

SOUND FIELD TRANSLATION METHODS FOR BINAURAL REPRODUCTION

Lachlan Birnie, Thushara Abhayapala, Prasanga Samarasinghe Vladimir Tourbabin

Audio & Acoustic Signal Processing Group
The Australian National University, Canberra, Australia

Facebook Reality Labs
Redmond, Washington, USA

ABSTRACT

Virtual-reality reproduction of real-world acoustic environments often fix the listener position to that of the microphone. In this paper, we propose a method for listener translation in a virtual reproduction that incorporates a mix of near-field and far-field sources. Compared to conventional plane-wave techniques, the mixed-source method offers stronger near-field reproduction and translation capabilities in the case of a sparse virtualization.

Index Terms— Sound field translation, virtual-reality reproduction, binaural synthesis, higher order microphone.

1. INTRODUCTION

Sound field translation can be used to enhance virtual-reality reproductions of real-world experiences, such as orchestral performances, or sitting in the stands of a sports stadium [1, 2]. Hardware and feasibility restrictions limit us from recording large acoustic areas [3, 4], and as a result, the listener is usually stuck in a fixed acoustic position when experiencing a virtual-reality reproduction [5]. Recently, techniques of expanding a captured acoustic environment into secondary virtual plane-waves [6], and translating the secondary virtual environment, have been developed to allow for listener head rotation and movement in virtual-reality reproductions [7, 8]. These expansion techniques, however, are still restricted by underlying hardware limitations [9], as listeners are only able to translate within the acoustic sweet spot of a few centimeters that is captured by a commercial higher order microphone [10]. Translating beyond this inherent region results in spectral distortions [11], loss of localization [12], and a poor perceptual listening experience.

In this paper, we propose an alternative method for sound field expansion and translation. The method expands a mode-limited recording into a mix of secondary near-field and far-field virtual sources, in an effort to create a more perceptually accurate virtual-reality reproduction. We compare the plane-wave expansion technique (Section 3) and our proposed mixed-source expansion (Section 4) through a preliminary simulation study (Section 5). We will show that the mixed-source method offers negligible difference to the plane-wave technique when the standard closed-form expression is used for virtualization. Both methods remain to be spatially restricted by the underlying modal limitation of the original recording. However, prospects improve when we consider a sparse expansion method using the least-absolute shrinkage and selection operator (Lasso), which has been proven to help extrapolate mode-limited sound fields [13, 14]. In this sparse case, we will show that the mixed-source method offers stronger near-field characteristics, while maintaining the far-field abilities of the plane-wave technique for virtual-reality reproduction. Furthermore, the mixed-source method is capable of modeling virtual near-field sound sources, which may provide a better perceptually immersive virtual-reality reproduction of near-field sound emitters and scatterers.

2. PROBLEM FORMULATION

Consider a listener within a virtual sound field reproduction, positioned at $\mathbf{d} \equiv (r, \theta, \phi)$, $\theta \in [0, \pi]$, and $\phi \in [0, 2\pi]$ with respect to the origin O . The binaural sound perceived by the listener can be synthesized by filtering the virtual source distribution with the listener's head-related transfer functions (HRTFs) [15], expressed as

$$P_{L,R}(k, \mathbf{d}) = \int \psi(k, \mathbf{y}; \mathbf{d}) H_{L,R}(k, \mathbf{y}) d\mathbf{y}, \quad (1)$$

where the integration is over the virtual source locations of \mathbf{y} for near-field sources, and over the virtual angles $\hat{\mathbf{y}}$ for far-field sources, $P_{L,R}(k, \mathbf{d})$ are the pressure signals at the left or right ear, $k = 2\pi f/c$ is the wave number, f is the frequency, c is the speed of sound, $H_{L,R}(k, \mathbf{y})$ are the listener's HRTFs, and $\psi(k, \mathbf{y}; \mathbf{d})$ is the source distribution due to a virtual source at \mathbf{y} as observed at \mathbf{d} . Similarly, the virtual signal field received at any arbitrary position \mathbf{x} is given by

$$P(k, \mathbf{x}) = \int \psi(k, \mathbf{y}; \mathbf{x}) d\mathbf{y}, \quad (2)$$

where $\psi(k, \mathbf{y}; \mathbf{x})$ is the source distribution observed at \mathbf{x} due to the same virtual sources at \mathbf{y} . It can be shown that any real-world or virtual homogeneous sound field like (2), can be expressed with the spherical harmonic decomposition of [16]

$$P(k, \mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \alpha_{nm}(k) j_n(k|\mathbf{x}|) Y_{nm}(\hat{\mathbf{x}}), \quad (3)$$

where $|\cdot| \equiv r$, $\hat{\cdot} \equiv (\theta, \phi)$, n and m index harmonic order and mode, respectively, $j_n(\cdot)$ is the first kind spherical Bessel function, $Y_{nm}(\cdot)$ are the spherical harmonic basis functions, and $\alpha_{nm}(k)$ are the sound field coefficients.

