Article

Particle-Velocity-Based Mixed-Source Sound Field Translation for Binaural Reproduction

Huanyu Zuo [1],[*], Lachlan I. Birnie [1], Prasanga N. Samarasinghe [1], Thushara D. Abhayapala [1] and Vladimir Tourbabin [2]

[1] Audio and Acoustic Signal Processing Group, College of Engineering and Computer Science, The Australian National University, Canberra, ACT 2601, Australia
[2] Meta Reality Labs Research, Redmond, WA 98052, USA
[*] Correspondence: huanyu.zuo@anu.edu.au

Abstract: Following the rise of virtual reality is a demand for sound field reproduction techniques that allow the user to interact and move within acoustic reproductions with six-degrees-of-freedom. To this end, a mixed-source model of near-field and far-field virtual sources has been introduced to improve the performance of sound field translation in binaural reproductions of spatial audio recordings. The previous works, however, expand the sound field in terms of the mixed sources based on sound pressure. In this paper, we develop a new mixed-source expansion based on particle velocity, which contributes to more precise reconstruction of the interaural phase difference and, therefore, contributes to improved human perception of sound localization. We represent particle velocity over space using velocity coefficients in the spherical harmonic domain, and the driving signals of the virtual mixed-sources are estimated by constructing cost functions to optimize the velocity coefficients. Compared to the state-of-the-art method, sound-pressure-based mixed-source expansion, we show through numerical simulations that the proposed particle-velocity-based mixed-source expansion has better reconstruction performance in sparse solutions, allowing for sound field translation with better perceptual immersion over a larger space. Finally, we perceptually validate the proposed method through a Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) experiment for a single source scenario. The experimental results support the better perceptual immersion of the proposed method.

Keywords: sound field translation; binaural reproduction; particle velocity; mixed-source model; MUSHRA

1. Introduction

Sound field translation plays an important role in binaural reproduction systems [1], which enables a listener to move about an acoustic space with a sustained perceptual immersion. A typical example of its application is in virtual-reality reproductions of real-world experiences, where the listener is allowed to virtually explore by moving their body/head and they can experience a spatially accurate perception that is the same as in the real-world space [2]. In this paper, we propose a particle-velocity-based mixed-source expansion that can provide a wide range of sound field translation with immersive human perception.

There have been various studies of sound field translation that allow for a listener to interact and move within a sound field reproduction, and they can be divided into two main categories: interpolation-based techniques [3] and extrapolation-based techniques [4]. The interpolation-based techniques aim to interpolate the sound field to listeners during reproduction using a grid of higher-order microphones (inpractice, higher-order microphones are designed from several pressure microphones equally spaced on a solid sphere) over space [5]. For example, a simple interpolation-based technique is shown in [6], where

the room impulse response is interpolated to enable real-time auralizations with head rotation. The interpolation-based techniques usually suffer from a significant localization error [5,7] and comb-filtering spectral distortions [8], in particular when the sound source is nearer to one microphone than to another. Although the comb-filtering spectral distortions can be avoided using a parametric Ambisonics room impulse response interpolation system [9], it requires extra effort to compensate for varying interference. In addition, the interpolation assumes multiple distributed microphones or even arrays of microphones, which is a significant hardware limitation in many Augmented Reality/Virtual Reality (AR/VR) applications [3,10]. To enhance the performance of perceptual source localization for displaced listeners, a modified method is proposed, where the directional components of the microphone nearest to the listener are emphasized [11]. This method also interpolates the sound field to listeners and is subject to the fundamental limitation. Methods that alleviate the fundamental limitation have been investigated; however, they usually require additional source localization and separation of direct sound field components [12–14]. On the other hand, the extrapolation-based techniques spherically extend the translation from the recording microphone to the nearest source, which can be achieved by Ambisonics [15,16], discrete virtual sources [17], plane-wave expansion [18–20], and the equivalent source method [21]. A comparison between different extrapolation-based techniques is given in [22]. To improve the performance of translation and source localization, especially for off-center positions, optimized decoding methods such as max-r_{E} (meaning max energy vector magnitude) decoding [23] and near-field-compensated decoding [24] have been proposed. However, most of the extrapolated-based techniques are limited by the order of the recordings, which is determined by both the upper frequency band and the microphone radius [25,26]. Consequently, for a low-order recording, the range of translation is significantly limited. Attempts to move beyond this limited range, even after extrapolation, result in spectral distortions [27–29], degraded source localization [22,30], and a poor perceptual listening experience.

Parametric models have been shown to provide an efficient way to describe sound scenes [14,31,32], which can be used in binaural reproduction. To overcome the limitation of translation range discussed above, a parametric model is built to describe first-order Ambisonics recordings for translation with known information of a listener's position and source distances [33]. A similar approach was described in [34], which extends to support arbitrary listener movement; however, with multiple source directions and distances. In [35], the authors improve the spatial localization accuracy by additional spatial information obtained from multi-perspective Ambisonics recordings. Higher order Ambisonics signals are used in [36], which translates the sound field based on multi-directional decomposition that estimates source distances. Another parametric decomposition from a higher order Ambisonics signal has been proposed in [37], offering excellent localization performance even for strong spatial displacements. However, it also requires a prior knowledge of source distance information and an additional analysis stage of direction of arrival estimation. The sound field translation with very few measurements in a room is investigated in [38]. Recently, an alternative secondary source model for extrapolated virtual reproduction has been developed, where it sparsely expands the recording into the equivalent virtual environment using a mix of near-field and far-field virtual sources [39]. It has been shown that this method can relax the limitation and offer an improved perceptual experience through perceptual listening tests [40]. However, this method expands the sound field based on sound pressure, which is not an acoustic quantity directly linked with human perception of sound localization.

Since Gerzon first developed velocity theory of sound localization based on binaural phase cues for reproducing psychoacoustically optimum sound [41], particle velocity (vector), which is recently controlled in sound field reproduction systems [42–44], has been one of the objective metrics predicting sound direction. Thus, having particle velocity as the design criterion contributes to a more precise reconstruction of the interaural phase difference, which is crucial for human perception of sound directions and, therefore,

contributes to improved localization of the reconstructed sound, especially for low frequencies (below 700 Hz) [45]. For example, a particle-velocity-based sound field reproduction method is proposed to improve the perception experience for listeners away from the center of the reproduction area [46]. Furthermore, the particle velocity-controlled sound field shows an accurate reproduction of sound intensity in the reproduction region [47,48], where sound intensity is another important acoustic quantity for localization perception; however, it is most appropriate at high frequencies (above 500 Hz) [49–51]. In this paper, we propose a new mixed-source expansion based on particle velocity for sound field translation and synthesis. We aim to reproduce a spatial acoustic scene by matching particle velocity of the sound field over a region with a sparse distribution of virtual sound sources placed in both the near and far field. We will show that sparsely expanding the sound field in terms of particle velocity offers a more immersive perceptual experience at the translated positions, especially for source localization.

The remainder of this paper is structured as follows. First, the sound field translation problem is formulated in Section 2. We have recently formulated continuous particle velocity expressions over space in the spherical harmonic domain [44], which contain the directivity information of a sound field over continuous spatial regions. In Section 3, we review the continuous particle velocity theory and develop the particle-velocity-based expansion by exploiting this theory in the mixed-source model. Section 4 translates the sound field and synthesizes binaural signals for translated listeners. In Section 5, we introduce two localization metrics, in addition to the relative pressure error, as evaluation criteria. With these criteria, we demonstrate the reproduction accuracy of the proposed method by comparing it with the state-of-the-art method, sound-pressure-based mixed-source expansion, through numerical simulations. The perceptual validation through a MUSHRA experiement is presented in Section 6. Finally, the work is concluded in Section 7.

2. Problem Formulation

Consider a listener within an acoustic environment, where the listening space of the environment is centered by the origin $\boldsymbol{o}$. The listener is positioned at $\boldsymbol{d} = (r, \theta, \phi)$, where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ in spherical coordinates, with respect to the origin. Let there be a total of K sound sources outside the listening space, with the κth source located at $\boldsymbol{b}_\kappa$. The free-field binaural sound perceived by the listener can be synthesized by filtering the source signals with the listener's head-related transfer functions (HRTFs), expressed as

$$P_{\mathrm{L,R}}(\boldsymbol{d}, k) = \sum_{\kappa=1}^{K} H_{\mathrm{L,R}}(k, \boldsymbol{b}_\kappa; \boldsymbol{d}) s_\kappa(k),\tag{1}$$

where $P_{\mathrm{L,R}}(\boldsymbol{d}, k)$ is the sound pressure at the listener's left and right ear, $H_{\mathrm{L,R}}(k, \boldsymbol{b}_\kappa; \boldsymbol{d})$ are the HRTFs between the κth source and the listener's ears, $s_\kappa(k)$ is the sound signal of the κth source, $k = 2\pi f / c$ is the wave number, f is frequency, and c is the speed of sound propagation.

We aim to record and reproduce the auditory experience of (1) for every possible listening position. The acoustic environment can be characterized using a spherical harmonic decomposition of the sound pressure, expressed as

$$P(\boldsymbol{x}, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \alpha_{nm}(k) j_n(k|\boldsymbol{x}|) Y_{nm}(\hat{\boldsymbol{x}}),\tag{2}$$

where $\boldsymbol{x}$ is an arbitrary position in the space, $|\cdot| \equiv r$, $\hat{\cdot} \equiv (\theta, \phi)$, $\alpha_{nm}(k)$ are the pressure coefficients which completely describe the source-free acoustic environment centered about

o, $j_n(\cdot)$ is the nth order spherical Bessel function of the first kind, and $Y_{nm}(\hat{x})$ is the spherical harmonic of order n and degree m, defined by

$$Y_{nm}(\hat{x}) \triangleq \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_{nm}(\cos\theta)e^{im\phi}, \tag{3}$$

where $P_{nm}(\cos\theta)$ is the associated Legendre function. In practice, the pressure coefficients describing the acoustic environment $\{\alpha_{nm}(k)\}$ can be estimated from recordings of a higher order microphone or a microphone array [52]. We now consider an Nth order receiver centered at o, such as a spherical array [53] or planar array [54], that we use to estimate the acoustic environment for $\{\alpha_{nm}(k)\}$ up to order N. However, the limited order N introduces a spatial reproduction constraint that the sound field can only be accurately reproduced within the receiver region (i.e., $|x| \leq R$), where R is the radius of the receiver region that is related by $N = \lceil kR \rceil$. This constraint restricts the listener to move within the receiver region (i.e., $|d| \leq R$). If the listener attempts to move beyond the receiver region, they would experience spectral distortions and a loss in perceptual immersion.

