Ear Centering for Near-Distance Head-Related Transfer Functions

1st Ayrton Urviola
Perception Research
Lima, Peru
aurviola@perception3d.com

2nd Shuichi Sakamoto
RIEC and GSIS
Tohoku University
Sendai, Japan
saka@ais.riec.tohoku.ac.jp

3rd César D. Salvador
Perception Research
Lima, Peru
salvador@perception3d.com

***Abstract*—The head-related transfer functions (HRTF) are a major tool in spatial sound technology for personal use. They are linear filters describing the transmission of sound from a point in space to the ears. The HRTFs are typically obtained for a sparse set of points at a single far distance, from which datapoints at near distances are synthesized using the spherical Fourier transform (SFT) and distance-varying filters (DVF). Ear centering is further required to match the center of the SFT (the center of the head) and the measurement positions (the ears). Hitherto, plane-wave (PW) translation operators have yield effective ear centering when synthesizing HRTFs at far distances. We propose to use spherical-wave (SW) translation operators for ear centering when synthesizing HRTFs at near distances. We contrasted the performance of SW and PW ear centering. Synthesis errors decreased consistently when applying SW ear centering and the enhancement was observed up to the maximum frequency determined by the input far-distance dataset.**

***Index Terms*—Head-related transfer functions, acoustic centering, translation operator, spherical Fourier transform, distance-varying filter.**

I. Introduction

The head-related transfer functions (HRTF) are a major tool in spatial sound technology for personal use [1]–[3]. They are linear filters describing the transmission of sound from a point in space to the eardrums of a listener [4]. The HRTFs are commonly obtained for a sparse set of points at a single far distance from the center of the head, a distance greater than 1 m. Besides far-distance datasets, there is a growing interest in accurately synthesizing HRTFs for arbitrary points close to the head [5], [6]; research interests include near-field auditory displays [7] and auditory attention experiments [8]. A promising synthesis approach extrapolates near-distance HRTFs starting from far-distance ones using the spherical Fourier transform (SFT) and distance-varying filters (DVF) [9]–[11]. However, when using the SFT to represent spherical HRTF datasets, the mismatch between the center of the SFT (the center of the head) and the measurement positions (the ears) demands 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 ear position and the SFT center in the framework of a more general technique called acoustic centering [12]–[16]. Ear centering is performed by means of translation operators that relate sound pressures at the head center and ears [17]–[22]. Translation operators can be applied in free-field [17]–[19] or include a rigid sphere [20]–[22]; they can also operate in the spatial domain (the unit sphere) [18]–[22] or in the SFT domain [17]. 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 [19]. However, when PW translation operators are used to synthesize near-distance 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 HRTFs for sound sources close to the head.

We propose to use a free-field translation operator based on a spherical-wave (SW) model for ear centering in near-distance HRTF synthesis. The reminder of this paper is organized as follows: Sec. II formulates ear centering for near-distance HRTFs using translation operators, Sec. III compares PW and SW translation operators, Sec. IV describes considerations for practical implementations, and Sec. V states the conclusions.

II. Ear centering for near-distance HRTFs

In spherical coordinates, 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 $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. Positions in front of the listener lie along the positive x-axis or the direction $(\theta = 0, \phi = 0)$. Positive θ is measured from the positive x-axis to the left. 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.

Figure 1 shows the top-view geometry for theoretical near-distance HRTF synthesis. The center of the head coincides with the origin $\mathbf{0} = (0, 0, 0)$ and the ear position is denoted by $\mathbf{r}_{\mathrm{ear}} = (r_{\mathrm{ear}}, \theta_{\mathrm{ear}}, \phi_{\mathrm{ear}})$. 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.

This study started when Ayrton Urviola and César D. Salvador were at Silicon Integrated (SI) Co., Ltd, China. This study was partly supported by SI, China and JSPS KAKENHI Grant Number 19H04145, Japan.

Figure 2 overviews the synthesis process with ear centering. The input is a continuous, spherical distribution of free-field HRTFs from $\mathbf{a}$ to $\mathbf{r}_{\text{ear}}$, denoted by $\mathcal{H}(\mathbf{a},\mathbf{r}_{\text{ear}})$, whereas the output is a synthesized free-field HRTF from $\mathbf{b}$ to $\mathbf{r}_{\text{ear}}$, denoted by $\hat{\mathcal{H}}(\mathbf{b},\mathbf{r}_{\text{ear}})$. For simplicity, only the left ear is considered, however, the formulations below that relate the output to the input hold for both ears.

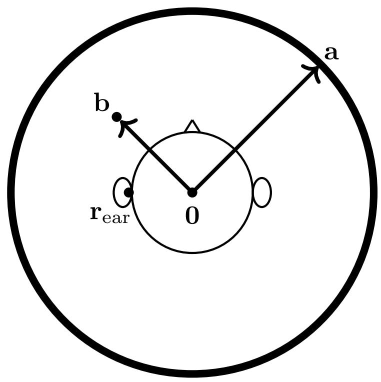

Fig. 1. Geometry for near-distance HRTF synthesis.