Let us now consider an N^{th} order receiver, such as a spherical or planar microphone array [10, 17], that we use to capture a real-world sound field (3) for $\{\alpha_{nm}(k)\}$, $n \in [0, N]$. We denote this measured sound field as $P_M(k, \mathbf{x})$ with respect to the center of the microphone when $\mathbf{x} = 0$, and stress that the mode limitation introduces a spatial reproduction constraint of

$$P_M(k, \mathbf{x}) \approx P(k, \mathbf{x}), \text{ for } |\mathbf{x}| < R_x, \quad (4)$$

where R_x is the receiver region related by $N = \lceil kR_x \rceil$ [18].

Our objective is to expand a source distribution $\psi(k, \mathbf{y}; \mathbf{d})$ from a mode limited sound field measurement $P_M(k, \mathbf{x}) \cong \{\alpha_{nm}(k)\}$, $n \in [0, N]$, with the intent to relax (4) and translate the recorded sound field to $|\mathbf{d}| > R_x$ for binaural reproduction. An illustration of this translation process is presented in Fig. 1.

3. PLANE WAVE EXPANSION

In this section, we review the plane-wave translation method presented in [7] (illustrated in Fig. 1). We first present the closed-form expression for expanding a measured sound field at \mathbf{x} to an equivalent virtual plane-wave distribution $\psi(k, \hat{\mathbf{y}}; \mathbf{x})$. We then translate this distribution to \mathbf{d} , finding $\psi(k, \hat{\mathbf{y}}; \mathbf{d})$ for binaural reproduction.

We aim to represent the measured sound field with an equivalent superposition of virtual plane-wave sources, given as

$$P_M(k, \mathbf{x}) \equiv P_{EQ}(k, \mathbf{x}) = \int \psi(k, \hat{\mathbf{y}}; \mathbf{x}) \frac{1}{4\pi} e^{-ik\hat{\mathbf{y}} \cdot \mathbf{x}} d\hat{\mathbf{y}}, \quad (5)$$

where $\psi(k, \hat{\mathbf{y}}; \mathbf{x})$ is the virtual plane-wave source distribution as seen by the receiver at \mathbf{x} . Given $\psi(k, \hat{\mathbf{y}}; \mathbf{x})$ is a spherical function, it can be decomposed into spherical harmonics by

$$\psi(k, \hat{\mathbf{y}}; \mathbf{x}) = \sum_{n'=0}^{\infty} \sum_{m'=-n'}^{n'} \beta_{n'm'}(k) Y_{n'm'}(\hat{\mathbf{y}}), \quad (6)$$

where $\beta_{n'm'}(k)$ are the virtual source harmonic coefficients expanded about the receiver region $\mathbf{x} = 0$. Substituting (6) and the plane-wave decomposition [16],

$$\frac{1}{4\pi} e^{-ik\hat{\mathbf{y}} \cdot \mathbf{x}} = \sum_{n=0}^{\infty} \sum_{m=-n}^n (-i)^n Y_{nm}^*(\hat{\mathbf{y}}) j_n(k|\mathbf{x}|) Y_{nm}(\hat{\mathbf{x}}), \quad (7)$$

into (5), gives the equivalent sound field harmonic decomposition,

$$P_{EQ}(k, \mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \underbrace{(-i)^n \beta_{nm}(k)}_{\alpha_{nm}(k)} j_n(k|\mathbf{x}|) Y_{nm}(\hat{\mathbf{x}}). \quad (8)$$

From (8), we find the relationship between the virtual source harmonics $\beta_{nm}(k)$ and the measured harmonic coefficients $\alpha_{nm}(k)$. Substituting $\beta_{nm}(k) = i^n \alpha_{nm}(k)$ into (6), gives us the desired expansion for a virtual source distribution that is equivalent to the measured sound field, expressed as

$$\psi(k, \hat{\mathbf{y}}; \mathbf{x}) = \sum_{n=0}^N \sum_{m=-n}^n i^n \alpha_{nm}(k) Y_{nm}(\hat{\mathbf{y}}). \quad (9)$$