The objective of this work is to perceptually enhance the performance of translation to $|d| > R$ for binaural reproduction from a mode-limited measurement so that a listener can move over a large space in a virtual reconstruction with a sustained perceptual immersion, through the proposed particle-velocity-based mixed-source expansion.

3. Particle-Velocity-Based Mixed-Source Expansion

In this section, we present the theory of particle-velocity-based mixed-source expansion. We first give a brief introduction of the background knowledge involving the proposed theory, i.e., the representation of particle velocity in the spherical harmonic domain [44] in Section 3.1, and the mixed-source model [39] in Section 3.2. The particle-velocity-based mixed-source expansion is developed in Section 3.3.

3.1. Spherical Harmonic Representation of Particle Velocity

Particle velocity is a vector and it can be expressed in spherical coordinates, as

$$\boldsymbol{V}(\boldsymbol{x},k) = V_r(\boldsymbol{x},k)\hat{\boldsymbol{r}} + V_\theta(\boldsymbol{x},k)\hat{\boldsymbol{\theta}} + V_\phi(\boldsymbol{x},k)\hat{\boldsymbol{\phi}}, \tag{4}$$

where $\hat{\boldsymbol{r}}$, $\hat{\boldsymbol{\theta}}$, and $\hat{\boldsymbol{\phi}}$ are the unit vectors in the r, θ, and ϕ directions, respectively. We have recently formulated continuous particle velocity in the spherical harmonic domain [44], which decomposes each component of the particle velocity vector in terms of spherical harmonic functions, as

$$V_\Psi(\boldsymbol{x},k) = \sum_{p=0}^{Q_\Psi} \sum_{q=-p}^{p} X_{pq}^{(\Psi)}(k,|\boldsymbol{x}|)Y_{pq}(\hat{x}); \quad \Psi \in \{r,\theta,\phi\}, \tag{5}$$

where Q_Ψ is the particle velocity truncation order (the spherical harmonic decomposition of particle velocity also contains infinite summation terms, and we truncate it to an explicit order for implementation) in the Ψ direction, and $X_{pq}^{(\Psi)}(k,|\boldsymbol{x}|)$ denote the velocity coefficients in the Ψ direction. According to the results in [44], the velocity coefficients are given as

$$X_{pq}^{(r)}(k,|\boldsymbol{x}|) = \frac{i}{k\rho_0 c}\alpha_{pq}(k)j_p'(k|\boldsymbol{x}|), \tag{6}$$

$$X_{pq}^{(\theta)}(k,|\boldsymbol{x}|) = \frac{2\pi i}{k\rho_0 c}H_{pq}\sum_{n=|q|}^{N}H_{nq}\alpha_{nq}(k)\frac{j_n(k|\boldsymbol{x}|)}{|\boldsymbol{x}|}G_1, \tag{7}$$

$$X_{pq}^{(\phi)}(k,|\boldsymbol{x}|) = -\frac{2\pi q}{k\rho_0 c}H_{pq}\sum_{n=|q|}^{N}H_{nq}\alpha_{nq}(k)\frac{j_n(k|\boldsymbol{x}|)}{|\boldsymbol{x}|}G_2, \tag{8}$$

where ρ_0 is the medium density, $j_p'(k|\boldsymbol{x}|)$ denotes the derivative of $j_p(k|\boldsymbol{x}|)$ with respect to r, and

$$H_{pq} = \frac{(-1)^{\frac{q+|q|}{2}}}{2^{|q|}(|q|)!}\sqrt{\frac{(2p+1)}{4\pi}\frac{(p+|q|)!}{(p-|q|)!}}, \tag{9}$$

$$
\begin{aligned}
G_1 &= (n+|q|+1)\mathcal{G}\left(a, b+\frac{1}{2}-\delta_{n+|q|+1}; -\mu_1-\delta_{n+|q|+1}, -\mu_2; \nu_1+\delta_{n+|q|}, \nu_2; \xi_1, \xi_2\right) \\
&\quad - (n+1)\mathcal{G}\left(a, b+\frac{1}{2}; -\mu_1, -\mu_2; \nu_1, \nu_2; \xi_1, \xi_2\right),
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
G_2 &= \mathcal{G}(a, b; -\mu_1, -\mu_2; \nu_1, \nu_2; \xi_1, \xi_2) \\
&= [(-1)^{2b+1}+1]\sum_{j_1=0}^{\mu_1}\sum_{j_2=0}^{\mu_2}\frac{(-\mu_1)_{j_1}(\nu_1)_{j_1}}{(\xi_1)_{j_1}j_1!}\frac{(-\mu_2)_{j_2}(\nu_2)_{j_2}}{(\xi_2)_{j_2}j_2!}\frac{B(j_1+j_2+a, b)}{2},
\end{aligned}
\tag{11}
$$

with

$$\delta_M = \begin{cases} 1, & \text{if } M \text{ is even} \\ 0, & \text{if } M \text{ is odd}, \end{cases} \tag{12}$$

$$(a)_j = \begin{cases} 1, & \text{if } j = 0 \\ a(a+1)\dots(a+j-1), & \text{if } j = 1, 2, \dots, \end{cases} \tag{13}$$

and $a = (2|q|+1)/2$, $b = (3-\delta_{n+|q|}-\delta_{p+|q|})/2$, $\mu_1 = -(1-n+|q|-\delta_{n+|q|})/2$, $\mu_2 = -(1-p+|q|-\delta_{p+|q|})/2$, $\nu_1 = (2+n+|q|-\delta_{n+|q|})/2$, $\nu_2 = (2+p+|q|-\delta_{p+|q|})/2$, $\xi_1 = \xi_2 = |q|+1$, and $B(\cdot)$ denotes the beta function. The detailed derivations of (6)–(8) can also be found in [44].

Note that for a sound field the truncation order of the velocity field (i.e., Q_Ψ) is usually larger than or equal to the truncation order of the pressure field (i.e., N) due to the truncation theorem of the velocity expressions [44]. The velocity coefficients represent the continuous particle velocity distribution of a sound field over space and they are directly determined by the pressure coefficients of the sound field, as well as the radial functions (i.e., the spherical Bessel functions). All other terms in the expressions of the velocity coefficients are either constants or indices-only dependent variables. In other words, we can obtain the continuous particle velocity distribution of a sound field over space directly from the pressure coefficients describing the sound field.

3.2. Mixed-Source Model

Different from the plane wave model [20], which represents a sound field with an equivalent superposition of virtual plane wave sources, the mixed-source model expands the sound field in terms of a mixture of near-field and far-field virtual sources [39]. This mixed-source model can overcome the difficulties of the plane wave expansion in synthesizing the near-field sources.

To blend the model of a near- and far-field virtual source together, we define a normalized point-source to be the mixed-source, defined as

$$P(\boldsymbol{x}, k; \boldsymbol{y}) = |\boldsymbol{y}|e^{-ik|\boldsymbol{y}|}\frac{e^{ik|\boldsymbol{y}-\boldsymbol{x}|}}{4\pi|\boldsymbol{y}-\boldsymbol{x}|}, \tag{14}$$

where $\boldsymbol{y}$ denotes the position of the virtual mixed source. In the spherical harmonic domain, it can be expressed as

$$P(\boldsymbol{x}, k; \boldsymbol{y}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\underbrace{ik|\boldsymbol{y}|e^{-ik|\boldsymbol{y}|}h_n(k|\boldsymbol{y}|)Y_{nm}^*(\hat{\boldsymbol{y}})}_{\beta_{nm}(k)}j_n(k|\boldsymbol{x}|)Y_{nm}(\hat{\boldsymbol{x}}), \tag{15}$$

where $h_n(\cdot)$ is the spherical Hankel function of the first kind, and $\beta_{nm}(k)$ are the pressure coefficients of the mixed source. We note that the mixed source has the property of [26]

$$\lim_{|\boldsymbol{y}| \to +\infty} P(\boldsymbol{x}, k; \boldsymbol{y}) = \frac{e^{-ik\hat{\boldsymbol{y}} \cdot \boldsymbol{x}}}{4\pi}, \tag{16}$$

which shows that the plane-wave expansion can be modeled using the mixed-source expansion placed in the far field.

We construct a virtual equivalent sound field using two concentric spheres of virtual mixed sources. Therefore, the sound field can be expressed, using the mixed-source model, as [39]

$$\begin{aligned}
P(\boldsymbol{x}, k) = & \int \psi(k, R_N\hat{\boldsymbol{y}}; \boldsymbol{o}) \underbrace{R_N e^{-ikR_N} \frac{e^{ik||R_N\hat{\boldsymbol{y}} - \boldsymbol{x}||}}{4\pi||R_N\hat{\boldsymbol{y}} - \boldsymbol{x}||}}_{P(\boldsymbol{x},k;R_N\hat{\boldsymbol{y}})} d\hat{\boldsymbol{y}} \\
& + \int \psi(k, R_F\hat{\boldsymbol{y}}; \boldsymbol{o}) \underbrace{R_F e^{-ikR_F} \frac{e^{ik||R_F\hat{\boldsymbol{y}} - \boldsymbol{x}||}}{4\pi||R_F\hat{\boldsymbol{y}} - \boldsymbol{x}||}}_{P(\boldsymbol{x},k;R_F\hat{\boldsymbol{y}})} d\hat{\boldsymbol{y}},
\end{aligned} \tag{17}$$

where R_N is the radius of the virtual sphere placed in the near field, R_F is the radius of the virtual sphere placed in the far field, $\psi(k, R_N\hat{\boldsymbol{y}}; \boldsymbol{o})$ and $\psi(k, R_F\hat{\boldsymbol{y}}; \boldsymbol{o})$ are the driving functions of the mixed-source distributions as observed at $\boldsymbol{o}$, and $P(\boldsymbol{x}, k; R_N\hat{\boldsymbol{y}})$ and $P(\boldsymbol{x}, k; R_F\hat{\boldsymbol{y}})$ are the mixed sources on the two spheres. Note that the mixed-source model expands the sound field with respect to $\boldsymbol{o}$, which allows us to observe the source distribution at $\boldsymbol{o}$ and estimate the sound at any translated position $\boldsymbol{x}$.

