On the Upscaling of Higher-Order Ambisonics Signals for Sound Field Translation

Maximilian Kentgens*, Shahd Al Hares†, Peter Jax‡

Institute of Communication Systems (IKS), RWTH Aachen University, Germany
*kentgens@iks.rwth-aachen.de, †shahd.al.hares@rwth-aachen.de, ‡jax@iks.rwth-aachen.de

***Abstract*—We propose a framework for improved sound field translation of higher-order Ambisonics signals extending the limits of conventional methods. In a first stage, the input signal is upscaled to an increased spherical harmonics truncation order using sparse plane wave recovery. This allows to provide additional spatial information to the second stage where any existing translation method can be used. In order to gain insights into the system's behavior, key parameters are assessed individually. We demonstrate the potential of the proposed framework by operating the plane wave translation with large displacements beyond its original capabilities. In an extensive simulation, the performance could be improved significantly. Our proposition may find application in recorded virtual reality applications in which the user is able to perform translational movements in the scene.**

***Index Terms*—spherical harmonics, higher-order Ambisonics, sound field translation, 3DoF+, sparse recovery**

I. Introduction

In virtual reality, sound field translation, i.e., the ability to evaluate the sound field at different positions from the original origin, is the technological key requirement to allow users to move freely in a recorded acoustic scene. This is where recordings from a single *spherical microphone array* (SMA) [1], [2] and the related *higher-order Ambisonics* (HOA) framework, i.e., the representation of sound fields using *spherical harmonics* (SH) [3]–[5], find their limitations. The *plane wave translation* straightly derived from theory [6], [7] fails to estimate the sound field at displacements larger than a few centimeters as SH truncation orders are low in practice due to physical constraints of today's recording technology.

In order to overcome these limitations, various strategies have been proposed which often exploit the properties of human spatial hearing. Besides fixed filter methods, e.g., [8], [9], several approaches perform an active decomposition of the signal to control an adaptive filter. The methods proposed in [10], [11] are highly engineered systems with a complex interaction between their individual components.

The authors of [10] employ methods from the mathematical framework of compressed sensing. In the context of spatial audio recordings, such methods have already been studied in the past, for instance to reduce spatial aliasing artifacts [12], to improve reproduction by increasing spatial resolution and sweet spot size [13], for beamforming [14] or to apply dereverberation [15].

Recently, Birnie et al. [16] proposed a method for spatially translated binaural reproduction. They showed how sparse recovery can be used to improve the performance of the plane wave translation method. However, their joint consideration of sound field translation and reproduction method does not allow to use alternative reproduction approaches such as playback on loudspeaker arrays. In the present paper, we propose a sparse-recovery-based translation framework which provides a more flexible HOA signal at its output. Moreover, we separate the sparse recovery from the actual translation process. This allows us to interchange the translation methods, e.g., by different of the above-mentioned fixed-filter translation methods, and to investigate the behavior of each functional stage individually. In the first stage, the order-limited input HOA signal is *upscaled* to an increased SH truncation, a methodology accredited to Wabnitz et al. [17], [18]. The upscaled signal has an increased spatial resolution which can be exploited to overcome shortcomings of the translation method in the second stage. We demonstrate how the upscaling allows to operate the plane wave translation beyond its original capabilities, i.e., a range extender so to speak.

The paper is organized as follows: the SH fundamentals and basic methods for sound field translation are recapped in Sec. II. In Sec. III, we introduce the concept of upscaling, present our novel framework and discuss key aspects. Sec. IV provides simulation results which show the influence of different parameters and reveal insights into the interaction of the individual system components. Finally, we draw an overall picture of the system's potentials and limitations.