In practice, the equivalent virtual sound field (5) can be approximated with a finite set of L virtual sources, at discrete positions of $\hat{\mathbf{y}}_\ell$, for $\ell \in [1, 2, \dots, L]$, distributed about a sphere. Solving (9) for each virtual source, gives us the reproduced virtual-reality sound field as

$$P_M(k, \mathbf{x}) \equiv P_{EQ}^{PW}(k, \mathbf{x}) = \sum_{\ell=1}^L \omega_\ell \psi(k, \hat{\mathbf{y}}_\ell; \mathbf{x}) \frac{1}{4\pi} e^{-ik\hat{\mathbf{y}}_\ell \cdot \mathbf{x}}, \quad (10)$$

where ω_ℓ are suitable sampling weights for the source distribution.

For the measurement position at $\mathbf{x} = 0$, we see that each term of (10) for $P_{EQ}^{PW}(k, \mathbf{x} + \mathbf{d})$ differs only by a phase-factor $e^{-ik\hat{\mathbf{y}} \cdot \mathbf{d}}$. Combining this phase-factor with the expanded distribution $\psi(k, \hat{\mathbf{y}}_\ell; \mathbf{x})$, gives us the translated virtual plane-wave source distribution of [19]

$$\psi(k, \hat{\mathbf{y}}_\ell; \mathbf{d}) = \psi(k, \hat{\mathbf{y}}_\ell; \mathbf{x}) e^{-ik\hat{\mathbf{y}}_\ell \cdot \mathbf{d}}. \quad (11)$$

Lastly, we synthesize the translated listener's binaural signals (1) from the virtual-reality reproduction by

$$P_{L,R}^{PW}(k, \mathbf{d}) = \sum_{\ell=1}^L \omega_\ell \psi(k, \hat{\mathbf{y}}_\ell; \mathbf{d}) H_{L,R}(k, \hat{\mathbf{y}}_\ell). \quad (12)$$

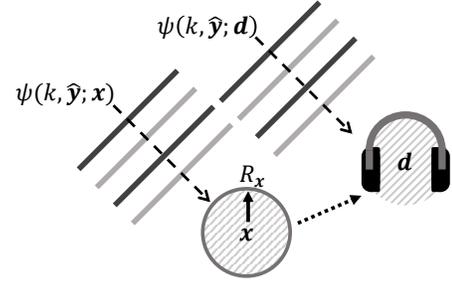


Figure 1: Virtual plane-wave expansion and translation for binaural reconstruction, where region R_x denotes the recorded sound field with an equivalent source distribution $\psi(k, \hat{\mathbf{y}}; \mathbf{x})$, and \mathbf{d} denotes the listener perceiving the translated source distribution $\psi(k, \hat{\mathbf{y}}; \mathbf{d})$.

4. MIXED SOURCE EXPANSION

In this section, we propose a mixed source distribution model of both near-field and far-field virtual sources, for sound field translation. First, we discuss the method of translating a mixed virtual source distribution $\psi(k, \mathbf{y}; \mathbf{x})$ to the listener position \mathbf{d} . Second, we present a closed-form method for expanding and translating the mixed source distribution from the measured coefficients $\alpha_{nm}(k)$. Finally, we present a sparse constrained source distribution expansion using the Lasso.

4.1. Mixed sound field translation

In this subsection, we derive an expression for the translated distribution $\psi(k, \mathbf{y}; \mathbf{d})$ in terms of the expanded distribution $\psi(k, \mathbf{y}; \mathbf{x} = 0)$, for a mix of near-field and far-field sources.

Consider a near-field point-source at \mathbf{y} , where the distribution at the source is denoted by $\psi(k, \mathbf{y})$. We can express the source distribution function observed at \mathbf{x} , with

$$\psi(k, \mathbf{y}; \mathbf{x}) = \psi(k, \mathbf{y}) \frac{e^{ik|\mathbf{y}-\mathbf{x}|}}{|\mathbf{y}-\mathbf{x}|}. \quad (13)$$

Evaluating (13) when $\mathbf{x} = 0$ gives the point-source distribution observed by the receiver, as

$$\psi(k, \mathbf{y}; \mathbf{x} = 0) = \psi(k, \mathbf{y}) \frac{e^{ik|\mathbf{y}|}}{|\mathbf{y}|}. \quad (14)$$