3.3. Particle-Velocity-Based Expansion

In this subsection, we develop the theory of particle-velocity-based mixed-source expansion. The particle velocity is related to the sound pressure by the Euler's equation [55],

$$\boldsymbol{V}(\boldsymbol{x}, k) = \frac{i}{k\rho_0 c} \vec{\nabla} P(\boldsymbol{x}, k), \tag{18}$$

where $\vec{\nabla} P(\boldsymbol{x}, k)$ denotes the gradient of $P(\boldsymbol{x}, k)$ in terms of r, θ, and ϕ. Similarly, for the mixed source, we have

$$\boldsymbol{V}(\boldsymbol{x}, k; R_X\hat{\boldsymbol{y}}) = \frac{i}{k\rho_0 c} \vec{\nabla} P(\boldsymbol{x}, k; R_X\hat{\boldsymbol{y}}), \tag{19}$$

where $R_X \in \{R_N, R_F\}$; therefore, taking the gradient operation on both sides of (17), we can obtain the particle-velocity-based mixed-source expansion

$$V_\Psi(\boldsymbol{x}, k) = \int \psi(k, R_N\hat{\boldsymbol{y}}; \boldsymbol{o}) V_\Psi(\boldsymbol{x}, k; R_N\hat{\boldsymbol{y}}) d\hat{\boldsymbol{y}} + \int \psi(k, R_F\hat{\boldsymbol{y}}; \boldsymbol{o}) V_\Psi(\boldsymbol{x}, k; R_F\hat{\boldsymbol{y}}) d\hat{\boldsymbol{y}}, \tag{20}$$

where $V_\Psi(\boldsymbol{x}, k; R_N\hat{\boldsymbol{y}})$ and $V_\Psi(\boldsymbol{x}, k; R_F\hat{\boldsymbol{y}})$ represent the particle velocity of the virtual sources in the near field and the far field, respectively, in the Ψ direction.

In practice, the particle-velocity-based mixed-source expansion (20) can be approximated by a finite set of L known virtual sources for each virtual sphere; therefore, (20) can be expressed as

$$V_\Psi(\boldsymbol{x}, k) = \sum_{\ell=1}^{L} w_\ell \psi(k, R_N\hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) V_\Psi(\boldsymbol{x}, k; R_N\hat{\boldsymbol{y}}_\ell) + \sum_{\ell=1}^{L} w_\ell \psi(k, R_F\hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) V_\Psi(\boldsymbol{x}, k; R_F\hat{\boldsymbol{y}}_\ell), \tag{21}$$

where w_ℓ are the the sampling weights for the source distribution, and $\hat{\boldsymbol{y}}_\ell$ denotes the position of the ℓth virtual source on each sphere.

Note that the parameters of the mixed-source expansion need to be carefully selected in the implementations. The near-field radius of the virtual sphere can be selected based on the desired maximum translation distance, because the listener cannot move beyond the near-field sphere. As for the far-field sphere, it can be placed anywhere so long as it is in the far-field. Moreover, the selection of the number of the virtual sources per sphere is a trade-off between angular resolution and computation cost. The more virtual sources there are, the more opportunity they have to match the true sound direction of arrival no matter the higher computation cost the system has.

By replacing $\alpha_{nm}(k)$ with $\beta_{nm}^{(\ell)}$ (i.e., the pressure coefficients of the ℓth virtual source) in (6)–(8), and then substituting them into (5), the particle velocity due to the ℓth virtual source can be written as

$$V_{\Psi}(\boldsymbol{x}, k; R_X \hat{\boldsymbol{y}}_{\ell}) = \sum_{p=0}^{Q_{\Psi}} \sum_{q=-p}^{p} X_{pq}^{(\Psi,\ell)}(k, |\boldsymbol{x}|; R_X \hat{\boldsymbol{y}}_{\ell}) Y_{pq}(\hat{\boldsymbol{x}}), \tag{22}$$

where $X_{pq}^{(\Psi,\ell)}(k, |\boldsymbol{x}|)$ are the velocity coefficients of the ℓth virtual source. Substituting (5) and (22) into (21), (21) can be simplified as

$$X_{pq}^{(\Psi)}(k, |\boldsymbol{x}|) = \sum_{\ell=1}^{2L} \omega_{\ell} \psi(k, R_{\ell} \hat{\boldsymbol{y}}_{\ell}; \boldsymbol{o}) X_{pq}^{(\Psi,\ell)}(k, |\boldsymbol{x}|; R_{\ell} \hat{\boldsymbol{y}}_{\ell}), \tag{23}$$

where $R_{\ell} = R_N$ for $\ell \in [1, L]$, and for $\ell \in [L+1, 2L]$ we have $R_{\ell} = R_F$, and $w_{\ell} = w_{\ell-L}$.

We denote (23), in matrix form, as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\psi}, \tag{24}$$

where $\boldsymbol{X} = [\boldsymbol{X}_r^T, \boldsymbol{X}_\theta^T, \boldsymbol{X}_\phi^T]^T$ is a $[(Q_r+1)^2 + (Q_\theta+1)^2 + (Q_\phi+1)^2]$ long vector containing all three components of the particle velocity vector with $\boldsymbol{X}_\Psi = [X_{00}^{(\Psi)}, X_{1(-1)}^{(\Psi)}, \ldots, X_{Q_\Psi Q_\Psi}^{(\Psi)}]^T$, $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_{\mathcal{L}}]^T$ is a $\mathcal{L} = 2L$ long vector with $\psi_{\mathcal{L}} = \psi(k, R_{\mathcal{L}} \hat{\boldsymbol{y}}_{\mathcal{L}}; \boldsymbol{o})$, and $\boldsymbol{A} = [\boldsymbol{A}_r^T, \boldsymbol{A}_\theta^T, \boldsymbol{A}_\phi^T]^T$ is the $[(Q_r+1)^2 + (Q_\theta+1)^2 + (Q_\phi+1)^2]$ by $\mathcal{L}$ expansion matrix with

$$\boldsymbol{A}_\Psi = \begin{bmatrix} w_1 X_{00}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{00}^{(\Psi,\mathcal{L})} \\ w_1 X_{1(-1)}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{1(-1)}^{(\Psi,\mathcal{L})} \\ \vdots & \ddots & \vdots \\ w_1 X_{Q_\Psi Q_\Psi}^{(\Psi,1)} & \cdots & w_{\mathcal{L}} X_{Q_\Psi Q_\Psi}^{(\Psi,\mathcal{L})} \end{bmatrix}. \tag{25}$$

Note that the dependence is omitted here for notational simplicity. The expansion is now reduced to calculate the driving function of the mixed source distribution $\psi(k, \boldsymbol{y}; \boldsymbol{o})$ for a given measured sound field. We introduce two solutions here.

3.3.1. Least Squares Solution

The least squares method is to estimate the driving function $\boldsymbol{\psi}$ so as to minimize the difference between $\boldsymbol{X}$ and $\boldsymbol{A}\boldsymbol{\psi}$, and it can be formulated as

$$\boldsymbol{\psi} = \arg \min_{\boldsymbol{\psi}} ||\boldsymbol{A}\boldsymbol{\psi} - \boldsymbol{X}||_2^2. \tag{26}$$

This problem can be solved using a Moore–Penrose inverse with Tikhonov regularization [56].

3.3.2. Sparse Solution

It is shown that modeling fewer virtual sources from propagation directions that are similar to the original sound will lead to better perceptual immersion [39]. Therefore, we also construct a sparse constrained objective function, expressed as

$$\psi = \arg\min_{\psi} ||A\psi - X||_2^2 + \lambda||\psi||_1, \tag{27}$$

where λ is the parameter controlling the strength of the sparsity constraint for the driving function ψ. This sparse linear regression problem can be solved using the least absolute shrinkage and selection operator (LASSO) [57]. Other optimization techniques such as the iteratively reweighted least squares (IRLS) [58] can also be applied to solving this problem. We direct readers to [59] for more details about compressive sensing. Note that the applicability of the above solutions will depend on the condition number of matrix A, which is determined by the distribution of the virtual sources. It shows that a well-conditioned A matrix should result from a geometry in which the virtual sources are maximally distributed over space in the literature [26].

4. Sound Field Translation and Synthesis for Binaural Reproduction

As shown in the flowchart of the velocity-based mixed-source sound field translation for binaural reproduction in Figure 1, in this section, we translate the sound field in the spherical harmonic domain, and then synthesize the left and right ear signals at the translated position.
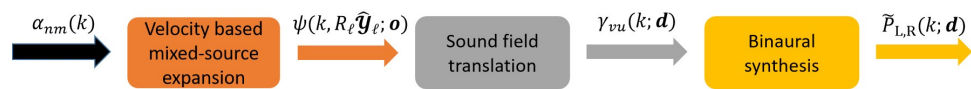

Figure 1. The flowchart of the velocity-based mixed-source sound field translation for binaural reproduction.

Solving (26) or (27) gives us the reproduced virtual-reality sound field, expressed as

$$\widetilde{P}(\boldsymbol{x}, k) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) R_\ell e^{-ikR_\ell} \frac{e^{ik||R_\ell \hat{\boldsymbol{y}}_\ell - \boldsymbol{x}||}}{4\pi||R_\ell \hat{\boldsymbol{y}}_\ell - \boldsymbol{x}||}. \tag{28}$$

With respect to the translated position $\boldsymbol{d}$, the virtual sources are located at $\boldsymbol{z}_\ell = R_\ell \hat{\boldsymbol{y}}_\ell - \boldsymbol{d}'$; therefore, the translated sound field can be written as

$$\widetilde{P}(\boldsymbol{x}, k; \boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) R_\ell e^{-ikR_\ell} \frac{e^{ik||\boldsymbol{z}_\ell - \boldsymbol{x}||}}{4\pi||\boldsymbol{z}_\ell - \boldsymbol{x}||}. \tag{29}$$

We decompose (29) in the spherical harmonic domain about the position $\boldsymbol{d}$, and the pressure coefficients of the translated sound field due to the virtual sources at $\boldsymbol{z}_\ell$ can expressed as

$$\gamma_{vu}(k; \boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) ikR_\ell e^{-ikR_\ell} h_v(k|\boldsymbol{z}_\ell|) Y_{vu}^*(\hat{\boldsymbol{z}}_\ell), \tag{30}$$

where v and u index the translated harmonics, centered at $\boldsymbol{d}$. The left and right ear signals can be synthesized, in the spherical harmonic domain, directly from the translated pressure coefficients $\gamma_{vu}(k; \boldsymbol{d})$ and the coefficients of the listener's HRTFs, $H_{\text{L,R}}^{vu}(k)$, with [60,61]

$$\widetilde{P}_{\text{L,R}}(k; \boldsymbol{d}) = \sum_{v=0}^{\infty} \sum_{u=-v}^{v} \gamma_{vu}(k; \boldsymbol{d}) H_{\text{L,R}}^{vu}(k). \tag{31}$$

Note that $H_{\text{L,R}}^{vu}(k)$ is obtained from the spherical harmonic decomposition of the HRTF $H_{\text{L,R}}(k, \boldsymbol{y}; \boldsymbol{d})$.