II. Fundamentals

A. Spherical Harmonics Description of the Sound Field

The complex-valued sound field pressure $\breve{p}(k, \vec{x})$ over Cartesian coordinates $\vec{x} \in \mathbb{R}^3$ can be modeled by superposition of plane wave sources $\breve{\psi}^{(\mathrm{PW})}(k, \vec{s}_\mathrm{u})$ impinging into the coordinate system's origin from directions $\vec{s}_\mathrm{u}$ as [19]

$$\breve{p}(k, \vec{x}) = \frac{1}{4\pi} \int_{\mathcal{S}^2} \breve{\psi}^{(\mathrm{PW})}(k, \vec{s}_\mathrm{u}) \mathrm{e}^{-\mathrm{i}k\vec{s}_\mathrm{u}\cdot\vec{x}} \, \mathrm{d}\vec{s}_\mathrm{u}. \tag{1}$$

Here, i is the imaginary unit. The wave number $k = \frac{\omega}{c}$ is proportional to the angular frequency ω in case of a constant speed of sound c. The notation $\vec{(\cdot)}_\mathrm{u}$ denotes a normalization of a vector to unit length. The surface element on the unit sphere $\mathcal{S}^2$ is $\mathrm{d}\vec{s}_\mathrm{u} \equiv \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$ with inclination angle θ and azimuth angle ϕ.

Equivalently, the sound field can also be described using a distribution of point sources $\psi^{(\mathrm{PS})}(k, s_0\vec{s}_u)$ at a fixed distance s_0 from the origin. The sound field is then given by [4], [19]

$$\breve{p}(k,\vec{x}) = \frac{1}{4\pi}\int_{\mathcal{S}^2} \breve{\psi}^{(\mathrm{PS})}(k, s_0\vec{s}_\mathrm{u}) \frac{\mathrm{e}^{\mathrm{i}\|s_0\vec{s}_\mathrm{u}-\vec{x}\|}}{\|s_0\vec{s}_\mathrm{u}-\vec{x}\|}\frac{s_0}{\mathrm{e}^{\mathrm{i}ks_0}}\,\mathrm{d}\vec{s}_\mathrm{u}. \quad (2)$$

Note that (2) describes an interior problem and is only valid for $\|\vec{x}\| < s_0$.

Both the signals $\breve{\psi}^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$ and $\breve{\psi}^{(\mathrm{PS})}(k, s_0\vec{s}_\mathrm{u})$ can be transformed to the SH domain in which we can express a relation between them.

We assume that $\breve{\psi}(k,\vec{s})$ is either $\breve{\psi}^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$ or $\breve{\psi}^{(\mathrm{PS})}(k, s_0\vec{s}_\mathrm{u})$. $\breve{\psi}(k,\vec{s})$ can then be expressed as a SH series, [19]

$$\breve{\psi}(k,\vec{s}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \psi_{nm}(k,\|\vec{s}\|) Y_n^m(\vec{s}_\mathrm{u}) \quad (3)$$

with SH basis functions $Y_n^m(\vec{s}_\mathrm{u}) \equiv Y_n^m(\theta,\phi)$ with order $n = 0, 1, 2, \ldots$ and degree m, $-n \le m \le n$. The inverse relation is given by

$$\psi_{nm}(k,\|\vec{s}\|) = \int_{\mathcal{S}^2} \breve{\psi}(k,\vec{s})\,[Y_n^m(\vec{s}_\mathrm{u})]^*\;\mathrm{d}\vec{s}_\mathrm{u}. \quad (4)$$

Here, $(\cdot)^*$ denotes the complex conjugate.

Equating (1) and (2) and performing the SH transform (4) yields the relation [8], [20]

$$\psi_{nm}^{(\mathrm{PW})}(k) = \frac{s_0}{\mathrm{e}^{\mathrm{i}ks_0}}\frac{\mathrm{i}kh_n(ks_0)}{(-\mathrm{i})^n}\psi_{nm}^{(\mathrm{PS})}(k, s_0). \quad (5)$$

$h_n(\cdot)$ is the spherical Hankel function of the first kind. $\psi_{nm}^{(\mathrm{PW})}(k)$ and $\psi_{nm}^{(\mathrm{PS})}(k, s_0)$ are asymptotically identical, i.e., for $ks_0 \gg n$. In other words, the point sources become plane waves for $s_0 \to \infty$.