Substituting for $\psi(k, \mathbf{y})$ in (13) from (14), and setting $\mathbf{x} = \mathbf{d}$ gives us the desired translated distribution expression

$$\psi(k, \mathbf{y}; \mathbf{d}) = \underbrace{\psi(k, \mathbf{y}; \mathbf{x} = 0)}_{\psi(k, \mathbf{y})} \frac{e^{ik|\mathbf{y}-\mathbf{d}|}}{|\mathbf{y}-\mathbf{d}|}. \quad (15)$$

We note that the $|\mathbf{y}|e^{-ik|\mathbf{y}|}$ factor can be considered as a normalization term. Furthermore, this normalization has the property of [20]

$$\lim_{|\mathbf{y}| \rightarrow \infty} |\mathbf{y}| e^{-ik|\mathbf{y}|} \frac{e^{ik|\mathbf{y}-\mathbf{x}|}}{|\mathbf{y}-\mathbf{x}|} = e^{-ik\hat{\mathbf{y}} \cdot \mathbf{x}}, \quad (16)$$

which allows virtual plane-wave distributions to be modeled by normalized point-sources positioned in the far-field. We will utilize the normalized point-source model (16) to expand and translate a mixed equivalent sound field of near-field and far-field virtual sources next.

4.2. Mixed virtual source reproduction

We propose a mixed equivalent sound field expansion consisting of two concentric spheres of virtual point-sources. The radii of the virtual spheres R_1 and R_2 are positioned in the near-field and far-field, respectively. We express this mixed equivalent sound field as

$$P_{EQ}(k, \mathbf{x}) = \sum_{q=1}^2 \int \psi(k, R_q \hat{\mathbf{y}}; \mathbf{x}) R_q e^{-ikR_q} \frac{e^{ik|R_q \hat{\mathbf{y}} - \mathbf{x}|}}{4\pi|R_q \hat{\mathbf{y}} - \mathbf{x}|} d\hat{\mathbf{y}}, \quad (17)$$

where q indexes each virtual sphere. The spherical harmonic decomposition of a normalized point-source is given by [16]

$$|\mathbf{y}|e^{-ik|\mathbf{y}|} \frac{e^{ik|\mathbf{y}-\mathbf{x}|}}{4\pi|\mathbf{y}-\mathbf{x}|} = \sum_{n=0}^{\infty} \sum_{m=-n}^n ik|\mathbf{y}|e^{-ik|\mathbf{y}|} h_n(k|\mathbf{y}|) Y_{nm}^*(\hat{\mathbf{y}}) j_n(k|\mathbf{x}|) Y_{nm}(\hat{\mathbf{x}}), \quad (18)$$

where $h_n(\cdot)$ are the first kind spherical Hankel functions. Following a similar procedure to Section 3, we can decompose the mixed-source distribution into harmonics of $\beta_{nm}(k, |\mathbf{y}|)$ similar to (6), and substitute them along with (18) into (17), to extract the relationship between the equivalent and measured coefficients,

$$\beta_{nm}(k, |\mathbf{y}|) = \frac{\alpha_{nm}(k)}{ik|\mathbf{y}|e^{-ik|\mathbf{y}|} h_n(k|\mathbf{y}|)}. \quad (19)$$

Substituting (19) back into (6), gives the closed-form expansion for the mixed virtual source distribution as

$$\psi(k, \mathbf{y}; \mathbf{x}) = \sum_{n=0}^N \sum_{m=-n}^n \frac{\alpha_{nm}(k)}{ik|\mathbf{y}|e^{-ik|\mathbf{y}|} h_n(k|\mathbf{y}|)} Y_{nm}(\hat{\mathbf{y}}). \quad (20)$$

We can once again approximate (17) with L discrete virtual point-sources on each sphere, giving the mixed equivalent sound field as $P_{\text{EQ}}^{\text{mix}}(k, \mathbf{x}) =$

$$\sum_{q=1}^2 \sum_{\ell=1}^L \omega_{\ell} \psi(k, R_q \hat{\mathbf{y}}_{\ell}; \mathbf{x}) R_q e^{-ikR_q} \frac{e^{ik|R_q \hat{\mathbf{y}}_{\ell} - \mathbf{x}|}}{4\pi|R_q \hat{\mathbf{y}}_{\ell} - \mathbf{x}|}. \quad (21)$$

