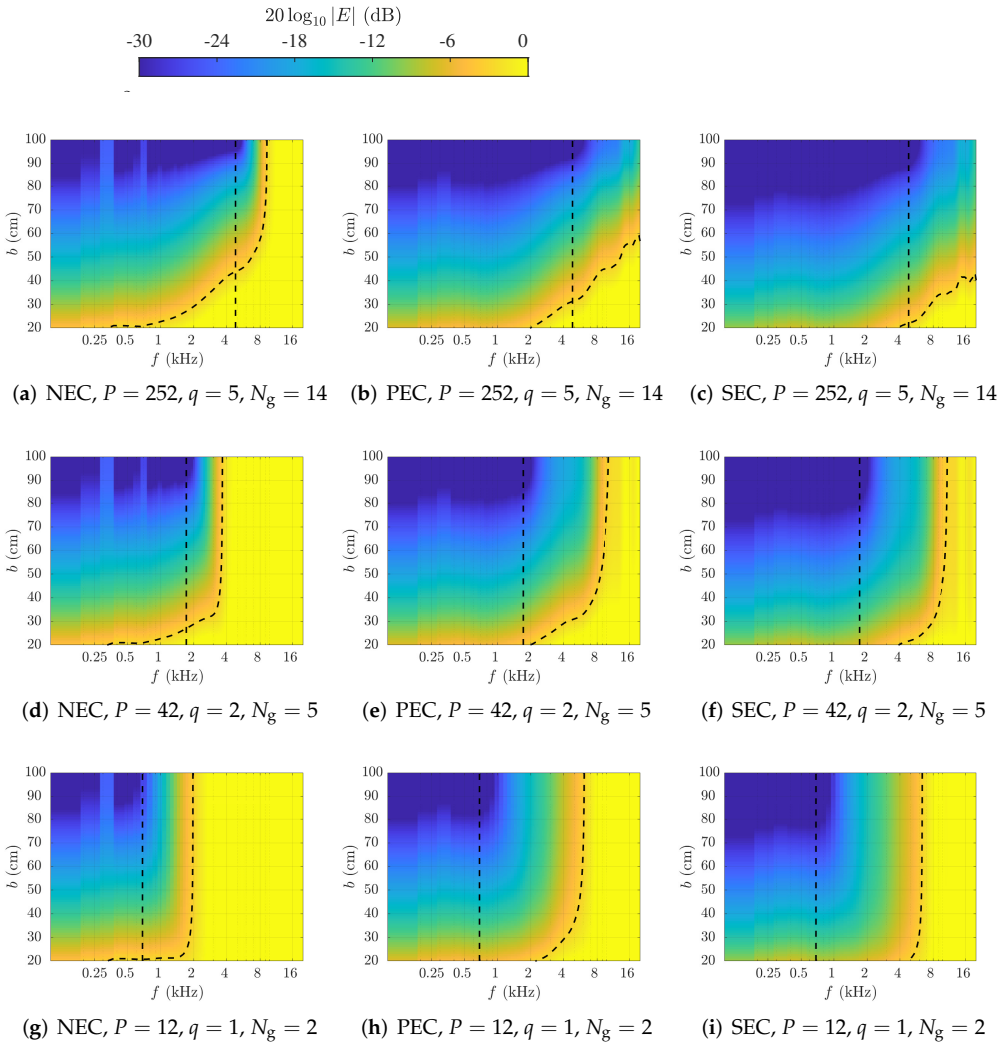
and the above synthesis error across all distances as

$$E(f_\kappa) = \underset{b_i}{\mathrm{RMS}} \left\{ \underset{s_\ell}{\mathrm{RMS}} \left\{ \frac{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H} - \hat{\boldsymbol{H}}\}}{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H}\}} \right\} \right\}, \tag{21}$$

where RMS stands for root mean square along either directions Ω_j, individual subjects s_ℓ or distances b_i.