Alternatively, the left and right ear signals can be obtained by applying the mixed-source driving functions to the HRTFs based on the listener's translated position, given as

$$\widetilde{P}_{\text{L,R}}(k;\boldsymbol{d}) = \sum_{\ell=1}^{2L} \omega_\ell \psi(k, R_\ell \hat{\boldsymbol{y}}_\ell; \boldsymbol{o}) H_{\text{L,R}}(k, \boldsymbol{y}_\ell; \boldsymbol{d}). \tag{32}$$

We also direct readers to [62,63] for other options of rendering the ear signals at the translated positions.

5. Simulation Study

In this section, we evaluate the performance of the proposed particle-velocity-based mixed-source expansion for binaural reproduction by comparing it with the state-of-the-art method, the sound-pressure-based mixed-source expansion, in simulated acoustic environments. The simulated environments and the evaluation criteria are explained in Section 5.1. The simulation results and discussions are given in Section 5.2 and Section 5.3, respectively.

5.1. Simulation Setup and Criteria

In this simulation, we firstly consider an acoustic environment that contains a single free-field point source with a location of $\boldsymbol{b}_1 = (1, \pi/3, \pi/4)$. We then consider another scenario where multiple sources exist, which is constructed by adding two more point sources with randomly selected locations of $\boldsymbol{b}_2 = (0.5, 3\pi/4, 4\pi/3)$ and $\boldsymbol{b}_3 = (1.8, \pi/2, \pi/2)$, respectively, in the former acoustic environment. The true sound pressure at an arbitrary position $\boldsymbol{x}$ due to the jth point source is

$$P_j(\boldsymbol{x}, k) = \frac{e^{ik||\boldsymbol{b}_j - \boldsymbol{x}||}}{4\pi||\boldsymbol{b}_j - \boldsymbol{x}||}, \quad j = \{1, 2, 3\}. \tag{33}$$

The total sound pressure due to multiple sources follows the principle of linear superposition. We assume that the receiver can record the sound field up to the 4th order (i.e., $N = 4$), and the radius of the receiver region is $R = 0.042$ m (We simulate the radius of the EigenMike, which is a popular commercial microphone array [53]). The pressure coefficients describing the acoustic environment $\{\alpha_{nm}(k)\}$ can be extracted from the receiver's recording. Here, for simplicity, we use the theoretical pressure coefficients up to the Nth order, expressed as

$$\alpha_{nm}(k) = \begin{cases} ikh_n(k|\boldsymbol{b}_1|)Y_{nm}^*(\hat{\boldsymbol{b}}_1), \ n \in [0, N], \text{ for S1,} \\\\ \sum_{j=1}^{3} ikh_n(k|\boldsymbol{b}_j|)Y_{nm}^*(\hat{\boldsymbol{b}}_j), \ n \in [0, N], \text{ for S2,} \end{cases} \tag{34}$$

where S1 denotes the single source scenario and S2 denotes the multiple sources scenario. The mixed-source model consists of two sets of $L = 36$ virtual sources distributed on the two spheres, both with the positions arranged on Fliege nodes [64]. The first set is placed in the near field at $R_N = 2$ m, and the second set is placed in far field at $R_F = 20$ m.

Throughout the simulations, the frequency of the sources is $f = 1500$ Hz, except the evaluation in terms of the frequency. The sound speed is $c = 343$ m/s and the air density is $\rho_0 = 1.29$ kg/m^3. The truncation orders of the velocity coefficients in the r, θ, ϕ direction are $Q_r = N$, $Q_\theta = 2N$, and $Q_\phi = 2N$, respectively, according to the truncation theorem of the velocity expressions [44]. For the sparse solution, we use a 500 iterations LASSO.

As the first objective performance measure for the reproduction method, we define the mismatch between the reproduced sound field and the true sound field at any point $\boldsymbol{x}$ as

$$\epsilon(\boldsymbol{x}, k) = \frac{|P(\boldsymbol{x}, k) - \widetilde{P}(\boldsymbol{x}, k)|^2}{|P(\boldsymbol{x}, k)|^2} \times 100(\%). \tag{35}$$

A value of 0% means the reproduced sound field is exactly the same as the true sound field at the point. The larger the percentage, the greater the difference between the reproduced sound field and the true sound field at the point. A good perception of sound localization requires accurate reproduction of sound direction. To evaluate the performance of the reproduction of sound direction at the translated position d, two localization metrics are introduced [22]. The first one is related to the velocity vector at the translated position, which is given by

$$\boldsymbol{r}_V(k; \boldsymbol{d}) = \text{Re}\left(\frac{\sum_{\ell=1}^{\mathcal{L}} \mathcal{U}_\ell(k)\hat{\boldsymbol{z}}_\ell}{\sum_{\ell=1}^{\mathcal{L}} \mathcal{U}_\ell(k)}\right), \tag{36}$$

where the subscript V denotes velocity, $\text{Re}(\cdot)$ denotes the real part, and $\mathcal{U}_\ell(k)$ are the effective source gains (accounting for point-source radiation), expressed as

$$\mathcal{U}_\ell(k) = \frac{e^{ik(|\boldsymbol{z}_\ell|-R_\ell)}}{|\boldsymbol{z}_\ell|/R_\ell}\psi(k, R_\ell\hat{\boldsymbol{y}}_\ell; \boldsymbol{o}). \tag{37}$$

$\boldsymbol{r}_V(k)$ is used to predict localization at low frequencies (below 700 Hz). For high frequencies (above 500 Hz), the second metric related to the intensity vector is given by

$$\boldsymbol{r}_E(k; \boldsymbol{d}) = \frac{\sum_{\ell=1}^{\mathcal{L}} |\mathcal{U}_\ell(k)|^2\hat{\boldsymbol{z}}_\ell}{\sum_{\ell=1}^{\mathcal{L}} |\mathcal{U}_\ell(k)|^2}, \tag{38}$$

where the subscript E denotes energy. The directions of $\boldsymbol{r}_V(k; \boldsymbol{d})$ and $\boldsymbol{r}_E(k; \boldsymbol{d})$ indicate the expected localization direction, and their magnitudes, $|\boldsymbol{r}_V(k; \boldsymbol{d})|$ and $|\boldsymbol{r}_E(k; \boldsymbol{d})|$, indicate the quality of localization [22]. For the point source at $\boldsymbol{b}$, the directional error $\zeta_{V,E}(k; \boldsymbol{d})$ at the translated position $\boldsymbol{d}$ is defined as (The direction error is related to the direction difference in angles between the two unit vectors in (39) by $\arccos(1 - \zeta^2/2)$.)

$$\zeta_{V,E}(k; \boldsymbol{d}) = \left|\frac{\boldsymbol{r}_{V,E}(k; \boldsymbol{d})}{|\boldsymbol{r}_{V,E}(k; \boldsymbol{d})|} - \frac{\boldsymbol{b} - \boldsymbol{d}}{|\boldsymbol{b} - \boldsymbol{d}|}\right|. \tag{39}$$

Ideally, $\boldsymbol{r}_V(k; \boldsymbol{d})$ and $\boldsymbol{r}_E(k; \boldsymbol{d})$ should be unit vectors that point to the direction of the point source $\boldsymbol{b}$ (i.e., $|\boldsymbol{r}_V(k; \boldsymbol{d})| = |\boldsymbol{r}_E(k; \boldsymbol{d})| = 1$, and $\zeta_{V,E}(k; \boldsymbol{d}) = 0$).

We also implement the two corresponding solutions (i.e., the closed-form solution and the sparse solution) of the state-of-the-art technique, the sound-pressure-based mixed-source expansion [39,40], for comparison.

5.2. Simulation Results

In this section, we first examine the accuracy of the proposed particle-velocity-based mixed-source expansion to reconstruct the pressure field in the two scenarios, and then we evaluate the localization performance in terms of the frequency and translated position using the metrics in Section 5.1 for the single source scenario.

5.2.1. Reproduced Pressure Field

In terms of the single source scenario, Figure 2 shows the true pressure field calculated by (33), and the recorded pressure field calculated using the theoretical pressure coefficients in (34). The black circle denotes the receiver region. The receiver can record the sound field within the receiver region. However, truncation error degrades the accuracy of the pressure field beyond the receiver region, which is the constraint in Section 2 that we wish to relax through the proposed method. To compare the proposed method with the sound-pressure-based expansion, we first show the results of the sound-pressure-based expansion for the closed-form solution and the sparse solution in Figure 3. We observe that the closed-form solution can guarantee the reproduction accuracy within the receiver region. However, for a position away from the receiver region, the mismatch percentage at the position quickly increases to a value close to 100 %, which is similar (but not exact)

to the recorded pressure field and is not good enough for translation. The sparse solution improves the performance and reconstructs the pressure field accurately around the receiver region, which allows for the listener to move to the position exterior to the receiver region. Figure 4 shows the reconstructed pressure field and the error field controlled by the least squares solution and the sparse solution using the particle-velocity-based expansion, and these two solutions have the exactly same pressure field mismatch, which means that the least squares solution of the particle-velocity-based expansion has the same performance as the closed-form solution of the sound-pressure-based expansion. However, for the sparse solution of the particle-velocity-based expansion, it provides a larger area with 0 % pressure field mismatch around the receiver region compared with the sparse solution of the sound-pressure-based expansion, which allows for the listener to translate further away from the original receiver region.
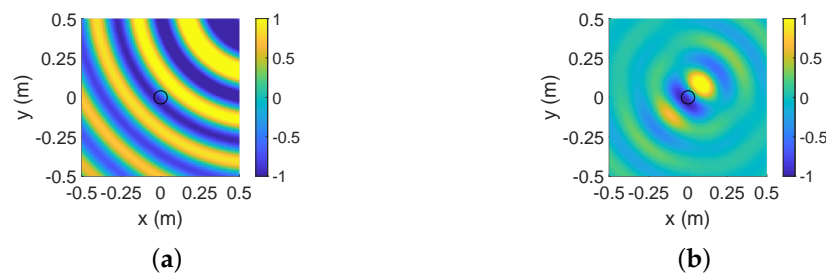

(a)

(b)

Figure 2. The true pressure field and recorded pressure field at 1500 Hz. The black circle denotes the receiver region. (**a**) the true pressure field; (**b**) the recorded pressure field.
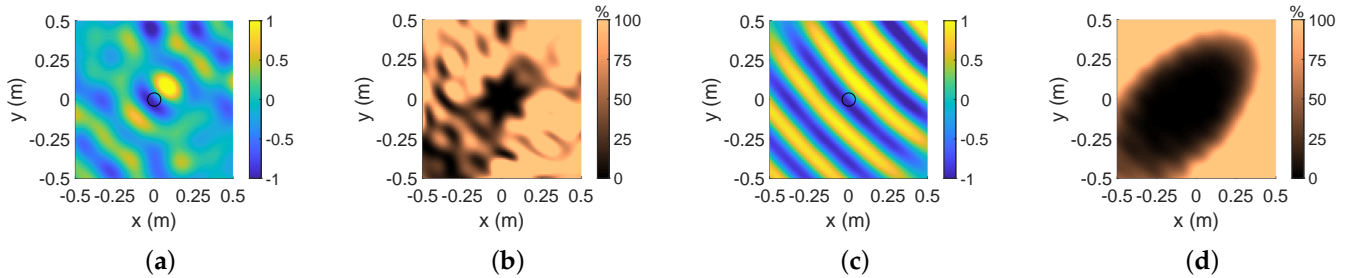