Translating the mixed-source sound field cannot be performed elegantly like the plane-wave phase-shift (11). Instead, we translate the sound field in the modal domain by decomposing $P_{\text{EQ}}^{\text{mix}}(k, \mathbf{d})$ into a shifted set of spherical harmonics about the position \mathbf{d} . The translated sound field coefficients with respect to \mathbf{d} are given by the mixed-sources at $\hat{\mathbf{y}} = \mathbf{y}_{\ell} - \mathbf{d}$, for $|\mathbf{d}| \leq R_1$, such that $\alpha_{\nu\mu}(k; \mathbf{d}) =$

$$\sum_{q=1}^2 \sum_{\ell=1}^L \omega_{\ell} \psi(k, R_q \hat{\mathbf{y}}_{\ell}; \mathbf{x}) ikR_q e^{-ikR_q} h_{\nu}(k|\hat{\mathbf{y}}|) Y_{\nu\mu}^*\left(\frac{\hat{\mathbf{y}}}{|\hat{\mathbf{y}}|}\right), \quad (22)$$

where ν and μ index the translated harmonics, centered at \mathbf{d} . The binaural signals (1) can be synthesized directly from these translated coefficients $\alpha_{\nu\mu}(k; \mathbf{d})$ and the listener's modal domain HRTFs $H_{\text{L,R}}^{\nu\mu}(k)$, with [21, 22, 15]

$$P_{\text{L,R}}^{\text{mix}}(k, \mathbf{d}) = \sum_{\nu=0}^{\infty} \sum_{\mu=-\nu}^{\nu} \alpha_{\nu\mu}(k; \mathbf{d}) H_{\text{L,R}}^{\nu\mu}(k). \quad (23)$$

4.3. Sparse expansion using the Lasso

While the closed-form expansion (20) constructs a mixed equivalent sound field, it does so by distributing energy $\psi(k, \mathbf{y}_{\ell}; \mathbf{x})$ across a majority of the virtual sources. In Section 5, we will show that this property may weaken the perceptual accuracy for a translated listener. To address this concern, we also examine a sparse constrained equivalent sound field expansion using the Lasso [23].

The measured coefficient $\alpha_{nm}(k)$ can be expressed in terms of the source distribution at the receiver $\psi(k, R_q \hat{\mathbf{y}}_{\ell}; \mathbf{x})$ with (18), as $\alpha_{nm}(k) =$

$$\sum_{q=1}^2 \sum_{\ell=1}^L \underbrace{ikR_q e^{-ikR_q} h_n(kR_q) Y_{nm}^*(\hat{\mathbf{y}}_{\ell})}_{a(nm, R_q, \hat{\mathbf{y}}_{\ell})} \times \psi(k, R_q \hat{\mathbf{y}}_{\ell}; \mathbf{x}). \quad (24)$$

We denote (24) in matrix form as

$$\boldsymbol{\alpha}(k) = \mathbf{A}(k) \boldsymbol{\psi}(k), \quad (25)$$

where $\boldsymbol{\alpha}(k) = [\alpha_{00}(k), \alpha_{1-1}(k), \dots, \alpha_{NN}(k)]^T$, $\boldsymbol{\psi}(k) = [\psi(k, R_1 \hat{\mathbf{y}}_1; 0), \dots, \psi(k, R_2 \hat{\mathbf{y}}_L; 0)]^T$, and $\mathbf{A}(k)$ is a $(N+1)^2$ by $2L$ matrix of elements $a(nm, R_q, \hat{\mathbf{y}}_{\ell})$.

A sparse equivalent source distribution $\boldsymbol{\psi}(k, \mathbf{y}_{\ell}; \mathbf{x})$ is constructed by solving the linear regression problem (25) using the Lasso [23]. In brief, the Lasso can be expressed by the sparse constrained objective function of [14]

$$\boldsymbol{\psi}(k) = \arg \min_{\boldsymbol{\psi}(k)} \|\boldsymbol{\alpha}(k) - \mathbf{A}(k) \boldsymbol{\psi}(k)\|_2^2 + \lambda \|\boldsymbol{\psi}(k)\|_1, \quad (26)$$

where the parameter λ controls the strength of the sparsity constraint for virtual signals $\boldsymbol{\psi}(k)$. We point the reader to [24, 25] for further information on compressive sensing using the Lasso. In the next section, we will compare the closed-form and sparse expansion methods in simulation, and show that the Lasso expansion is able to relax the spatial constraint of (4) for binaural synthesis.

5. SIMULATION ANALYSIS

In this section, we study the spatial characteristics of virtual source expansions for the application of translated binaural reproduction. First, we examine the point-source and mixed-source closed-form expansions. Second, we investigate the use of sparse constrained equivalent sound field expansions using the Lasso.