3.3. Results

All panels in Figure 4 show synthesis errors calculated with (20) and displayed in a logarithmic scale, from −30 dB to 0 dB, to contrast with the perceivable HRTF dynamic range of around 30 dB, as reported in [32]. The black-dashed curves highlight the −3 dB values and are used as an indicator to compare among panels. We use the −3 dB indicator because this value is commonly considered as a perceivable difference. The black-dashed lines indicate $f_{\max}$ as formulated in (16).



(a) NEC, $P = 252$, $q = 5$, $N_g = 14$ **(b)** PEC, $P = 252$, $q = 5$, $N_g = 14$ **(c)** SEC, $P = 252$, $q = 5$, $N_g = 14$

(d) NEC, $P = 42$, $q = 2$, $N_g = 5$ **(e)** PEC, $P = 42$, $q = 2$, $N_g = 5$ **(f)** SEC, $P = 42$, $q = 2$, $N_g = 5$

(g) NEC, $P = 12$, $q = 1$, $N_g = 2$ **(h)** PEC, $P = 12$, $q = 1$, $N_g = 2$ **(i)** SEC, $P = 12$, $q = 1$, $N_g = 2$

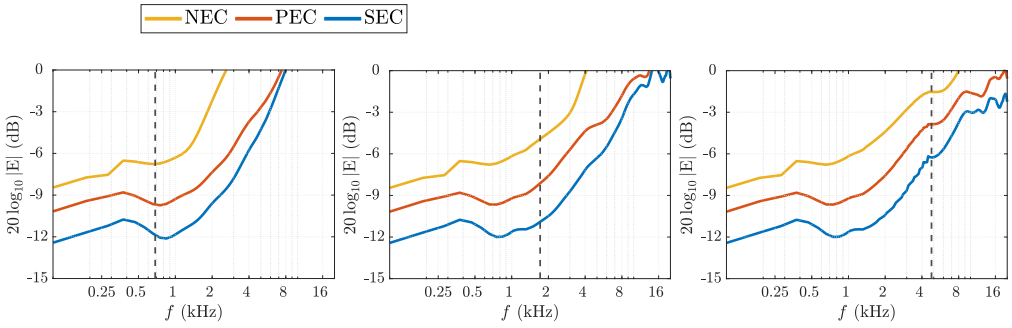
Figure 4. Synthesis error calculated with (20). Black-dashed curves indicate −3 dB values. Black-dashed lines indicate $f_{\max}$ in (16). Left column: NEC. Center column: PEC. Right column: SEC.

In Figure 4, panels (a), (d), and (g), correspond to synthesis with NEC; panels (b), (e), and (h), to synthesis with PEC; and panels (c), (f), and (i), to synthesis with SEC. Panels (a), (b), and (c) correspond to synthesis from $P = 252$ points ($q = 5$, $N_g = 14$); panels (d), (e), and (f), to synthesis from $P = 42$ points ($q = 2$, $N_g = 5$); and panels (g), (h), and (i), to synthesis from $P = 12$ points ($q = 1$, $N_g = 2$).

When comparing panels along rows in Figure 4, it is observed that applying SEC outperforms the accuracy across all spherical resolutions when contrasted with PEC and NEC. The enhancements of SEC are more noticeable at near distances and their benefits extend even beyond the corresponding $f_{\max}$. A closer inspection of the black-dash curves (−3 dB) at distances close to the head (around 20 cm) shows that, for the same error level, SEC yields a bandwidth improvement of around 2 kHz when compared to PEC. At the same

distance, we can obtain an accurate synthesis up to a higher frequency; in other words, at the same accuracy, we can obtain a synthesis up to a distance closer to the head.

Panels in Figure 5 show the overall synthesis errors across distances calculated with (21) and displayed in a logarithmic scale, from −15 dB to 0 dB. The black-dashed lines indicate $f_{\max}$ in (16) once again. Panel (a) corresponds to synthesis from $P = 12$ points ($q = 1$, $N_g = 2$), panel (b) to synthesis from $P = 42$ points ($q = 2$, $N_g = 5$), and panel (c) to synthesis from $P = 252$ points ($q = 5$, $N_g = 14$). Yellow curves correspond to synthesis with NEC, red curves to synthesis with PEC, and blue curves to synthesis with SEC.