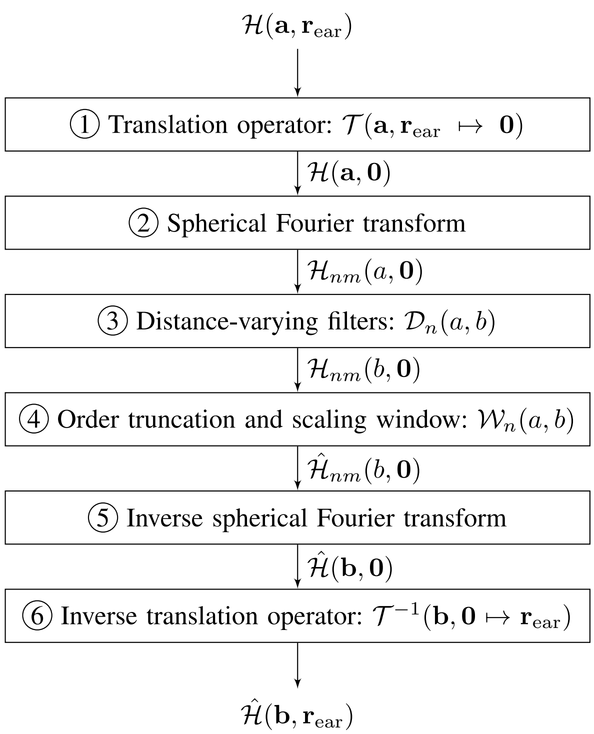

Fig. 2. Near-distance HRTF synthesis with ear centering.

Direct ear centering is performed by an operator $\mathcal{T}$ that translates the reference of the input from $\mathbf{r}_{\text{ear}}$ to $\mathbf{0}$ as follows:

$$\mathcal{H}(\mathbf{a},\mathbf{0}) = \mathcal{T}(\mathbf{a},\mathbf{r}_{\text{ear}} \mapsto \mathbf{0})\mathcal{H}(\mathbf{a},\mathbf{r}_{\text{ear}}). \tag{1}$$

The notation $\mathcal{H}(\mathbf{a},\mathbf{0})$ is used as conceptual support and by no means it indicates that an HRTF is obtained at the head center. If the translation was required to be applied to the SFT basis functions instead of the spherical data, it would be required a translation operator in the opposite direction, from $\mathbf{0}$ to $\mathbf{r}_{\text{ear}}$, as formulated in a more general manner in [12]. Formulations in this paper, however, are delimited to translations of spherical data in the spatial domain.

The PW translation operator in [19] is formulated as

$$\mathcal{T}_{\text{PW}}(\mathbf{a},\mathbf{r}_{\text{ear}} \mapsto \mathbf{0}) = e^{-jkr_{\text{ear}}\cos\Theta_{\mathbf{a},\mathbf{r}_{\text{ear}}}}, \tag{2}$$

where $\Theta_{\mathbf{a},\mathbf{r}_{\text{ear}}}$ denotes the angle between $\mathbf{a}$ and $\mathbf{r}_{\text{ear}}$. Considering a PW emanating from $\mathbf{a}$, (2) stems from the ratio of PW observations at $\mathbf{0}$ and $\mathbf{r}_{\text{ear}}$.

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

$$\mathcal{T}_{\text{SW}}(\mathbf{a},\mathbf{r}_{\text{ear}} \mapsto \mathbf{0}) = \frac{\|\mathbf{a}-\mathbf{r}_{\text{ear}}\|}{a} e^{-jk(a-\|\mathbf{a}-\mathbf{r}_{\text{ear}}\|)}, \tag{3}$$

where $\|\cdot\|$ denotes Euclidean norm. Considering a SW emanating from $\mathbf{a}$, (3) stems from the ratio of SW observations at $\mathbf{0}$ and $\mathbf{r}_{\text{ear}}$.

The SFT of ear-centered $\mathcal{H}(\mathbf{a},\mathbf{0})$ is defined by

$$\mathcal{H}_{nm}(a,\mathbf{0}) = \int_{-\pi}^{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathcal{H}(\mathbf{a},\mathbf{0}) Y_n^m(\theta_a,\phi_a)\cos(\phi_a)d\phi_a d\theta_a. \tag{4}$$

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

$$Y_n^m(\theta,\phi) = N_{nm}P_n^{|m|}(\sin\phi)\begin{cases}1, & m=0,\\ \sqrt{2}\cos(m\theta), & m>0,\\ \sqrt{2}\sin(|m|\theta), & m<0,\end{cases} \tag{5}$$

where P_n^m is the non-normalized associated Legendre polynomial [23] and N_{nm} is the following normalization factor

$$N_{nm} = (-1)^{|m|}\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}. \tag{6}$$

The real-valued basis functions in (5) are preferred to the complex-valued ones in [24] to avoid phase modifications during SFT representations.