We use a sound field reconstruction error metric to gain insight on the virtual source expansion methods, defined as

$$\epsilon_{\text{true}}(k, \mathbf{x}) = \frac{|P(k, \mathbf{x}) - P_{\text{EQ}}(k, \mathbf{x})|^2}{|P(k, \mathbf{x})|^2}, \quad (27)$$

where $P(k, \mathbf{x})$ is the original sound field, and $P_{\text{EQ}}(k, \mathbf{x})$ is a plane-wave or mixed-source equivalent expansion. We reinforce that the aim of this paper is not to achieve perfect sound field reconstruction, but, to create a virtual acoustic environment that is immersive and perceptually accurate for a human listener. Therefore, while we use reconstruction error as a tool to gain insight, it is not a complete representation of performance.

We simulate a simple sound field for a point source at $\mathbf{y}_s = (2\text{m}, \pi/3, \pi/4)$, with a frequency of $f = 1000\text{Hz}$. The sound field expansion has $L = 625$ virtual sources, with the positions \mathbf{y}_{ℓ} arranged on Fliege nodes [26]. The mixed equivalent sound field has a near-field virtual sphere at $R_1 = 3\text{m}$, and a far-field virtual sphere at $R_2 = 20\text{m}$. The recorded sound field $P_{\text{M}}(k, \mathbf{x})$ is taken at the origin $O = [0, 0, 0]\text{m}$, and is mode limited to $N = 4$, corresponding to a spatial reproduction constraint (4) of $R_x = 0.22\text{m}$. The translated listening region is positioned at $\mathbf{d} = (0.5\text{m}, \pi/2, \pi/2)$, and is designated the same R_x region size. Sparse equivalent sound field expansions use a $\lambda = 0.001$, 500 iteration Lasso [23].

5.1. Closed-form expansion

The mixed-source and plane-wave closed-form expansions are observed to have identical sound field characteristics, so we have selected to only present the plane-wave equivalent sound field here. We show the closed-form plane-wave expansion of the xy-horizontal plane in Fig. 2(c), with the true incident sound field given in Fig. 2(a). We comment that the closed-form expansion (9) (as well as (20)) in Fig. 2(c) is an almost exact reconstruction of the mode limited recording shown in Fig. 2(b). This result of $P_{\text{EQ}}(k, \mathbf{x}) \equiv P_{\text{M}}(k, \mathbf{x})$, shows that while an equivalent sound field is not explicitly mode limited, it does, however, remain to be spatially limited (4) due to its strict approximation of the mode limited recording. Figure 2(d) illustrates the closed-form expansion's spatial limitation. We observed that the virtual reproduction has a sweet spot isolated about the recording region. Furthermore, reconstruction in the translated listening region about \mathbf{d} is seen to be erroneous, and it is believed that this indicates poor perceptual reproduction.

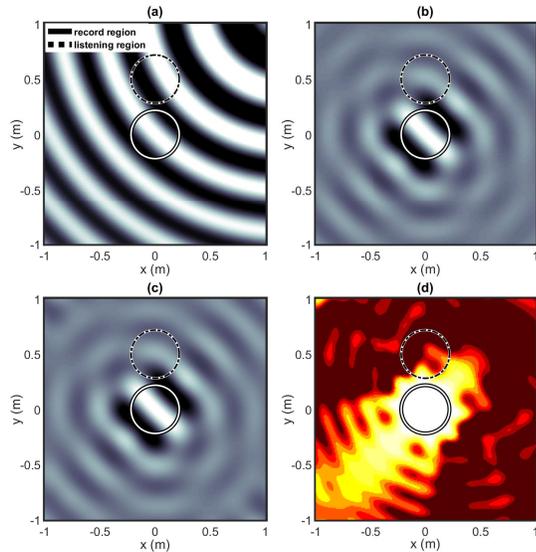


Figure 2: Equivalent plane-wave sound field, where (a): is the true incident sound field $P(k, \mathbf{x})$, (b): is the higher order recording $P_M(k, \mathbf{x})$, (c): is $P_{EQ}^{PW}(k, \mathbf{x})$ from a closed-form expansion, and (d): is the virtual plane-wave reproduction error $\epsilon_{true}^{PW}(k, \mathbf{x})$. We note that a closed-form mixed-source expansion has similar characteristics to the closed-form plane-wave expansion presented here.

5.2. Sparse expansion using the lasso

We present the sparse plane-wave and mixed-source equivalent sound fields using the Lasso in Fig. 3(a) and (c). We comment on how the Lasso expansion results in a uniform virtual equivalent sound field for the two methods. As a result, we see that the wave form about d better represents the true sound field in Fig. 2(a), potentially leading to a perceptually accurate reproduction for the translated listener. This strong outcome is in contrast to the distorted reproduction we observed for the closed-form expansion.