(a)

(b)

(c)

(d)

Figure 3. The reconstructed pressure field and the error field controlled by the closed-form solution (P-CFS) and the sparse solution (P-SS) using the sound-pressure-based expansion. (**a**) Reconstructed pressure field controlled by P-CFS; (**b**) error field controlled by P-CFS; (**c**) reconstructed pressure field controlled by P-SS; (**d**) error field controlled by P-SS.
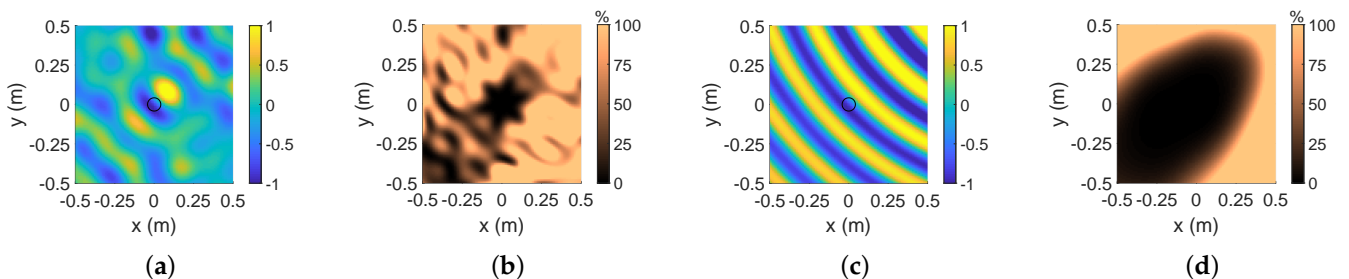

(a)

(b)

(c)

(d)

Figure 4. The reconstructed pressure field and the error field controlled by the least squares solution (V-LSS) and the sparse solution (V-SS) using the particle-velocity-based expansion. (**a**) Reconstructed pressure field controlled by V-LSS; (**b**) error field controlled by V-LSS; (**c**) reconstructed pressure field controlled by V-SS; (**d**) error field controlled by V-SS.

As for the multiple sources scenario, the true pressure field and recorded pressure field are given in Figure 5. The pressure field difference between them also increases gradually (i.e., the value of the mismatch gradually increases from 0% to 100%) as the

observation point moves away from the target region due to the truncation error. Similar to the evaluation for the single source scenario, we first show the reconstructed pressure fields controlled by the closed-form solution and the sparse solution based on the sound-pressure-based expansion in the multiple sources scenario. The results with the error fields are shown in Figure 6. It shows that the closed-form solution still can only guarantee accurate reproduction around the target region, whereas the sparse solution can provide greater translation area. However, the area with 0% pressure field mismatch is much smaller than that in the single source scenario. Figure 7 shows the results of the least squares solution and the sparse solution based on the particle-velocity-based expansion. Similarly, the result of the least squares solution under the particle-velocity-based expansion is identical to the result of the closed-form solution under the sound-pressure-based expansion in terms of the pressure field mismatch. Although the sparse solution of particle-velocity-based expansion in the multiple sources scenario can provide a smaller translation area than that in the single source scenario, it still has an improved performance compared to the sparse solution of the sound-pressure-based expansion in the multiple sources scenario.

Figure 5. The true pressure field and recorded pressure field at 1500 Hz in the multiple sources scenario. The black circle denotes the receiver region. (**a**) the true pressure field; (**b**) the recorded pressure field.
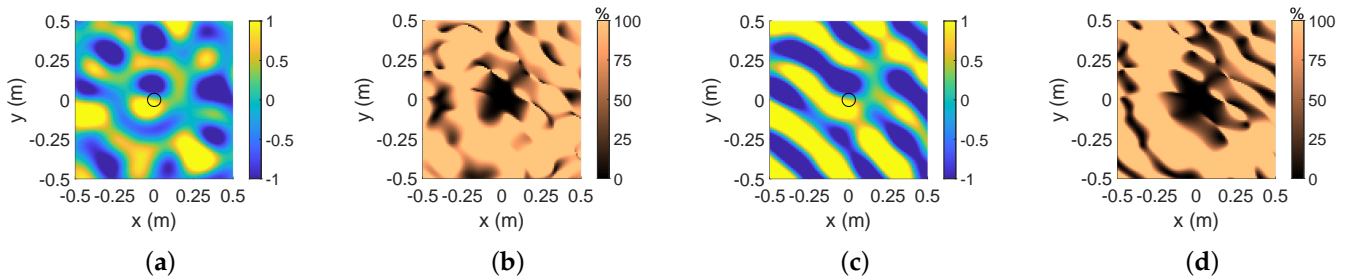

Figure 6. The reconstructed pressure field and the error field controlled by P-CFS and P-SS using the sound-pressure-based expansion in the multiple sources scenario. (**a**) Reconstructed pressure field controlled by P-CFS; (**b**) error field controlled by P-CFS; (**c**) reconstructed pressure field controlled by P-SS; (**d**) error field controlled by P-SS.
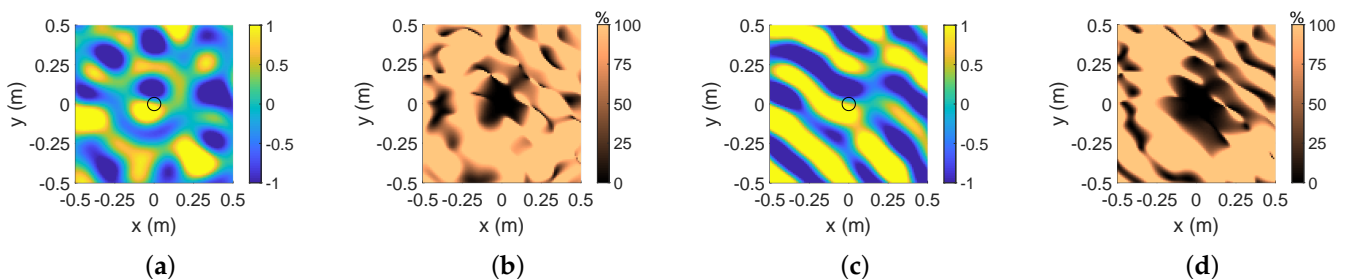

Figure 7. The reconstructed pressure field and the error field controlled by V-LSS and V-SS using the particle-velocity-based expansion in the multiple sources scenario. (**a**) Reconstructed pressure field controlled by V-LSS; (**b**) error field controlled by V-LSS; (**c**) reconstructed pressure field controlled by V-SS; (**d**) error field controlled by V-SS.

5.2.2. Frequency

To evaluate the broadband performance of the proposed method on direction reproduction at a translated position, we calculate the directional error $\zeta_{V,E}$ and the magnitude of $r_{V,E}$ (i.e., $|r_{V,E}|$) using the two localization metrics introduced in Section 5.1, with respect to the change in frequency. The translated position is $(0.2, 0.3, 0)$ m in Cartesian coordinates. Figure 8 shows the results of ζ_V and $|r_V|$. As expected, the sparse solution of the particle-velocity-based expansion is better than all the other solutions, especially for low frequencies. The results of ζ_E and $|r_E|$ are given in Figure 9. We observe that the sparse solution of the particle-velocity-based expansion has the least directional error, and the $|r_E|$ for this solution is closest to 1 among all the evaluated methods. Compared to curves for the other method, the curve for the sparse solution of the particle-velocity-based expansion has little fluctuation in terms of frequency, which means the sparse solution of the particle-velocity-based expansion can provide a uniform performance for broadband signals. In addition, the least squares solution of the particle-velocity-based expansion is the same as the closed-form solution of the sound-pressure-based expansion for the performance in terms of frequency.
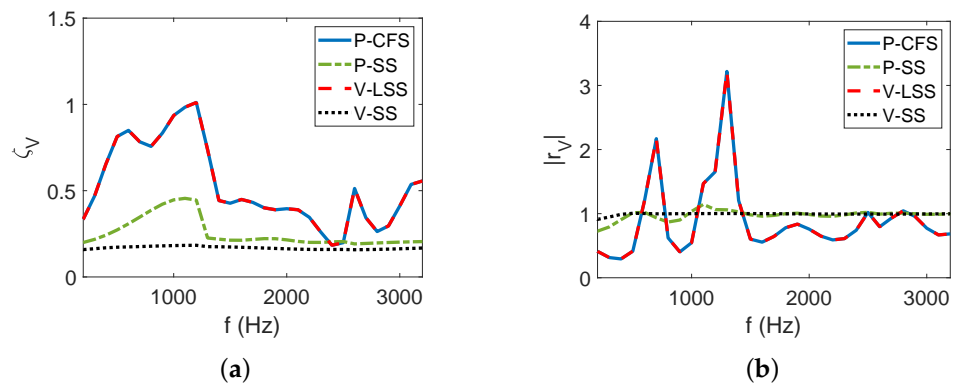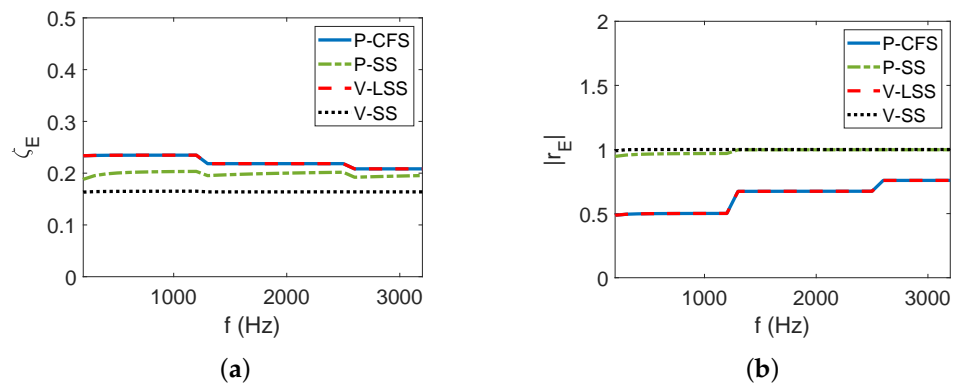

Figure 8. The directional error ζ_V and the magnitude of r_V with respect to frequency at the translated position of $(0.2, 0.3, 0)$ m. (**a**) Directional error ζ_V; (**b**) magnitude of r_V.