(**a**) $P = 12$, $q = 1$, $N_g = 2$ (**b**) $P = 42$, $q = 2$, $N_g = 5$ (**c**) $P = 252$, $q = 5$, $N_g = 14$

Figure 5. Synthesis error calculated with (21) in dB. Black-dashed lines indicate $f_{\max}$ in (16).

When comparing the curves in each panel in Figure 5 at $f_{\max}$, we observe the following approximate improvements. In panel (a), PEC outperforms NEC by 3 dB, and SEC outperforms PEC by 2 dB. In panel (b), PEC outperforms NEC by 3 dB and SEC outperforms PEC by 3 dB. In panel (c), PEC outperforms NEC by 2 dB, and SEC outperforms PEC by 2 dB.

In summary, in the whole distance range from 20 to 100 cm and at frequencies below $f_{\max}$, applying SEC yields an approximate improvement of 2 dB when compared to PEC.

4. Effectiveness of SW Ear Centering in the Encoding Stage of HRTF Synthesis

In the far field, spherical waves are similar to plane waves, thus the reason why many models consider any far-field source as a plane wave. This similarity may tempt us to change the SW direct translation operator by the PW one, while maintaining the inverse SW translation operator because in the near-field the similarities between plane waves and spherical waves disappear. However, we investigate the possible reasons why this approach is not a viable option for our ear centering scheme.

Plane waves and spherical waves are similar in the far field. However, translation operators are neither plane nor spherical waves, but ratios of them. This subtle difference greatly impacts the synthesis performance.

Magnitudes and phases of direct translation operators are identified in (3) and (4) as

$$|\mathcal{T}_{\mathrm{PW}}(\mathbf{a}, \mathbf{r}_{\mathrm{ref}} \mapsto \mathbf{r}_{\mathrm{ear}})| = 1, \tag{22}$$

$$\angle\mathcal{T}_{\mathrm{PW}}(\mathbf{a}, \mathbf{r}_{\mathrm{ref}} \mapsto \mathbf{r}_{\mathrm{ear}}) = k(r_{\mathrm{ref}} \cos\Theta_{\mathbf{a},\mathbf{r}_{\mathrm{ref}}} - r_{\mathrm{ear}} \cos\Theta_{\mathbf{a},\mathbf{r}_{\mathrm{ear}}}), \tag{23}$$

$$|\mathcal{T}_{\mathrm{SW}}(\mathbf{a}, \mathbf{r}_{\mathrm{ref}} \mapsto \mathbf{r}_{\mathrm{ear}})| = \frac{\|\mathbf{a} - \mathbf{r}_{\mathrm{ear}}\|}{\|\mathbf{a} - \mathbf{r}_{\mathrm{ref}}\|}, \tag{24}$$

$$\angle\mathcal{T}_{\mathrm{SW}}(\mathbf{a}, \mathbf{r}_{\mathrm{ref}} \mapsto \mathbf{r}_{\mathrm{ear}}) = -k(\|\mathbf{a} - \mathbf{r}_{\mathrm{ref}}\| - \|\mathbf{a} - \mathbf{r}_{\mathrm{ear}}\|). \tag{25}$$

In Figure 6, the top panel is a color legend for panels (a), (b), (c), and (d). The color dots in the top panel indicate the positions $\mathbf{a}$ of twelve point sources placed at the vertices of the icosahedron, at a distance of 100 cm from the center of the head $\mathbf{r}_{\mathrm{ref}}$ marked with a circle. The coordinate system used in the top panel is the same described in Figure 1. The positive x-axis indicates the front, and the y-axis is the interaural axis. The left ear $\mathbf{r}_{\mathrm{ear}}$ lies on the positive y-axis and is marked with an asterisk. Point sources with the same color

indicate that they are at the same distance from $\mathbf{r}_{\text{ear}}$. The point sources lie in three regions of space:

- Blue and yellow dots: $\|\mathbf{a} - \mathbf{r}_{\text{ear}}\| > \|\mathbf{a} - \mathbf{r}_{\text{ref}}\|$,
- Green dots: $\|\mathbf{a} - \mathbf{r}_{\text{ear}}\| \approx \|\mathbf{a} - \mathbf{r}_{\text{ref}}\|$,
- Purple and red dots: $\|\mathbf{a} - \mathbf{r}_{\text{ear}}\| < \|\mathbf{a} - \mathbf{r}_{\text{ref}}\|$.