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

$$\mathcal{H}_{nm}(b,\mathbf{0}) = \mathcal{D}_n(a,b)\mathcal{H}_{nm}(a,\mathbf{0}). \tag{7}$$

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

$$\mathcal{D}_n(a,b) = \frac{h_n^{(1)}(kb)}{h_n^{(1)}(ka)}, \tag{8}$$

where $h_n^{(1)}$ is the spherical Hankel function of the first kind and order n [25]. Because the ideal DVFs in (8) yield excessive values for higher orders and lower frequencies [11], their action need to be limited according to

$$\hat{\mathcal{H}}_{nm}(b,\mathbf{0}) = \mathcal{W}_n(a,b)\mathcal{H}_{nm}(b,\mathbf{0}) \tag{9}$$

with an order truncation and scaling window defined as [26]:

$$\mathcal{W}_n(a,b) = \begin{cases}\frac{b}{a}e^{-jk(b-a)}, & n \le \min(\lceil kr_{\text{h}}\rceil, N),\\ 0, & \text{otherwise.}\end{cases} \tag{10}$$

Here, r_{h} is the radius of a sphere fully containing the head, the rule $n \leq \lceil kr_{\mathrm{h}} \rceil$ indicates the far-to-near field transition, and n ranges from 0 to the maximum order N constrained by the spherical sampling scheme used in practice.

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

$$\hat{\mathcal{H}}(\mathbf{b}, \mathbf{0}) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \mathcal{H}_{nm}(b, \mathbf{0}) Y_n^m(\theta_b, \phi_b). \tag{11}$$

The maximum order N indicates that the sum is truncated up to the first $(N+1)^2$ SFT basis functions.

Finally, inverse ear centering is performed with the inverse operator $\mathcal{T}^{-1}$ that translates the reference from $\mathbf{0}$ to $\mathbf{r}_{\mathrm{ear}}$:

$$\mathcal{H}(\mathbf{b}, \mathbf{r}_{\mathrm{ear}}) = \mathcal{T}^{-1}(\mathbf{b}, \mathbf{0} \mapsto \mathbf{r}_{\mathrm{ear}}) \mathcal{H}(\mathbf{b}, \mathbf{0}). \tag{12}$$

The inverse PW translation operator [19] can be expressed as

$$\mathcal{T}_{\mathrm{PW}}^{-1}(\mathbf{b}, \mathbf{0} \mapsto \mathbf{r}_{\mathrm{ear}}) = e^{jkr_{\mathrm{ear}} \cos \Theta_{\mathbf{b}, \mathbf{r}_{\mathrm{ear}}}}. \tag{13}$$

Considering a PW emanating from $\mathbf{b}$, (13) stems from the ratio of PW observations at $\mathbf{r}_{\mathrm{ear}}$ and $\mathbf{0}$.

The proposed inverse SW translation operator takes the form

$$\mathcal{T}_{\mathrm{SW}}^{-1}(\mathbf{b}, \mathbf{0} \mapsto \mathbf{r}_{\mathrm{ear}}) = \frac{b}{\|\mathbf{b} - \mathbf{r}_{\mathrm{ear}}\|} e^{-jk(\|\mathbf{b} - \mathbf{r}_{\mathrm{ear}}\| - b)}. \tag{14}$$

Considering a SW emanating from $\mathbf{b}$, (14) stems from the ratio of SW observations at $\mathbf{r}_{\mathrm{ear}}$ and $\mathbf{0}$.

III. Evaluation of PW and SW translation

This section compares the performance of PW and SW translation operators in the framework of a spherical sampling of the theory presented in Sec. II. The SFT and DVF algorithms described in [27] and [11], respectively, were adapted to the purposes of our evaluations.

A. Conditions

Left-ear HRTFs for two head models (without torso) of two individuals subjects available in the calculated near-distance dataset in [28] were used in evaluations. The datasets have 512 samples along time, sampled at 48 kHz. 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. The number of points P in an icosahedral grid, generated with a subdivision factor q, is

$$P = 10q^2 + 2. \tag{15}$$

For almost regular spherical samplings, such as the icosahedral ones, the maximum SFT order achievable with P points is

$$N_{\mathrm{grid}} = \lfloor \sqrt{P} \rfloor - 1. \tag{16}$$

It ensures reliable synthesis up to a maximum frequency

$$f_{\max} = \frac{cN_{\mathrm{grid}}}{2\pi r_{\mathrm{h}}}, \tag{17}$$

where r_{h} is the same radius used in (10) and c is the speed of sound in air.

Datasets at a distance $a = 100$ cm were used as inputs. Four icosahedral grids with $P = 12, 42, 162, 252$, correspondingly $q = 1, 2, 4, 5$, and $N_{\mathrm{grid}} = 2, 5, 11, 14$, 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:

$$N = \min(\lceil kr_{\mathrm{h}} \rceil, N_{\mathrm{grid}}). \tag{18}$$

In (10), (17), and (18), r_{h} should ideally be the radius of the smallest sphere containing a head model. However, this theoretical limit yielded artifacts due to the discontinuities of the truncation rule $n \leq \lceil kr_{\mathrm{h}} \rceil$. To reduce these artifacts, we have empirically chosen $r_{\mathrm{h}} = 16$ cm as a convenient value for the two individual head models in [28]. The speed of sound in air, $c = 344$ m/s, was the same used in [28].

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; ear centering with the PW translation operators in (2) and (13); and ear centering with the SW translation operators in (3) and (14).