Next, we compare the plane-wave, Fig. 3(a), and the mixed-source, Fig.3(c), virtual reproductions. The plane-wave reproduction is observed to be far-field like in nature. Whereas, the mixed-source reproduction is seen to have a more prominent curvature, representative of the true near-field sound source. This near-field nature of the mixed-source reproduction is believed to provide a superior auditory experience for the translated listener. Furthermore, the sound field reconstruction errors in Fig. 3(b) and (c) reinforce this notion. The plane-wave sweet spot is seen to extrapolate from the recording region, but struggles to reach the translated listener at d , instead extending predominantly in the wave field direction. The mixed-source reconstruction is shown to behave similarly, but with a larger extrapolated sweet spot that is able to cover the translated listening position.

Finally, we present the distribution of activated virtual sound sources in closed-form and Lasso expansions. For this simulation, we introduce a second sound source in the far-field, such that the true incident sound field $P(k, \mathbf{x})$ is excited by point-sources at $\mathbf{y}_{s1} = (2m, \pi/3, \pi/4)$, $\mathbf{y}_{s2} = (50m, \pi/3, 3\pi/4)$. Figure 4 depicts the distribution of virtual signals $|\psi(k)|$ (26), normalized to $[0, 1]$ over ϕ , for the plane-wave and mixed-source equivalent sound fields. We observe a significant spread of activated sources for the closed-form expansions in Fig. 4(a). Furthermore, the mixed-source distribution is seen to utilize both the inner sphere and outer sphere virtual sources for both the near-field and far-field incident

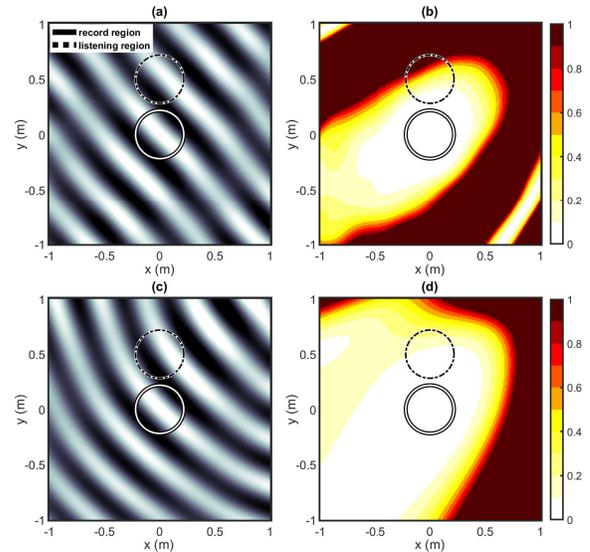


Figure 3: Virtual reproductions (left) and their reconstruction errors $\epsilon_{true}(k, \mathbf{x})$ (right). For (a) & (b): plane-wave expansion using the Lasso, and (c) & (d): mixed-source expansion using the Lasso.

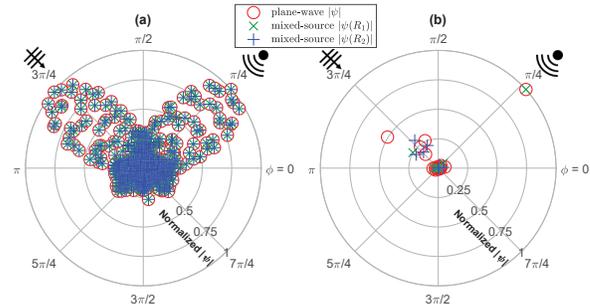


Figure 4: Distribution of activated virtual sound sources in plane-wave and mixed-source equivalent sound fields, for (a): closed-form expansion, and (b): expansion using the Lasso, where the true incident sound field is excited by a near-field source at $(2m, \pi/3, \pi/4)$ and a far-field source at $(50m, \pi/3, 3\pi/4)$.

sound components. On the other hand, the mixed-source Lasso expansion is able to differentiate incident near-field and far-field components with similarly distant virtual sources, as shown in Fig. 4(b). This attribute of modeling near-field components with close virtual sources, and far-field components with far sources, is believed to offer a stronger perceptual experience for the listener.