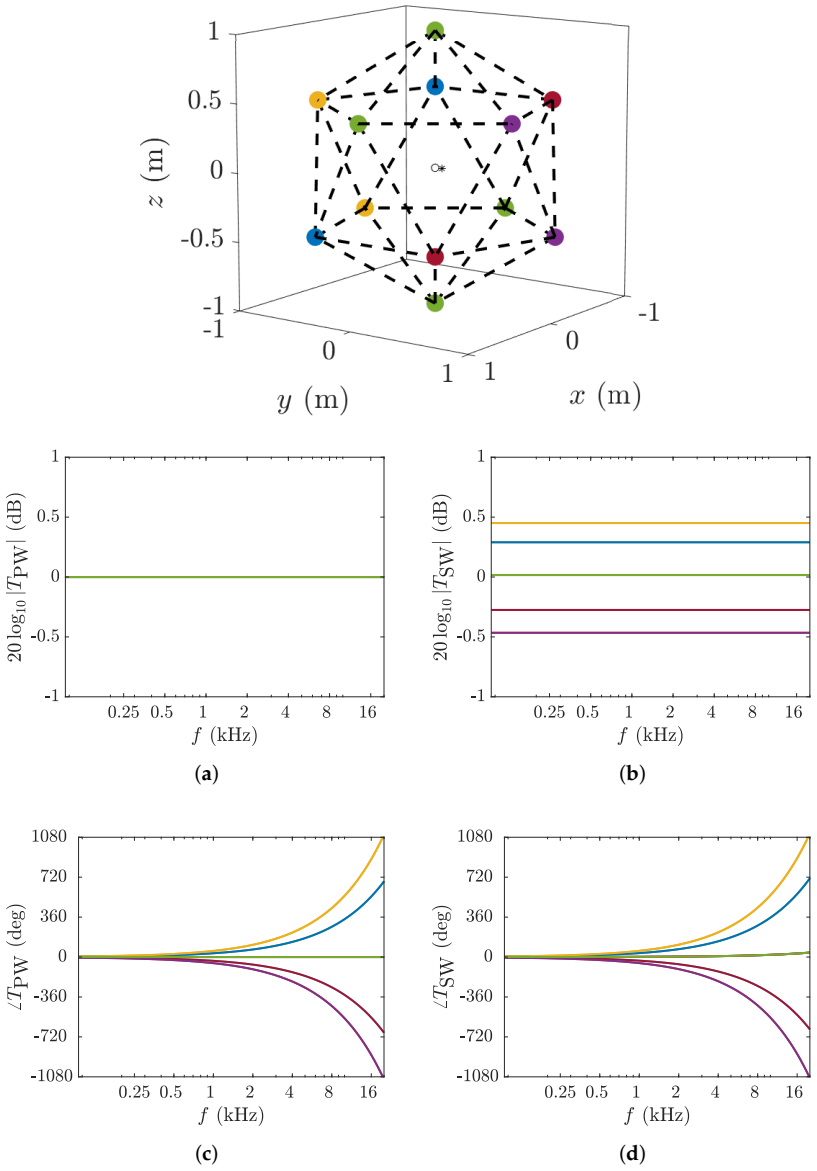


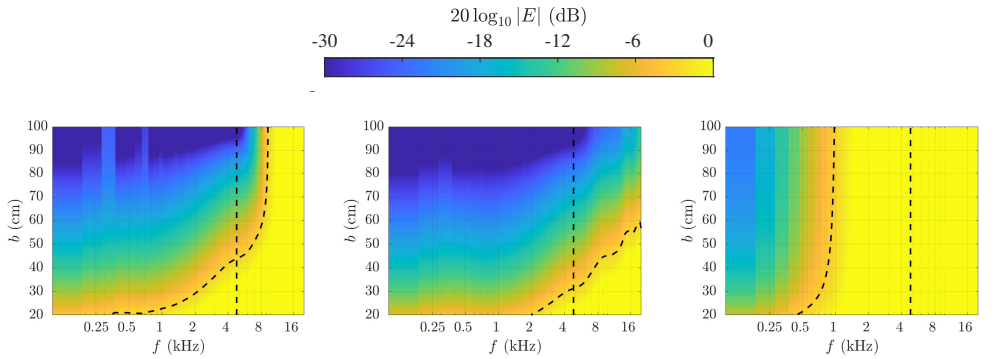
Figure 6. PW and SW direct translation operators for twelve different directions along the vertices of the icosahedron (top panel), at a distance of 100 cm from the center of the head. (**a**) Magnitude in (22); (**b**) Magnitude in (24); (**c**) Phase in (23); (**d**) Phase in (25).

In Figure 6, panels (a) and (b) show the magnitude of the PW and SW translation operators identified in (22) and (24) respectively, expressed in logarithmic scale from -1 dB to 1 dB; and panels (c) and (d) show the phase of the PW and SW translation operators identified in (23) and (25), respectively, expressed in degrees from -1080 to 1080 degrees.

When comparing panels (a) and (b) in Figure 6, we observe that panel (b) shows changes in amplitude of around ± 0.5 dB. There are no changes in panel (a) from point to point because the magnitude of the PW direct translation operator, identified in (22), remains the same for every direction. In panel (b), we observe that blue and yellow

point sources yield positive magnitudes in dB, green point sources are in the median plane and yield approximately 0 dB in magnitude, and purple and red point sources yield negative magnitudes in dB. This is consistent with the magnitude of the SW direct translation operator identified in (24). On the other hand, when comparing panels (c) and (d) in Figure 6, we observe negligible changes in phase between the phases of the direct PW and SW translation operators.

All panels in Figure 7 show synthesis errors calculated with (20) and displayed in a logarithmic scale from -30 dB to 0 dB. The black-dashed curves highlight the -3 dB values and are used as an indicator to compare among panels. The black-dashed lines indicate $f_{\max}$ in (16). The conditions for synthesis are the same as the ones mentioned in Section 3.1, and all panels correspond to synthesis from $P = 252$ points ($q = 5$, $N_{\mathrm{g}} = 14$).



(**a**) NEC (**b**) Direct SEC and inverse SEC (**c**) Direct PEC and inverse SEC