B. 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 $i = 1, 2, ..., 81$ indicates radial distances; index $j = 1, 2, ..., 642$, directions on the sphere with $\Omega_j = (\theta_j, \phi_j)$; index $\kappa = 1, 2, ..., 257$, frequency bins; and index $\ell = 1, 2$, individual subjects. The synthesis error is defined as

$$E(b_i, f_\kappa) = \underset{s_\ell}{\mathrm{RMS}} \left\{ \frac{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H} - \hat{\boldsymbol{H}}\}}{\underset{\Omega_j}{\mathrm{RMS}}\{\boldsymbol{H}\}} \right\}, \tag{19}$$

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

C. Results

All panels in Fig. 3 show synthesis errors calculated with (19) and displayed in a logarithmic scale, from 0 to -30 dB, to contrast with the perceivable HRTF dynamic range of around 30 dB, as reported in [29]. The black-dashed curves highlight the -15 dB values and are used as an indicator to compare among panels. Values around -15 dB for the metric in (19) have also been used in previous research on HRTF synthesis [17] and sound field reconstruction [30], [31]. The black-dashed lines indicate $f_{\max}$ in (17).

In Fig. 3, left-column panels (a), (d), (g), and (j) corresponds to synthesis without ear centering; center-column panels (b), (e), (h), and (k), to ear centering with the PW translation operators in (2) and (13); and right-column panels (c), (f), (i), and (l), to ear centering with the SW translation operators in (3) and (14). First-row panels (a), (b), and (c) correspond to synthesis from $P = 252$ points ($q = 5$, $N_{\mathrm{grid}} = 14$); second-row panels (d), (e), and (f), to synthesis from $P = 162$ points

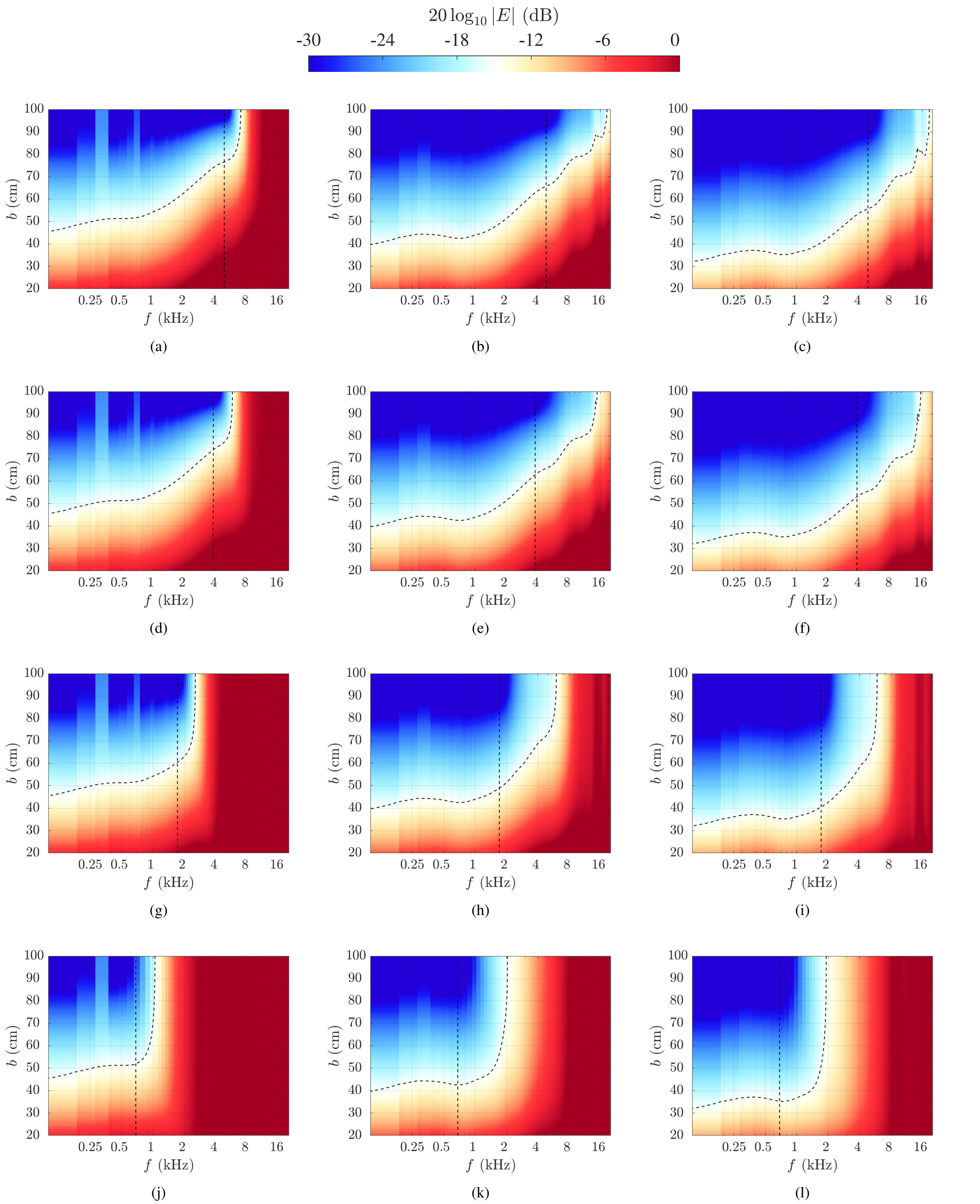

Fig. 3. Synthesis error (19) in dB. Black-dashed curves indicate -15 dB values. Black-dashed lines indicate $f_{\max}$ in (17). Left-column panels: No ear centering. Center-column panels: Ear centering with PW translation in (2) and (13). Right-column panels: Ear centering with SW translation in (3) and (14). Panels (a), (b), and (c): $P = 252$, $q = 5$, and $N_{\text{grid}} = 14$. Panels (d), (e), and (f): $P = 162$, $q = 4$, and $N_{\text{grid}} = 11$. Panels (g), (h), and (i): $P = 42$, $q = 2$ and $N_{\text{grid}} = 5$. Panels (j), (k), and (l): $P = 12$, $q = 1$ and $N_{\text{grid}} = 2$.