Figure 9. The directional error ζ_E and the magnitude of r_E with respect to frequency at the translated position of $(0.2, 0.3, 0)$ m. (**a**) Directional error ζ_E; (**b**) magnitude of r_E.

5.2.3. Translated Position

We also evaluate the performance of each method when the translated position moves away from the receiver region. Figures 10 and 11 show the results of ζ_V and $|r_V|$ at 300 Hz, and the results of ζ_E and $|r_E|$ at 1500 Hz, respectively, as a function of translation distance along the positive y-axis. Note that we plot Figures 10 and 11 at different frequencies because r_V is a good predictor for low frequencies, whereas r_E is a good predictor for high frequencies, as mentioned in Section 5.1. From Figures 10a and 11a, we show that all the methods have little error at the origin where the receiver is positioned; however,

the performance worsens as the translation distance increases for all the methods. Compared to the sparse solution of the sound-pressure-based expansion, the sparse solution of the particle-velocity-based expansion has less directional error and this advantage becomes more significant for a further translated position, which leads to a better direction reproduction, especially for the translated position far away from the receiver region. The results of Figures 10b and 11b also indicate that the sparse solution of the particle-velocity-based expansion can provide a better quality of localization for the listener than the sparse solution of the sound-pressure-based expansion. Once again, we notice that the least squares solution and the closed-form solution have same performance, and both are worse than the the sparse solutions of the particle-velocity-based expansion and the sound-pressure-based expansion.

From the above analysis, we conclude that the least squares solution of the particle-velocity-based expansion is identical to the closed-form solution of the sound-pressure-based expansion, whereas the sparse solution of the particle-velocity-based expansion has better overall performance than the sparse solution of the sound-pressure-based expansion.
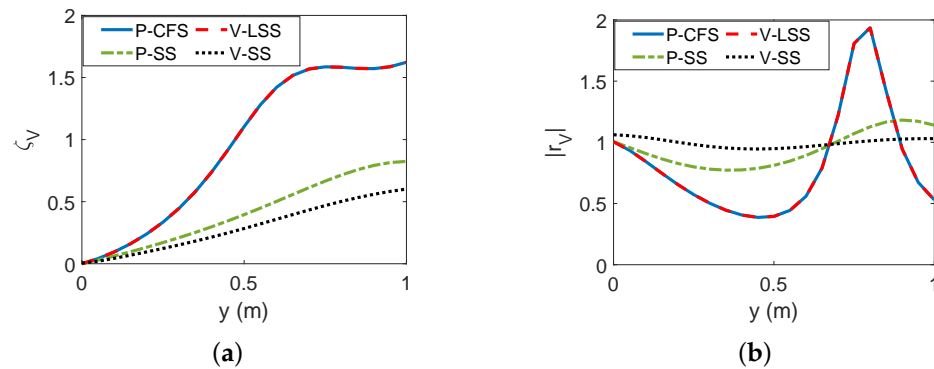

Figure 10. The directional error ζ_V and the magnitude of $\boldsymbol{r}_V$ as a function of translation distance along the positive y-axis. The frequency of the source is 300 Hz. (**a**) Directional error ζ_V; (**b**) magnitude of $\boldsymbol{r}_V$.

Figure 11. The directional error ζ_E and the magnitude of $\boldsymbol{r}_E$ as a function of translation distance along the positive y-axis. The frequency of the source is 1500 Hz. (**a**) Directional error ζ_E; (**b**) magnitude of $\boldsymbol{r}_E$.
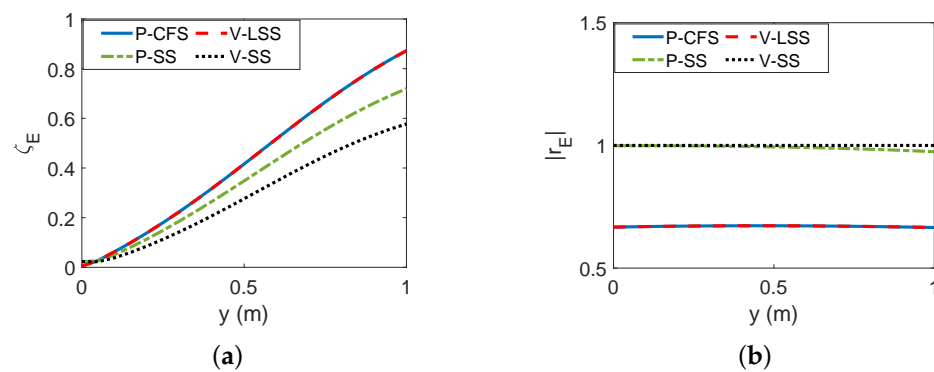
5.3. Discussion

Based on the above simulation results, we give the following comments:

- The particle-velocity-based expansion is derived directly from the sound-pressure-based expansion by the gradient calculation. Therefore, the least squares solution of the particle-velocity-based expansion is mathematically parallel to the closed-form solution of the sound-pressure-based expansion, both of which aim to reconstruct the original truncated recording by distributing energy throughout all virtual sources and

inherit the spatial artifacts caused by the truncated measurement. This leads to poor reproduction outside the receiver region.

- Sparse solutions can improve the performance of reproduction outside the receiver region due to the fact that most sound fields can be reproduced accurately by a single or a few virtual sources, and thus exhibit sparsity in space. Therefore, provided that the desired sound is sparse in space, the sparse solutions can be used to relax the restriction we mentioned in Section 2. We should note that the region of accurate translation in the multiple sources scenario becomes smaller than that in the single source scenario for the sparse solution. In addition, the sparse solution is less applicable to highly reverberant fields where the sparsity does not hold.

- Particle velocity contains the direction information of a sound field. For the sparse solutions, we can achieve more accurate sound field reproduction by controlling particle velocity than sound pressure. In other words, the particle-velocity-based solution outperforms the sound-pressure-based solution for sound field reproduction in the cases where there are a limited number of sources. Furthermore, particle velocity reflects the interaural time difference (ITD); therefore, the velocity vector is directly related to the localization predictor for human perception at low frequencies. Particle velocity is also one of the quantities that determine the sound intensity vector (the localization predictor at high frequencies), which reflects another human localization cue of interaural level difference (ILD). Therefore, the sparse solution of particle-velocity-based expansion is expected to provide an enhanced perceptual immersion for the listener.

- Sparse solutions extrapolate the sound field at the translated positions beyond the receiver region according to the sound field within the receiver region by the mixed-source expansion where we use multiple virtual sources to estimate the original source. Therefore, the more the virtual sources are used, the less error the mixed-source expansion has; however, the higher computation cost the system has. In addition, the particle-velocity-based solutions would be more computationally expensive than the sound-pressure-based solutions due to the the extra spherical harmonic decomposition of particle velocity.

We examine the perceptual advantages of the proposed sparse solution of particle-velocity-based expansion by experiments in the next section.

6. Experimental Verification

In this section, we evaluate the perceptual performance of the proposed method against the state-of-the-art method in a listening experiment. The experimental methodology is introduced in detail in Section 6.1, and then we present the statistical results in Section 6.2.

6.1. Experimental Methodology

To generate the binaural test signals, we construct a virtual experimental environment, where there is a 4th order receiver located at the origin and a free-field point source located at $(1,0,0)$ m, with respect to the origin. The receiver region is a spherical region with a radius of 0.042 m, to represent the spatial abilities of common spherical microphone arrays. We assume the microphone array works ideally, and the theoretical pressure coefficients (34) can be extracted from the ideal microphone recording. The parameters of the mixed-source model are the same as those in Section 5.1. We reproduce the sound image for the listener away from the recording area by extrapolation using the proposed method. The listener faces the positive x-axis, and the listener's head is aligned vertically with the point source on the x-y plane. We used the HRTFs of the FABIAN head and torso simulator [65] from the HUTUBS dataset [66,67]. The source signal is a clip of dry instrumental music with 10 s duration and its spectrum energy mainly distributes in the frequency band below 8 kHz. We process the signals at a frame size of 4096 with 50% overlap and a 16 kHz sampling frequency. We carry out a MUSHRA [68] experiment to

compare four translation methods; therefore, for a translated position there are a total of six binaural test signals in this experiment:

- Reference/Hidden reference: The ground truth, which is obtained directly from filtering the theoretical point source signal with the HRTFs.
- Anchor: Signals of the truncated recording at the origin, which are simulated using the pressure coefficients up to order N in (34). No translation is processed, but is filtered by the HRTFs.
- P-CFS: Signals reconstructed using the closed-form solution of the sound-pressure-based expansion and rendered by the HRTFs.
- V-LSS: Signals reconstructed using the least squares solution of the particle-velocity-based expansion and rendered by the HRTFs.
- P-SS: Signals reconstructed using the sparse solution of the sound-pressure-based expansion and rendered by the HRTFs.
- V-SS: Signals reconstructed using the sparse solution of the particle-velocity-based expansion and rendered by the HRTFs.

We generate the above binaural test signals at six different translated positions outside the receiver region. All the translated positions are on the x-y plane, and they are $(0, 1, 0)$ m, $(0, -1, 0)$ m, $(0.5, 0.8, 0)$ m, $(0.5, -0.8, 0)$ m, $(1, 1, 0)$ m, and $(1, -1, 0)$ m, respectively. In total, the scores of 31 subjects (16 females and 15 males) with normal hearing are collected for this listening experiment. Among the participants there are 14 experienced listeners with an audio research background and 17 common listeners without any audio research background. A random set out of the six sets of test signals is selected for each hearing subject. The test signals are played on a computer and the hearing subjects need to wear headphones to listen to the test signals. We provide the participants with Matlab programs containing the test signals with a user interface, and guide them to perform the experiment using their own computer and headphone through online meetings. The experiment includes two tests, and they are the source localization test and the basic audio quality test, respectively. The order in which the test signals are played is also random for each test. For the source localization test, the hearing subjects are asked to score all the test signals against the reference for the perceived direction of the sound source, whereas the hearing subjects need to score on the spectral distortions and other audible processing artifacts with respect to the reference for the basic audio quality test. During the test, the participants can arbitrarily switch between different test signals, and the duration of the experiment is usually within 20 min. The score ranges from 0 to 100. We disqualify the hearing subjects who score the hidden reference below 90, and their results are removed from the final dataset of scores.