Figure 7. Comparison of performance with synthesis error (20) in dB ($P = 252$, $q = 5$, $N_{\mathrm{g}} = 14$). (**a**) synthesis with NEC. Panel; (**b**) synthesis with direct and inverse SEC; (**c**) synthesis with direct PEC and inverse SEC.

When comparing panels in Figure 7, we observe that the synthesis performance degrades dramatically in panel (c), even compared to NEC in panel (a), and cannot manage to reach the theoretical maximum frequency of synthesis $f_{\max}$ at a -3 dB synthesis error at any distance. The reason why the small variations in amplitude between panels (b) and (c) of Figure 6 produce the large synthesis errors shown in panel (c) of Figure 7 is that the DVFs in (9) are based on a spherical-wave model and are prone to inconsistencies in the translation model.

These results show the importance of respecting the model used for ear centering in both the direct and inverse processes. Both operations must be consistent between each other to allow for a successful synthesis.

5. Conclusions

We proposed new spherical-wave translation operators for ear centering to be used in the synthesis of head-related transfer functions for sound sources located close to the head. The new operators translate the reference position and maintain fixed ear positions; they are therefore consistent with the definition of HRTFs. Our formulations rely on complex-valued spherical basis functions, which allows for more precise evaluations.

We contrasted the performance of our proposal and the existing plane-wave translation operators. Synthesis accuracy increased consistently with our proposal. Enhancements were observed across distinct spherical resolutions and up to frequencies above the range of operation determined by the input spherical grid.

We also addressed the open question of whether direct PW and SW translation operators are interchangeable when encoding HRTF datasets in the far field. We found that an accurate synthesis requires using the same type of translation operator in both the direct and inverse stages of ear centering.