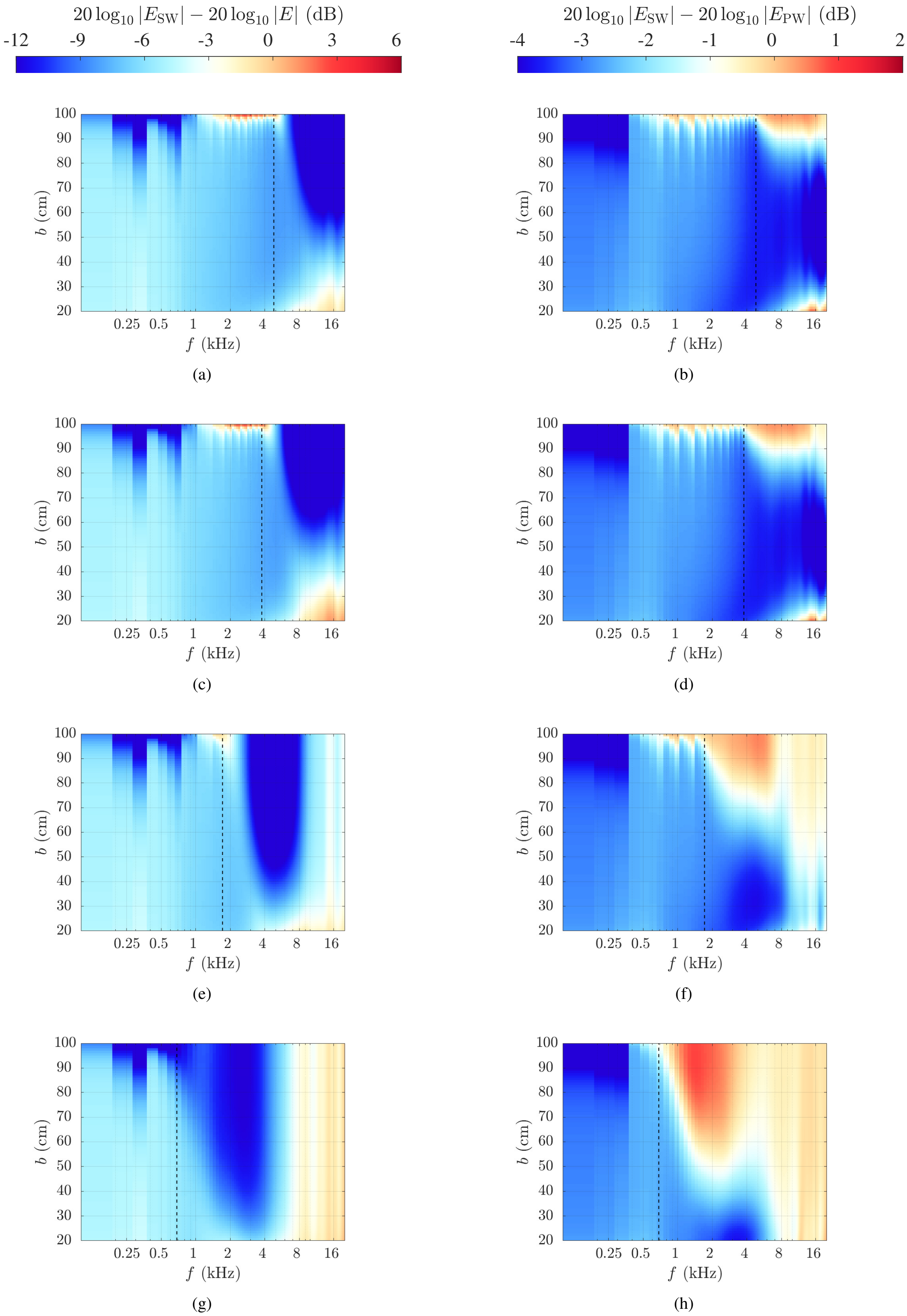

Fig. 4. Difference between results in Fig. 3. Black-dashed lines indicate $f_{\max}$ in (17). Left-column panels: Difference between SW ear centering and No ear centering. Right-column panels: Difference between SW ear centering and PW ear centering. Panels (a) and (b): $P = 252$, $q = 5$, and $N_{\mathrm{grid}} = 14$. Panels (c) and (d): $P = 162$, $q = 4$, and $N_{\mathrm{grid}} = 11$. Panels (e) and (f): $P = 42$, $q = 2$ and $N_{\mathrm{grid}} = 5$. Panels (g) and (h): $P = 12$, $q = 1$ and $N_{\mathrm{grid}} = 2$.