6. CONCLUSION

We have proposed an alternative approach for virtual-reality reproduction of real-world environments using a mix of near-field and far-field virtual sources. The mixed-source and conventional plane-wave translation techniques remained to be spatially restricted by the recording's underlying modal limit, for closed-form expansion methods. However, for sparse constrained expansion methods, the mixed-source model is shown to be beneficial. It offered near-field like sound field reconstruction with similarly positioned virtual sources, which may lead to stronger perceptual accuracy for translated listeners. We have not presented any examinations with room acoustics, or any listening tests applying the translation methods with real HRTF data. These are assignments we hope to address in the near future.

7. REFERENCES

- [1] J. G. Tylka and E. Y. Choueiri, "Models for evaluating navigational techniques for higher-order ambisonics," in *Proc. Meetings on Acoust.*, vol. 30, no. 1. ASA, 2017, p. 050009.
- [2] Y. Suzuki, T. Okamoto, J. Trevino, Z.-L. Cui, Y. Iwaya, S. Sakamoto, and M. Otani, "3d spatial sound systems compatible with human's active listening to realize rich high-level kansei information," *Interdisciplinary information sciences*, vol. 18, no. 2, pp. 71–82, 2012.
- [3] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2. IEEE, 2002, pp. II–1949.
- [4] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, 2005.
- [5] C. D. Salvador, S. Sakamoto, J. Trevino, and Y. Suzuki, "Spatial accuracy of binaural synthesis from rigid spherical microphone array recordings," *Acoustical Science and Technology*, vol. 38, no. 1, pp. 23–30, 2017.
- [6] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. G., "Plane-wave decomposition analysis for spherical microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.* IEEE, 2005, pp. 150–153.
- [7] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *Proc. of 52nd Intl. Aud. Eng. Soc. Conf. on Sound Field Control-Engineering and Perception.* Audio Engineering Society, 2013.
- [8] T. Pihlajamaki and V. Pulkki, "Synthesis of complex sound scenes with transformation of recorded spatial sound in virtual reality," *J. Audio Eng. Soc.*, vol. 63, no. 7/8, pp. 542–551, 2015.
- [9] J. G. Tylka and E. Choueiri, "Comparison of techniques for binaural navigation of higher-order ambisonic soundfields," in *Audio Eng. Soc. Conf. 139.* Audio Engineering Society, 2015.
- [10] M. Acoustics, "Em32 eigenmike microphone array release notes (v17. 0)," 25 Summit Ave, Summit, NJ 07901, USA, 2013.
- [11] N. Hahn and S. Spors, "Modal bandwidth reduction in data-based binaural synthesis including translatory head-movements," in *Proc. German Annu. Conf. Acoust.(DAGA)*, 2015, pp. 1122–1125.
- [12] F. Winter, F. Schultz, and S. Spors, "Localization properties of data-based binaural synthesis including translatory head-movements," in *Proc. Forum Acusticum*, vol. 31, 2014.
- [13] Y. Maeno, Y. Mitsufuji, and T. D. Abhayapala, "Mode domain spatial active noise control using sparse signal representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2018, pp. 211–215.
- [14] Y. Hu, P. N. Samarasinghe, T. D. Abhayapala, and G. Dickins, "Modeling characteristics of real loudspeakers using various acoustic models: Modal-domain approaches," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* IEEE, 2019, pp. 561–565.
- [15] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions," *EURASIP J. Audio, Speech, Music Process.*, vol. 2019, no. 1, p. 5, 2019.
- [16] E. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography.* London, UK: Academic Press, 1999.
- [17] H. Chen, T. D. Abhayapala, and W. Zhang, "Theory and design of compact hybrid microphone arrays on two-dimensional planes for three-dimensional soundfield analysis," *J. Acoust. Soc. Amer.*, vol. 138, no. 5, pp. 3081–3092, 2015.
- [18] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, "Intrinsic limits of dimensionality and richness in random multipath fields," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2542–2556, 2007.
- [19] D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," *J. Acoust. Soc. Amer.*, vol. 121, no. 3, pp. 1559–1563, 2007.
- [20] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech, Audio Process.*, vol. 9, no. 6, pp. 697–707, 2001.
- [21] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Regularized hrtf fitting using spherical harmonics," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.* IEEE, 2009, pp. 257–260.
- [22] W. Zhang, T. D. Abhayapala, R. A. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation," *J. Acoust. Soc. Amer.*, vol. 127, no. 4, pp. 2347–2357, 2010.
- [23] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 8, pp. 1902–1912, 2010.
- [24] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [25] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [26] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA J. Numer. Anal.*, vol. 19, no. 2, pp. 317–334, 1999.