6.2. Experimental Results

After removing the scores of the disqualified subjects, we collect the scores of 30 subjects in total for the source localization test, and 29 scores for the basic audio quality test. Box plots showing the results for both tests are given in Figure 12. On each box, the central short red line indicates the median value, and the edges of the box are the 25th and 75th percentile of the scores. The extremes of the whiskers correspond to $q_3 + 1.5(q_3 - q_1)$ and $q_1 - 1.5(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentile of the scores, respectively. If a score goes beyond the extremes of the whiskers, it is considered an outlier, indicated by the symbol + in the plots. The symbol $*$ denotes the mean value, and the v-shaped notches refer to the 95% confidence interval. The endpoints of the notches are calculated by $q_2 - 1.57(q_3 - q_1)/\sqrt{N_s}$ and $q_2 + 1.57(q_3 - q_1)/\sqrt{N_s}$, where q_2 is the median value and N_s is the number of the scores. The intervals can be used to compare the median values. If two intervals do not overlap, their true medians are different with 95% confidence.
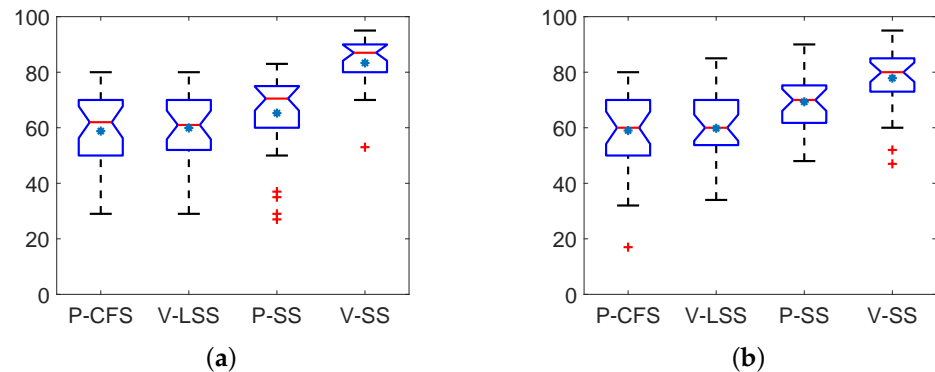
Figure 12. Results of the listening experiment scores for the source localization test and the basic audio quality test. Each box bounds the 25th and 75th percentile of the scores with the central red line indicating the median value, and the whiskers are the outward extension of the 25th and 75th percentile by 1.5 times of the interquartile range. The v-shaped notches represent the 95% confidence intervals. The symbols + denote the outliers and the symbols ∗ denote the mean values. (**a**) Source localization test; (**b**) basic audio quality test.

From Figure 12a, we observe that, for the source localization, the closed-form solution of the sound-pressure-based expansion and the least squares solution of the particle-velocity-based mixed-source solution have similar mean values and median values, and their confidence intervals have nearly 100% overlapping. The sparse solutions of both expansions have higher mean values and median values. However, the sparse solution of the particle-velocity-based expansion has a more significant improvement in score and a more centralized score distribution compared to the sparse solution of the sound-pressure-based expansion. Similar results are shown for the basic audio quality test in Figure 12b, where no significant difference is found between the closed-form solution and the least squares solution. For the sparse solutions, the particle-velocity-based expansion is still observed to perform better than the sound-pressure-based expansion; however, both of outperform the closed-form solution and the least squares solution. We also notice that for the sparse solution of the particle-velocity-based expansion the scores in the source localization test are relatively higher than those in the basic audio quality test.

According to the above experimental results, we conclude that in terms of the perceptual criteria of source localization and audio quality the sparse solutions have better performance than the closed-formed solution and the least squares solution, both of which have similar performance. In the sparse solutions, the particle-velocity-based expansion shows an improvement against the sound-pressure-based expansion, especially for the source localization.

6.3. Discussion

Based on the above experimental results, we give the following comments:

- Statistically, the least squares solution of the particle-velocity-based expansion is equivalent to the closed-form solution of the sound-pressure-based expansion, whereas the sparse solution of the the particle-velocity-based expansion provides better perceptual performance than the sparse solution of the sound-pressure-based expansion at a random translated position. The experimental results are consistent with the discussions in Section 5.3.
- Having particle velocity as the design criterion contributes to more precise reconstruction of the interaural phase difference than sound pressure, which is of utmost important for direction perception and therefore, contributes to improved localization of reconstructed sound when there are only a limited number of sources (the sparse solution). In addition to direction perception, audio quality can also be improved by

controlling particle velocity for the sparse solution, which may be due to the more accurate reproduction of the sound field.

- There are some limitations for the experimental verification. We only evaluate the single source scenario in the perceptual experiment, where the sound field exhibits sparsity. The perceptual performance may degrade for the multiple source scenario or the highly reverberant sound fields. In addition, the source signal is a single type of music, and the performance of the proposed method in terms of the type of audio remains unexplored. We should note that the equipment used (e.g., headphone and sound card) and the listening environment may differ for different subjects due to the remote experiment.

7. Conclusions

We have proposed a new mixed-source expansion, which exploits particle velocity to enhance the performance of sound field translation for binaural reproduction. We describe spatial particle velocity using the velocity coefficients in the spherical harmonic domain, and expand the sound field in terms of the velocity coefficients. We compare the proposed method with the state-of-the-art expansion based on sound pressure. The simulation results show that the least squares solution of the particle-velocity-based expansion is parallel to the closed-form solution of the sound-pressure-based expansion, whereas the sparse solution of the particle-velocity-based expansion can reproduce the sound field over a larger area with less error, and has more accurate direction reproduction at the translated positions than the sparse solution of the sound-pressure-based expansion. Finally, the results of the MUSHRA experiment verify that the particle-velocity-based expansion outperforms the sound-pressure-based expansion for the sparse solution at the translated positions in terms of both source localization and audio quality for the single source scenario. The improvement for source localization is greater than audio quality.

Author Contributions: Conceptualization, H.Z., L.I.B., P.N.S., T.D.A. and V.T.; methodology, H.Z., P.N.S. and T.D.A.; software, H.Z. and L.I.B.; validation, H.Z.; formal analysis, H.Z., P.N.S. and T.D.A.; investigation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, L.I.B., P.N.S., T.D.A. and V.T.; supervision, P.N.S. and T.D.A.; project administration, T.D.A.; funding acquisition, T.D.A. and P.N.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The ethical aspects of this research have been approved by the Australian National University Human Research Ethics Committee (Protocol 2019/767).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR/VR	Augmented Reality/Virtual Reality
MUSHRA	Multiple Stimulus with Hidden Reference and Anchor
HRTF	Head-related transfer function
LASSO	Least absolute shrinkage and selection operator
IRLS	Iteratively reweighted least squares
ITD	Interaural time difference
ILD	Interaural level difference

References

1. Rafaely, B.; Tourbabin, V.; Habets, E.; Ben-Hur, Z.; Lee, H.; Gamper, H.; Arbel, L.; Birnie, L.; Abhayapala, T.; Samarasinghe, P. Spatial audio signal processing for binaural reproduction of recorded acoustic scenes—Review and challenges. *Acta Acust.* **2022**, *6*, 47. [CrossRef]
2. Tylka, J.G.; Choueiri, E.Y. Models for evaluating navigational techniques for higher-order ambisonics. *Proc. Meet. Acoust.* **2017**, *30*, 050009.
3. Tylka, J.G.; Choueiri, E.Y. Fundamentals of a parametric method for virtual navigation within an array of ambisonics microphones. *J. Audio Eng. Soc.* **2020**, *68*, 120–137. [CrossRef]
4. Tylka, J.G.; Choueiri, E.Y. Performance of linear extrapolation methods for virtual sound field navigation. *J. Audio Eng. Soc.* **2020**, *68*, 138–156. [CrossRef]
5. Mariette, N.; Katz, B. Sounddelta—Large scale, multi-user audio augmented reality. In Proceedings of the EAA Symposium on Auralization, Espoo, Finland, 15–17 June 2009; pp. 15–17.
6. Southern, A.; Wells, J.; Murphy, D. Rendering walk-through auralisations using wave-based acoustical models. In Proceedings of the 17th European Signal Processing Conference, Glasgow, UK, 24–28 August 2009; pp. 715–719.
7. Mariette, N.; Katz, B.F.; Boussetta, K.; Guillerminet, O. Sounddelta: A study of audio augmented reality using wifi-distributed ambisonic cell rendering. In *Audio Engineering Society Convention 128*; Audio Engineering Society: New York, NY, USA, 2010.
8. Tylka, J.G.; Choueiri, E. Soundfield navigation using an array of higher-order ambisonics microphones. In Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality, Los Angeles, CA, USA, 30 Septemeber–1 October 2016.
9. Müller, K.; Zotter, F. Auralization based on multi-perspective ambisonic room impulse responses. *Acta Acust.* **2020**, *4*, 25. [CrossRef]
10. Samarasinghe, P.; Abhayapala, T.; Poletti, M. Wavefield analysis over large areas using distributed higher order microphones. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2014**, *22*, 647–658. [CrossRef]
11. Patricio, E.; Ruminski, A.; Kuklasinski, A.; Januszkiewicz, L.; Zernicki, T. Toward six degrees of freedom audio recording and playback using multiple ambisonics sound fields. In *Audio Engineering Society Convention 146*; Audio Engineering Society: New York, NY, USA, 2019.
12. Wang, Y.; Chen, K. Translations of spherical harmonics expansion coefficients for a sound field using plane wave expansions. *J. Acoust. Soc. Amer.* **2018**, *143*, 3474–3478. [CrossRef]
13. Thiergart, O.; Galdo, G.D.; Taseska, M.; Habets, E.A. Geometry-based spatial sound acquisition using distributed microphone arrays. *IEEE Trans. Audio Speech Lang. Process.* **2013**, *21*, 2583–2594. [CrossRef]
14. Mccormack, L.; Politis, A.; Mckenzie, T.; Hold, C.; Pulkki, V. Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple ambisonic receivers. *J. Audio Eng. Soc.* **2022**, *70*, 355–372. [CrossRef]
15. Noisternig, M.; Sontacchi, A.; Musil, T.; Holdrich, R. A 3D ambisonic based binaural sound reproduction system. In Proceedings of the 24th International Conference: Multichannel Audio, The New Reality, Banff, AL, Canada, 26–28 June 2003.
16. Menzies, D.; Al-Akaidi, M. Ambisonic synthesis of complex sources. *J. Audio Eng. Soc.* **2007**, *55*, 864–876.
17. Pihlajamaki, T.; Pulkki, V. Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality. *J. Audio Eng. Soc.* **2015**, *63*, 542–551. [CrossRef]
18. Duraiswami, R.; Li, Z.; Zotkin, D.N.; Grassi, E.; Gumerov, N.A. Plane-wave decomposition analysis for spherical microphone arrays. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 16–19 October 2005; pp. 150–153.
19. Menzies, D.; Al-Akaidi, M. Nearfield binaural synthesis and ambisonics. *J. Acoust. Soc. Amer.* **2007**, *121*, 1559–1563. [CrossRef] [PubMed]
20. Schultz, F.; Spors, S. Data-based binaural synthesis including rotational and translatory head-movements. In *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control-Engineering and Perception*; Audio Engineering Society: New York, NY, USA, 2013.
21. Fernandez-Grande, E. Sound field reconstruction using a spherical microphone array. *J. Acoust. Soc. Amer.* **2016**, *139*, 1168–1178. [CrossRef]
22. Tylka, J.G.; Choueiri, E. Comparison of techniques for binaural navigation of higher-order ambisonic soundfields. In *Audio Engineering Society Convention 139*; Audio Engineering Society: New York, NY, USA, 2015.
23. Frank, M. Phantom Sources Using Multiple Loudspeakers in the Horizontal Plane. Ph.D. Thesis, University of Music and Performing Arts, Graz, Austria, 2013.
24. Daniel, J. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format. In Proceedings of the 23rd International Conference: Signal Processing in Audio Recording and Reproduction, Copenhagen, Denmark, 23–25 May 2003.
25. Poletti, M.A. Three-dimensional surround sound systems based on spherical harmonics. *J. Audio Eng. Soc.* **2005**, *53*, 1004–1025.
26. Ward, D.B.; Abhayapala, T.D. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Trans. Speech Audio Process.* **2001**, *9*, 697–707. [CrossRef]
27. Hahn, N.; Spors, S. Modal bandwidth reduction in data-based binaural synthesis including translatory head-movements. In Proceedings of the German Annual Conference on Acoustics (DAGA), Nurnberg, Germany, 16–19 March 2015; pp. 1122–1125.