($q = 4$, $N_{\text{grid}} = 11$); third-row panels (g), (h), and (i), to synthesis from $P = 42$ points ($q = 2$, $N_{\text{grid}} = 5$); and last-row panels (j), (k), and (l), to synthesis from $P = 12$ points ($q = 1$, $N_{\text{grid}} = 2$).

When comparing panels along rows in Fig. 3, it is observed that applying SW ear centering outperforms the accuracy across all distances when SW results are contrasted with PW ear centering and no ear centering. The enhancements of SW ear centering are more noticeable at near distances and their benefits extend even beyond the corresponding $f_{\max}$. A closer inspection of the intersections between black-dashed lines ($f_{\max}$) and black-dash curves (-15 dB) shows that, below $f_{\max}$ and for the same error levels, SW ear centering yields an improvement of nearly 10 cm closer to the head when compared to PW ear centering.

Panels (b), (e), (h), and (k) in Fig. 3 show that, for frequencies up to $f_{\max}$ in panel (k), PW ear centering yields similar accuracies across all distances. For the same value of $f_{\max}$, now in panel (l), it is observed that the results of SW ear centering in panel (l) outperforms those of PW ear centering in panels (b), (e), (h), and (k) across all distances. SW ear centering, therefore, outperforms PW ear centering in the sense of enabling the reduction of the required number of points in the input HRTF dataset without compromising the accuracy.

All panels in Fig. 4 show differences between the synthesis errors in Fig. 3. The errors obtained with PW ear centering and No ear centering are subtracted from the errors obtained with SW ear centering. Negative values in dB towards the blue colors indicate the regions were SW ear centering outperforms No ear centering and PW ear centering. The black-dashed lines indicate $f_{\max}$ in (17). Left-column panels (a), (c), (e), and (g) show differences between SW ear centering and No ear centering. Right-column panels (b), (d), (f), and (h) show differences between SW ear centering and PW ear centering. First-row panels (a) and (b) correspond to synthesis from $P = 252$ points ($q = 5$, $N_{\text{grid}} = 14$); second-row panels (c) and (d), to synthesis from $P = 162$ points ($q = 4$, $N_{\text{grid}} = 11$); third-row panels (e) and (f), to synthesis from $P = 42$ points ($q = 2$, $N_{\text{grid}} = 5$); and last-row panels (g), and (h), to synthesis from $P = 12$ points ($q = 1$, $N_{\text{grid}} = 2$).

Panels (a), (c), (e), and (g) in Fig. 4 show that, in frequencies below $f_{\max}$, SW ear centering outperforms No ear centering across all distances, yielding an overall improvement of 6 dB. Panels (b), (d), (f), and (h) show that, in frequencies below $f_{\max}$, SW ear centering also outperforms PW ear centering across all distances, offering an overall enhancement of 3 dB. Moreover, at distances below 30 cm, the 3 dB enhancement holds beyond $f_{\max}$.

IV. Considerations for practical implementations

By adding few computational power, steps ① and ⑥ of the proposal in Fig. 2, to a standard method of HRTF synthesis, steps ② to ⑤ of Fig. 2, we can obtain more accurate HRTFs for spatialization applications. Furthermore, these additional steps are easy to implement and to incorporate into already available spatializers as they are independant of the SFT.

Figure 5 illustrates the generation of binaural signals from the convolution of a monofonic signal with head-related impulse responses (HRIR) for an arbitrary position $\mathbf{b}$. HRIRs at $\mathbf{b}$ are calculated with the proposal in Fig. 2, which is divided into two stages: off-line analysis and on-line synthesis. Off-line analysis takes a sparse HRIR dataset together with the far source positions $\mathbf{a}$ and ear positions $\mathbf{r}_{\text{ears}}$ as inputs; a fast Fourier transform (FFT) along time converts HRIRs into HRTFs; subsequently, the steps ① and ② of Fig. 2 provide a SFT representation. Off-line analysis is only updated when the sparse HRIR dataset changes among available generic and individual options. On-line synthesis, on the other hand, is updated in real-time as $\mathbf{b}$ changes. On-line synthesis consists of applying the steps from ③ to ⑥ of Fig. 2, followed by an inverse fast Fourier transform (IFFT) that finally converts HRTFs into HRIRs as required by the convolution engine. Algorithms to implement real-time convolution engines can be found in [32].

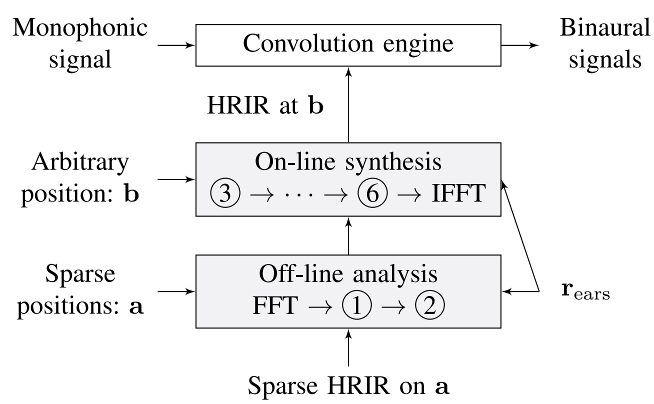

Fig. 5. Spatialization with near-distance HRIRs.