Extensions to this work might consider validations using bottom-up auditory models to simulate localization tasks. Subjective tests could also provide more insight into the validity of the suggested approach.

Author Contributions: All authors contributed equally to the conceptualization, methodology, validation, analysis, and writing of this research. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by JSPS KAKENHI Grant Nos. 19H04145 and 22H00523.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used as input to generate the results reported in this manuscript can be found at https://cesardsalvador.github.io/download.html. The last day of access was 4 August 2022.

Acknowledgments: The authors wish to thank the SI audio algorithm team for fruitful discussions. The authors also wish to thank Jorge Trevino for his insights and constructive debates.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Algazi, V.; Duda, R. Headphone-based spatial sound. *IEEE Signal Process. Mag.* **2011**, *28*, 33–42. [CrossRef]
2. Salvador, C.D.; Sakamoto, S.; Treviño, J.; Suzuki, Y. Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets. *Acoust. Sci. Technol.* **2017**, *38*, 51–62. [CrossRef]
3. Zhang, W.; Samarasinghe, P.N.; Chen, H.; Abhayapala, T.D. Surround by sound: A review of spatial audio recording and reproduction. *Appl. Sci.* **2017**, *7*, 532. [CrossRef]
4. Schörkhuber, C.; Zaunschirm, M.; Holdrich, R. Binaural rendering of ambisonic signals via magnitude least squares. In Proceedings of the DAGA German Annual Conference on Acoustics, Munich, Germany, 19–22 March 2018; Volume 44, pp. 339–342.
5. Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization*; Revised edition; MIT Press: Cambridge, MA, USA; London, UK, 1997.
6. Prepeliţă, S.T.; Bolaños, J.G.; Pulkki, V.; Savioja, L.; Mehra, R. Numerical simulations of near-field head-related transfer functions: Magnitude verification and validation with laser spark sources. *J. Acoust. Soc. Am.* **2020**, *148*, 153–166. [CrossRef]
7. Arend, J.M.; Liesefeld, H.R.; Pörschmann, C. On the influence of non-individual binaural cues and the impact of level normalization on auditory distance estimation of nearby sound sources. *Acta Acust. United Acust.* **2021**, *5*, 10. [CrossRef]
8. Armstrong, C. *Improvements in the Measurement and Optimisation of Head Related Transfer Functions for Binaural Ambisonics*. Ph.D. Thesis, University of York, York, UK, 2019.
9. Brungart, D.S. Near-Field Virtual Audio Displays. *Presence Teleop. Virt. Environ.* **2002**, *11*, 93–106. [CrossRef]
10. Sakamoto, S.; Monasterolo, F.; Salvador, C.D.; Cui, Z.; Suzuki, Y. Effects of target speech distance on auditory spatial attention in noisy environments. In Proceedings of the ICA 2019 and EAA Euroregio, Aachen, Germany, 9–13 September 2019; pp. 2177–2181.
11. Duraiswami, R.; Zotkin, D.N.; Gumerov, N.A. Interpolation and range extrapolation of HRTFs. *Proc. IEEE ICASSP* **2004**, *4*, 45–48. [CrossRef]
12. Pollow, M.; Nguyen, K.V.; Warusfel, O.; Carpentier, T.; Müller-Trapet, M.; Vorländer, M.; Noisternig, M. Calculation of head-related transfer functions for arbitrary field points using spherical harmonics. *Acta Acust. United Acust.* **2012**, *98*, 72–82. [CrossRef]
13. Salvador, C.D.; Sakamoto, S.; Treviño, J.; Suzuki, Y. Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values. *Acoust. Sci. Technol.* **2017**, *38*, 1–13. [CrossRef]
14. Gumerov, N.A.; Duraiswami, R. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*; Elsevier Series in Electromagnetism; Elsevier: Rockville, MD, USA, 2004.
15. Ben Hagai, I.; Pollow, M.; Vorländer, M.; Rafaely, B. Acoustic centering of sources measured by surrounding spherical microphone arrays. *J. Acoust. Soc. Am.* **2011**, *130*, 2003–2015. [CrossRef]
16. Shabtai, N.R.; Vorländer, M. Acoustic centering of sources with high-order radiation patterns. *J. Acoust. Soc. Am.* **2015**, *137*, 1947–1961. [CrossRef]
17. Wang, Y.; Chen, K. Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions. *J. Acoust. Soc. Am.* **2018**, *143*, 3474–3478. [CrossRef] [PubMed]