B. Spherical Harmonics Band Limitation

In practice, only a finite number of coefficients $\psi_{nm}(k,\|\vec{s}\|)$ up to a certain truncation order $n \le N$ are available, e.g., $N = 4$ [21]. In this case, (3) is approximated using

$$\psi(k,\vec{s}) = \sum_{n=0}^{N}\sum_{m=-n}^{n} \psi_{nm}(k,\|\vec{s}\|) Y_n^m(\vec{s}_\mathrm{u}). \quad (6)$$

The signal $\psi(k,\vec{s})$ can be considered a spatially low-passed version of the true signal $\breve{\psi}(k,\vec{s})$.

When evaluating (6) on a nearly-uniform discrete sampling grid $\mathcal{Q} = \{\vec{s}_{\mathrm{u},q} | q = 1,\ldots,Q\}$ with Q sampling points, (6) can be rephrased in vector notation as

$$\boldsymbol{\psi} = \boldsymbol{Y}_{N,\mathcal{Q}}\,\boldsymbol{\psi_{nm}} \quad (7)$$

with $\boldsymbol{\psi} = [\psi(k,\vec{s}_1),\ldots,\psi(k,\vec{s}_Q)]^\mathrm{T}$ and $\boldsymbol{\psi_{nm}} = [\psi_{0,0}(k,\|\vec{s}\|),\ldots,\psi_{N,N}(k,\|\vec{s}\|)]^\mathrm{T}$. We omit the dependency on k for better readability. Note that when $\boldsymbol{\psi_{nm}}$ is $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$, the coefficient vector is equivalent to a HOA signal in the frequency domain. For the case of $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PS})}$, the coefficients can be considered a *near-field-compensated* HOA (NFC-HOA) signal [20]. The SH matrix $\boldsymbol{Y}_{N,\mathcal{Q}}$ has dimension $Q\times(N+1)^2$.

The coefficient vector $\boldsymbol{\psi_{nm}}$ can be reconstructed from $\boldsymbol{\psi}$ if $Q \ge (N+1)^2$ and $\boldsymbol{Y}_{N,\mathcal{Q}}$ has full rank,

$$\boldsymbol{\psi_{nm}} = \boldsymbol{Y}_{N,\mathcal{Q}}^\dagger\,\boldsymbol{\psi}. \quad (8)$$

Here, $(\cdot)^\dagger$ denotes the Moore-Penrose pseudoinverse.

Applying the band-limited signals $\psi^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$ or $\psi^{(\mathrm{PS})}(k, s_0\vec{s}_\mathrm{u})$ in (1) or (2), respectively, results in a sound field $p(k,\vec{x})$ which is an adequate approximation of $\breve{p}(k,\vec{x})$ in the region

$$k\|\vec{x}\| < N \quad (9)$$

as shown in [4]. This *sweet spot* size strongly depends on frequency and truncation order. For $N = 4$, it corresponds to a diameter of $87\,\mathrm{cm}$ at $500\,\mathrm{Hz}$ and only $8.7\,\mathrm{cm}$ at $5\,\mathrm{kHz}$.

C. Sound Field Translation

Sound field translation methods aim to find a signal $\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}}$ from a given signal $\boldsymbol{\psi_{nm}}$ which is an appropriate estimate of the sound field around a shifted position $\vec{d}$, i.e., $\tilde{p}(k,\vec{x}) \approx \breve{p}(k,\vec{x}+\vec{d})$. In the following, we are going to recap two state-of-the-art methods which both assume a linear relation between $\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}}^{(\mathrm{PW})}$ and $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$ using a translation matrix $\boldsymbol{T}_N$,

$$\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}}^{(\mathrm{PW})} = \boldsymbol{T}_N\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}. \quad (10)$$

The *plane wave translation* method is straightforward to derive by replacing $\vec{x}$ by $\vec{x}+\vec{d}$ in (1) which results in a phase rotation of $\breve{\psi}^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$. However, since $\breve{\psi}^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$ is unavailable, the operation is performed on $\psi^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})$ resulting in [6], [8], [22]

$$\tilde{\psi}^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u}) = \psi^{(\mathrm{PW})}(k,\vec{s}_\mathrm{u})\,\mathrm{e}^{-\mathrm{i}k\vec{s}_\mathrm{u}\cdot\vec{d}}, \quad (11)$$

which is practically limited to very small translations $k\|\vec{d}\| \ll N$ (cf. (9)). This poses a major practical limitation.