For each frequency bin, Table I details the process shown in Fig. 2 for one ear. From left to right, the first column states each one of the six steps in Fig. 2. The second column describes the operations involved in each step. The third column displays the dimensions of the operands for each operation in the previous column. The last column shows the algorithmic complexity of each operation in big-O notation $\mathcal{O}$, considering the complex-domain multiplication as the constant time complexity $\mathcal{O}(1)$ [33]. The algorithmic complexities shown take into account $N_{\text{grid}} \geq N$ and $(N_{\text{grid}}+2)^2 > P \geq (N\text{grid}+1)^2$.

The off-line process consists of steps ① and ②. Step ① takes one vector of P elements, the sampled HRTF, and another vector of P elements, the translation operator, performs an element-wise multiplication, and returns one vector of P elements, the translated HRTF, with P described in (15). Step ② takes one $P \times (N+1)^2$ matrix, the spherical harmonics, and one $P \times 1$ vector, the translated HRTF, performs a matrix inversion and then a matrix multiplication between the inverted $(N+1)^2 \times P$ matrix and the $P \times 1$ vector, and returns one $(N+1)^2 \times 1$ vector, the translated SFT coefficients of the HRTF, with N described in (18). The overall complexity of the off-line process is given by the complexity of the matrix

TABLE I
COMPLEXITY OF OPERATIONS IN FIG. 2

Step	Operation	Dimensions of operands	Algorithmic Complexity
1	Element-wise multiplication	2 vectors of P elements	$\mathcal{O}(N_{\text{grid}}^2)$
2	Matrix inversion	$P \times (N+1)^2$ matrix	$\mathcal{O}(N_{\text{grid}}^4 N^2)$
	Matrix multiplication	$(N+1)^2 \times P$ matrix, $P \times 1$ vector	$\mathcal{O}(N_{\text{grid}}^2 N^2)$
3	Element-wise multiplication	2 vectors of $(N+1)^2$ elements	$\mathcal{O}(N^2)$
4	Element-wise multiplication	2 vectors of $(N+1)^2$ elements	$\mathcal{O}(N^2)$
5	Dot product	2 vectors of $(N+1)^2$ elements	$\mathcal{O}(N^2)$
6	Complex-domain multiplication	2 complex-domain numbers	$\mathcal{O}(1)$

inversion operation $\mathcal{O}(N_{\text{grid}}^4 N^2)$, which for high frequencies when $N \to N_{\text{grid}}$ becomes $\mathcal{O}(N_{\text{grid}}^6)$.

The on-line process consists of steps 3, 4, 5 and 6. Step 3 takes one vector of $(N+1)^2$ elements, the translated SFT coefficients of the HRTF, and another vector of $(N+1)^2$ elements, the DVFs, performs an element-wise multiplication, and returns one vecto of $(N+1)^2$ elements, the near-distance translated SFT coefficients of the HRTF. Step 4 takes one vector of $(N+1)^2$ elements, the near-distance translated SFT coefficients of the HRTF, and another vector of $(N+1)^2$ elements, the scaling window, performs an element-wise multiplication, and returns one vector of $(N+1)^2$ elements, the scaled near-distance translated SFT coefficients of the HRTF. Step 5 takes one vector of $(N+1)^2$ elements, the spherical harmonics, and another vector of $(N+1)^2$ elements, the scaled near-distance translated SFT coefficients of the HRTF, performs a dot product, and returns one complex number, the near-distance translated HRTF. Step 6 takes one complex number, the inverse translation operator, and another complex number, the near-distance translated HRTF, performs a complex-domain multiplication, and returns one complex number, the HRTF at an arbitrary position $\mathbf{b}$. The overall complexity of the on-line process is $\mathcal{O}(N^2)$, which for high frequencies when $N \to N_{\text{grid}}$ becomes $\mathcal{O}(N_{\text{grid}}^2)$.

V. CONCLUSION

We proposed a spherical-wave translation operator that performs ear centering when synthesizing head-related transfer functions for sound sources close to the head. We contrasted the performance of our proposal and the existing plane-wave translation operator. Synthesis accuracy increased consistently with spherical-wave ear centering when compared to plane-wave ear centering. Enhancements were observed at near distances and for frequencies within the range of operation determined by the spherical resolution of the input dataset.

Extensions to this work might include regularization techniques to optimize the use of basis functions in spherical Fourier transforms during the synthesis process. Another extension might consider the inclusion of distance information in non-free-field translation operators such as the ones based on a rigid sphere. Perceptual evaluations by means of detectability of differences and localization tests could also provide more insight into the validity of the suggested approach.

ACKNOWLEDGMENT

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

REFERENCES

[1] V. Algazi and R. Duda, “Headphone-based spatial sound,” *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 33–42, Jan. 2011.

[2] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Design theory for binaural synthesis: Combining microphone array recordings and head-related transfer function datasets,” *Acoust. Sci. Technol.*, vol. 38, no. 2, pp. 51–62, Mar. 2017.

[3] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, “Surround by sound: a review of spatial audio recording and reproduction,” *Appl. Sci.*, vol. 7, no. 5, 2017.

[4] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*, revised ed. Cambridge, MA, USA; London, England.: MIT Press, 1997.