28. Hahn, N.; Spors, S. Physical properties of modal beamforming in the context of data-based sound reproduction. In *Audio Engineering Society Convention 139*; Audio Engineering Society: New York, NY, USA, 2015.

29. Kuntz, A.; Rabenstein, R. Limitations in the extrapolation of wave fields from circular measurements. In Proceedings of the 15th European Signal Processing Conference, Poznan, Poland, 3–7 September 2007; pp. 2331–2335.

30. Winter, F.; Schultz, F.; Spors, S. Localization properties of data-based binaural synthesis including translatory head-movements. In Proceedings of the Forum Acusticum, Krakow, Poland, 12–14 September 2014.

31. Kowalczyk, K.; Thiergart, O.; Taseska, M.; Galdo, G.D.; Pulkki, V.; Habets, E.A.P. Parametric spatial sound processing: A flexible and efficient solution to sound scene acquisition, modification, and reproduction. *IEEE Signal Proc. Mag.* **2015**, *32*, 31–42. [CrossRef]

32. Laitinen, T.; Pihlajamäki, M.V.a.; Erkut, C.; Pulkki, V. Parametric time-frequency representation of spatial sound in virtual worlds. *ACM Trans. Appl. Percept. (TAP)* **2012**, *9*, 1–20. [CrossRef]

33. Plinge, A.; Schlecht, S.J.; Thiergart, O.; Robotham, T.; Rummukainen, O.; Habets, E.A.P. Six-degrees-of-freedom binaural audio reproduction of first-order ambisonics with distance information. In Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality, Redmond, WA, USA, 20–22 August 2018.

34. Stein, E.; Goodwin, M.M. Ambisonics depth extensions for six degrees of freedom. In Proceedings of the AES International Conference on Headphone Technology, San Francisco, CA, USA, 27–29 August 2019.

35. Blochberger, M.; Zotter, F. Particle-filter tracking of sounds for frequency-independent 3D audio rendering from distributed b-format recordings. *Acta Acust.* **2021**, *5*, 20. [CrossRef]

36. Allen, A.; Kleijn, B. Ambisonics soundfield navigation using directional decomposition and path distance estimation. In Proceedings of the International Conference on Spatial Audio, Graz, Austria, 7–10 September 2017.

37. Kentgens, M.; Behler, A.; Jax, P. Translation of a higher order ambisonics sound scene based on parametric decomposition. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 151–155.

38. Werner, S.; Klein, F.; Neidhardt, A.; Sloma, U.; Schneiderwind, C.; Brandenburg, K. Creation of auditory augmented reality using a position-dynamic binaural synthesis system—Technical components, psychoacoustic needs, and perceptual evaluation. *Appl. Sci.* **2021**, *11*, 1150. [CrossRef]

39. Birnie, L.; Abhayapala, T.; Samarasinghe, P.; Tourbabin, V. Sound field translation methods for binaural reproduction. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 20–23 October 2019; pp. 140–144.

40. Birnie, L.I.; Abhayapala, T.D.; Tourbabin, V.; Samarasinghe, P. Mixed source sound field translation for virtual binaural application with perceptual validation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2021**, *29*, 1188–1203. [CrossRef]

41. Gerzon, M.A. Optimal reproduction matrices for multispeaker stereo. In Proceedings of the 91st Audio Engineering Society Convention, New York, NY, USA, 4–8 October 1991.

42. Buerger, M.; Maas, R.; Löllmann, H.W.; Kellermann, W. Multizone sound field synthesis based on the joint optimization of the sound pressure and particle velocity vector on closed contours. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 18–21 October 2015; pp. 1–5.

43. Buerger, M.; Hofmann, C.; Kellermann, W. Broadband multizone sound rendering by jointly optimizing the sound pressure and particle velocity. *J. Acoust. Soc. Amer.* **2018**, *143*, 1477–1490. [CrossRef] [PubMed]

44. Zuo, H.; Abhayapala, T.D.; Samarasinghe, P.N. Particle velocity assisted three dimensional sound field reproduction using a modal-domain approach. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 2119–2133. [CrossRef]

45. Gerzon, M.A. General metatheory of auditory localisation. In Proceedings of the 92nd Audio Engineering Society Convention, Vienna, Austria, 24–27 March 1992.

46. Wang, S.; Hu, R.; Chen, S.; Wang, X.; Peng, B.; Yang, Y.; Tu, W. Sound physical property matching between non central listening point and central listening point for nhk 22.2 system reproduction. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 436–440.

47. Shin, M.; Fazi, F.M.; Nelson, P.A.; Seo, J. Control of velocity for sound field reproduction. In Proceedings of the 52nd International Conference: Sound Field Control-Engineering and Perception, Guildford, UK, 2–4 September 2013.

48. Shin, M.; Nelson, P.A.; Fazi, F.M.; Seo, J. Velocity controlled sound field reproduction by non-uniformly spaced loudspeakers. *J. Sound Vib.* **2016**, *370*, 444–464. [CrossRef]

49. Arteaga, D. An ambisonics decoder for irregular 3-D loudspeaker arrays. In Proceedings of the 134th Audio Engineering Society Convention, Rome, Italy, 4–7 May 2013.

50. Scaini, D.; Arteaga, D. Decoding of higher order ambisonics to irregular periphonic loudspeaker arrays. In Proceedings of the 55th International Conference: Spatial Audio, Helsinki, Finland, 27–29 August 2014.

51. Zuo, H.; Samarasinghe, P.N.; Abhayapala, T.D. Intensity based spatial soundfield reproduction using an irregular loudspeaker array. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2020**, *28*, 1356–1369. [CrossRef]

52. Abhayapala, T.D.; Ward, D.B. Theory and design of high order sound field microphones using spherical microphone array. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 13–17 May 2002; pp. II-1949–II-1952.

53. MH Acoustics. *Em32 Eigenmike Microphone Array Release Notes (v17. 0)*; Tech. Rep.; MH Acoustics: Summit, NJ, USA, 2013.

54. Chen, H.; Abhayapala, T.D.; Zhang, W. Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis. *J. Acoust. Soc. Amer.* **2015**, *138*, 3081–3092. [CrossRef]
55. Williams, E.G. *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*; Elsevier: Amsterdam, The Netherlands, 1999.
56. Loan, C.F.V.; Golub, G.H. *Matrix Computations*; Johns Hopkins University Press: Baltimore, MD, USA, 1983.
57. Lilis, G.N.; Angelosante, D.; Giannakis, G.B. Sound field reproduction using the lasso. *IEEE Trans. Audio Speech Lang. Process.* **2010**, *18*, 1902–1912. [CrossRef]
58. Chartrand, R.; Yin, W. Iteratively reweighted algorithms for compressive sensing. In Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 3869–3872.
59. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 21–30. [CrossRef]
60. Zotkin, D.N.; Duraiswami, R.; Gumerov, N.A. Regularized hrtf fitting using spherical harmonics. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, USA, 18–21 October 2009; pp. 257–260.
61. Zhang, W.; Abhayapala, T.D.; Kennedy, R.A.; Duraiswami, R. Insights into head-related transfer function: Spatial dimensionality and continuous representation. *J. Acoust. Soc. Amer.* **2010**, *127*, 2347–2357. [CrossRef]
62. Bernschütz, B.; Giner, A.V.; Pörschmann, C.; Arend, J. Binaural reproduction of plane waves with reduced modal order. *Acta Acust.* **2014**, *100*, 972–983. [CrossRef]
63. Schörkhuber, C.; Zaunschirm, M.; Höldrich, R. Binaural rendering of ambisonic signals via magnitude least squares. In *Proc. German Annu. Conf. Acoust. (DAGA)* **2018**, *44*, 339–342.
64. Fliege, J.; Maier, U. The distribution of points on the sphere and corresponding cubature formulae. *IMA J. Numer. Anal.* **1999**, *19*, 317–334. [CrossRef]
65. Lindau, A.; Hohn, T.; Weinzierl, S. Binaural resynthesis for comparative studies of acoustical environments. In Proceedings of the 122nd Audio Engineering Society Convention, Vienna, Austria, 5–8 May 2007.
66. Brinkmann, F.; Dinakaran, M.; Pelzer, R.; Grosche, P.; Voss, D.; Weinzierl, S. A cross-evaluated database of measured and simulated hrtfs including 3D head meshes, anthropometric features, and headphone impulse responses. *J. Audio Eng. Soc.* **2019**, *67*, 705–718. [CrossRef]
67. Fabian, B.; Manoj, D.; Robert, P.; Joschka, W.J.; Fabian, S.; Daniel, V.; Peter, G.; Stefan, W. The Hutubs Head-Related Transfer Function (Hrtf) Database. 2019. Available online: https://depositonce.tu-berlin.de/items/dc2a3076-a291-417e-97f0-7697e332c960 (accessed on 13 January 2021).
68. ITU Radiocommunication Assembly. *Itu-r bs. 1534-3: Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems*; Tech. Rep.; ITU Radiocommunication Assembly: Dubai, United Arab Emirates, 2015.