18. Kentgens, M.; Jax, P. Translation of a higher-order ambisonics sound scene by space warping. In Proceedings of the Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality. Audio Engineering Society, Virtual, 17–19 August 2020.
19. Richter, J.G.; Pollow, M.; Wefers, F.; Fels, J. Spherical harmonics based HRTF datasets: Implementation and evaluation for real-time auralization. *Acta Acust. United Acust.* **2014**, *100*, 667–675. [CrossRef]
20. Zaunschirm, M.; Schörkhuber, C.; Höldrich, R. Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. *J. Acoust. Soc. Am.* **2018**, *143*, 3616–3627. [CrossRef] [PubMed]
21. Ben-Hur, Z.; Alon, D.L.; Mehra, R.; Rafaely, B. Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment. *IEEE Trans. Audio Speech Lang. Process.* **2019**, *27*, 2249–2262. [CrossRef]
22. Pörschmann, C.; Arend, J.M.; Brinkmann, F. Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *27*, 1060–1071; Correction in **2020**, *28*, 2194–2194. [CrossRef]
23. Arend, J.M.; Brinkmann, F.; Pörschmann, C. Assessing spherical harmonics interpolation of time-aligned head-related transfer functions. *J. Audio Eng. Soc.* **2021**, *69*, 104–117. [CrossRef]
24. Urviola, A.; Sakamoto, S.; Salvador, C.D. Ear centering for near-distance head-related transfer functions. In Proceedings of the International Conference—Immersive and 3D Audio (I3DA): From Architecture to Automotive, Bologna, Italy, 8–10 September 2021; IEEE: Bologna, Italy, 2021. [CrossRef]
25. Olver, F.W.J.; Daalhuis, A.B.O.; Lozier, D.W.; Schneider, H.S. (Eds.) *NIST Digital Library of Mathematical Functions*, 1.1.0 ed.; 2020. Available online: http://dlmf.nist.gov/ (accessed on 4 August 2022).
26. Rehmann, U. *Encyclopedia of Mathematics*; 2020. Available online: https://encyclopediaofmath.org/ (accessed on 4 August 2022).
27. Salvador, C.D.; Sakamoto, S.; Treviño, J.; Suzuki, Y. Boundary matching filters for spherical microphone and loudspeaker arrays. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 461–474. [CrossRef]
28. Rui, Y.; Yu, G.; Xie, B.; Liu, Y. Calculation of individualized near-field head-related transfer function database using boundary element method. In Proceedings of the 134th Convention of Audio Engineering Society, Rome, Italy, 4–7 May 2013.
29. Salvador, C.D.; Sakamoto, S.; Treviño, J.; Suzuki, Y. Dataset of near-distance head-related transfer functions calculated using the boundary element method. In Proceedings of the Audio Engineering Society International Conference on Spatial Reproduction—Aesthetics and Science, Tokyo, Japan, 7–9 August 2018; Audio Engineering Society: Tokyo, Japan, 2018.
30. Algazi, V.R.; Duda, R.O.; Duraiswami, R.; Gumerov, N.A.; Tang, Z. Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.* **2002**, *112*, 2053–2064. [CrossRef]
31. Salvador, C.D.; Sakamoto, S.; Treviño, J.; Suzuki, Y. Validity of distance-varying filters for individual HRTFs on the horizontal plane. In Proceedings of the Spring Meeting Acoustic Society of Japan, Kawasaki, Japan, 15–17 March 2017; Acoustical Society of Japan: Kawasaki, Japan, 2017.
32. Rasumow, E.; Blau, M.; Hansen, M.; van de Par, S.; Doclo, S.; Mellert, V.; Püschel, D. Smoothing individual head-related transfer functions in the frequency and spatial domains. *J. Acoust. Soc. Am.* **2014**, *135*, 2012–2025. [CrossRef]