[5] S. T. Prepeliţă, J. G. Bolaños, V. Pulkki, L. Savioja, and R. Mehra, “Numerical simulations of near-field head-related transfer functions: Magnitude verification and validation with laser spark sources,” *J. Acoust. Soc. Am.*, vol. 148, no. 1, pp. 153–166, 2020.

[6] J. M. Arend, H. R. Liesefeld, and C. Pörschmann, “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 Ac.*, vol. 5, p. 10, 2021.

[7] D. S. Brungart, “Near-Field Virtual Audio Displays,” *Presence: Teleop. Virt. Env.*, vol. 11, no. 1, pp. 93–106, Feb. 2002.

[8] S. Sakamoto, F. Monasterolo, C. D. Salvador, Z. Cui, and Y. Suzuki, “Effects of target speech distance on auditory spatial attention in noisy environments,” in *Proc. ICA 2019 and EAA Euroregio*, Aachen, Germany, Sep. 2019, pp. 2177–2181.

[9] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, “Interpolation and range extrapolation of HRTFs,” in *Proc. IEEE ICASSP*, vol. 4, May 2004, pp. 45–48.

[10] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, “Calculation of head-related transfer functions for arbitrary field points using spherical harmonics,” *Acta Acust. United Ac.*, vol. 98, no. 1, pp. 72–82, Jan. 2012.

[11] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, “Distance-varying filters to synthesize head-related transfer functions in the horizontal plane from circular boundary values,” *Acoust. Sci. Technol.*, vol. 38, no. 1, pp. 1–13, Jan. 2017.

[12] N. A. Gumerov and R. Duraiswami, *Fast multipole methods for the Helmholtz equation in three dimensions*, ser. Elsevier Series in Electromagnetism. Maryland, USA: Elsevier, 2004.

[13] I. Ben Hagai, M. Pollow, M. Vorländer, and B. Rafaely, “Acoustic centering of sources measured by surrounding spherical microphone arrays,” *J. Acoust. Soc. Am.*, vol. 130, no. 4, pp. 2003–2015, 2011.

[14] N. R. Shabtai and M. Vorländer, "Acoustic centering of sources with high-order radiation patterns," *J. Acoust. Soc. Am.*, vol. 137, no. 4, pp. 1947–1961, 2015.
[15] Y. Wang and K. Chen, "Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3474–3478, 2018.
[16] M. Kentgens and P. Jax, "Translation of a higher-order ambisonics sound scene by space warping," in *Proc. Audio Eng. Soc. Int. Conf. Audio for Virtual and Augmented Reality*, Aug. 2020.
[17] J.-G. Richter, M. Pollow, F. Wefers, and J. Fels, "Spherical harmonics based HRTF datasets: Implementation and evaluation for real-time auralization," *Acta Acust. United Ac.*, vol. 100, no. 4, pp. 667–675, Jul. 2014.
[18] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627, 2018.
[19] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment," *IEEE Trans. Audio, Speech, Language Process.*, pp. 1–1, 2019.
[20] C. Pörschmann, J. M. Arend, and F. Brinkmann, "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 6, pp. 1060–1071, 2019.
[21] ——, "Correction to "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling" [Jun 19 1060-1071]," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2194–2194, 2020.
[22] J. M. Arend, F. Brinkmann, and C. Pörschmann, "Assessing spherical harmonics interpolation of time-aligned head-related transfer functions," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 104–117, Jan. 2021.
[23] F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, and H. S. Schneider, Eds., *NIST Digital Library of Mathematical Functions*, 1st ed., Dec. 2020. [Online]. Available: http://dlmf.nist.gov/
[24] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. London, UK: Academic Press, 1999.
[25] U. Rehmann, *Encyclopedia of Mathematics*, 2020. [Online]. Available: https://encyclopediaofmath.org/
[26] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, "Validity of distance-varying filters for individual HRTFs on the horizontal plane," in *Proc. Spring Meeting Acoust. Soc. Jpn.* Kawasaki: Acoustical Society of Japan, Mar. 2017.
[27] ——, "Boundary matching filters for spherical microphone and loudspeaker arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 461–474, Mar. 2018.
[28] ——, "Dataset of near-distance head-related transfer functions calculated using the boundary element method," in *Proc. Audio Eng. Soc. Int. Conf. Spatial Reproduction —Aesthetics and Science—*. Tokyo, Japan: Audio Engineering Society, Aug. 2018. [Online]. Available: https://cesardsalvador.github.io/download.html
[29] E. Rasumow, M. Blau, M. Hansen, S. van de Par, S. Doclo, V. Mellert, and D. Püschel, "Smoothing individual head-related transfer functions in the frequency and spatial domains," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 2012–2025, 2014.
[30] D. B. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 6, pp. 697–707, Sep. 2001.
[31] T. Shimizu, J. Treviño, S. Sakamoto, Y. Suzuki, and T. Ise, "Evaluation of the extension and coloration of multiple listening zones synthesized by the shared field reproduction system," *Acoust. Sci. Technol.*, vol. 40, no. 4, pp. 241–249, 2019.
[32] F. Wefers, *Partitioned convolution algorithms for real-time auralization*. Logos Verlag Berlin GmbH, 2015, vol. 20.
[33] R. R. Howell, "On asymptotic notation with multiple variables," *Tech. Rep.*, 2008. [Online]. Available: https://people.cs.ksu.edu//~rhowell/asymptotic.pdf