As a consequence, other methods were developed to overcome the limitation of (11). The *space warping-based translation* presented in [8] defines

$$\tilde{\psi}^{(\mathrm{PW})}(k,\tilde{\theta},\phi) = g(\tilde{\theta})\psi^{(\mathrm{PW})}(k, f^{-1}(\tilde{\theta}),\phi). \quad (12)$$

f^{-1} is the inverse of the warping function $\tilde{\theta} = f(\theta)$. The choice [8]

$$f(\theta) = \cos^{-1}\left(\frac{\cos(\theta)-\beta}{\sqrt{1-2\beta\cos(\theta)+\beta^2}}\right) \quad (13)$$

has the intent of mapping the original source directions to the source directions seen from the translated position according to a geometric model where all sources are assumed to be in a constant distance from the origin s_0. The warping factor $\beta = d/s_0, |\beta| < 1$ describes the displacement relative to the sources. The effect of the operation is a squeezing of angles for a movement away from the sources and angular spreading for a movement towards a source. Note that for the sake of simplified notation, (12) is limited to displacements in the direction of the spherical coordinate system's zenith, i.e., $\vec{d} = d\vec{e}_z$, without loss of generality.

In [8], different strategies for the choice of the equalization function $g(\tilde{\theta})$ were proposed. We here stick to the *amplitude-adjusted* variant. This particular variant neglects physically

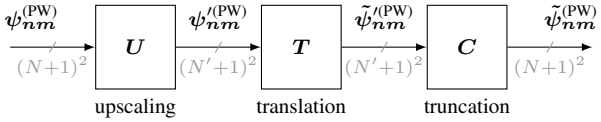


Fig. 1. Block diagram

correct phase adjustments in favor of an expected improvement in psychoacoustic performance of the translation method even for translations beyond $k\|\vec{d}\| > N$.

III. Proposed Framework

In Sec. II-C, we learned that a low truncation order N poses a major limitation on the translation of the sound field. To overcome this limitation, the idea of the method proposed in this section is to perform an *order upscaling* of the given signal to an increased order $N' > N$ prior to the translation operation as illustrated in Fig. 1. The upscaling is based on a sparse source model, which will be introduced in the following.

A. Sparse Source Model

Instead of reconstructing the spatially low-pass filtered signal $\boldsymbol{\psi}$ from $\boldsymbol{\psi}_{nm}$ using (7), we aim at reconstructing the original signal $\breve{\psi}(k,\vec{s})$ from (3) sampled on a dense grid $\mathcal{Q}$, denoted as $\breve{\boldsymbol{\psi}}$. That is solving the underdetermined problem

$$\boldsymbol{Y}_{N,\mathcal{Q}}^{\dagger}\,\breve{\boldsymbol{\psi}} = \boldsymbol{\psi}_{\boldsymbol{nm}} \tag{14}$$

for $\breve{\boldsymbol{\psi}}$. We choose the number of sampling points Q sufficiently high to later (in Sec. III-B) reconstruct the upscaled signal with $(N'+1)^2$ coefficients, i.e., $Q \geq (N'+1)$. In order to improve the estimate for $\breve{\boldsymbol{\psi}}$ over the least-squares solution (7), we introduce a source model with a sparse source distribution $\boldsymbol{\chi}$ consisting of $I \ll (N+1)^2$ sources χ_i impinging from $(\theta_{\chi_i},\phi_{\chi_i})$ and additive noise $\boldsymbol{\nu}$,

$$\begin{aligned}\breve{\boldsymbol{\psi}} &= \boldsymbol{\chi} + \boldsymbol{\nu} \\ &= \sum_{i=1}^{I} \chi_i \boldsymbol{e}_{q|(\theta_q,\phi_q)=(\theta_{\chi_i},\phi_{\chi_i})} + \boldsymbol{\nu}.\end{aligned} \tag{15}$$

Here, $\boldsymbol{e}_q$ is the Q-dimensional standard basis vector which is all zero except for the q-th entry. Further, we assume that $\boldsymbol{\chi}$ is dominant, i.e., $\mathrm{E}\{\boldsymbol{\chi}^{\mathrm{H}}\boldsymbol{\chi}\} \gg \mathrm{E}\{\boldsymbol{\nu}^{\mathrm{H}}\boldsymbol{\nu}\}$, and that $\boldsymbol{\nu}$ is diffuse, i.e., $\mathrm{E}\{\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{H}}\} = \frac{1}{Q}\,\mathrm{E}\{\boldsymbol{\nu}^{\mathrm{H}}\boldsymbol{\nu}\}\boldsymbol{I}$. Here, $\boldsymbol{I}$ denotes the identity matrix.

Note that the source model (15) implies that all sources are at a constant distance from the origin, which is infinity in case of $\boldsymbol{\psi}_{\boldsymbol{nm}}$ being $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$, or s_0 in case of $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PS})}$. Although this limitation can be overcome by using source distributions on multiple shells at different radii as proposed in [16], the simple model (15) facilitates the systematic evaluation in the remainder of this paper.

B. Upscaling

Order upscaling [17], [18] of a truncated signal to an order $N' > N$ aims to recover a matrix $\boldsymbol{\Omega}$ such that $\boldsymbol{\Omega}\,\boldsymbol{\psi}_{\boldsymbol{nm}} = \breve{\boldsymbol{\psi}}$ and therefore solves (14). Instead of minimizing the square norm, which yields a solution in which the energy is distributed evenly among the coefficients, we want to exploit the sparsity assumption on our source model (15). Since the noise $\boldsymbol{\nu}$ violates the assumption of sparsity, L consecutive observations are taken into account to reduce the impact on the recovery of $\boldsymbol{\Omega}$. Problem (14) is then extended to

$$\boldsymbol{Y}_{N,\mathcal{Q}}^{\dagger}\,\breve{\boldsymbol{\Psi}} = \boldsymbol{\Psi}_{\boldsymbol{nm}} \tag{16}$$

with $\breve{\boldsymbol{\Psi}} = [\breve{\boldsymbol{\psi}}(1),\breve{\boldsymbol{\psi}}(2),\ldots,\breve{\boldsymbol{\psi}}(L)]$ and $\boldsymbol{\Psi}_{\boldsymbol{nm}} = [\boldsymbol{\psi}_{\boldsymbol{nm}}(1),\boldsymbol{\psi}_{\boldsymbol{nm}}(2),\ldots,\boldsymbol{\psi}_{\boldsymbol{nm}}(L)]$. We use the regularized-MFOCUSS algorithm, described in [23], to recover $\boldsymbol{\Omega}$ by solving the minimization problem

$$\begin{aligned}\hat{\boldsymbol{\Omega}} = \underset{\boldsymbol{\Omega}}{\operatorname{argmin}} \Big(&\Big\| \boldsymbol{Y}_{N,\mathcal{Q}}^{\dagger}\,\boldsymbol{\Omega}\,\boldsymbol{\Psi}_{\boldsymbol{nm}} - \boldsymbol{\Psi}_{\boldsymbol{nm}} \Big\|_F^2 \\ &+ \lambda J(\boldsymbol{\Omega}\,\boldsymbol{\Psi}_{\boldsymbol{nm}}) \Big)\end{aligned} \tag{17}$$

where $\|.\|_F$ is the Frobenius norm, λ is a regularization factor, and $J(\cdot)$ is the ℓ_{12}-norm. It first calculates the ℓ_2-norm over the L observations, and then computes the ℓ_1-norm over the sampling grid $\mathcal{Q}$, i.e.,

$$J(\breve{\boldsymbol{\Psi}}) = \sum_{q=1}^{Q} \left(\sum_{l=1}^{L} \left| \breve{\psi}_{ql} \right|^2 \right)^{\frac{1}{2}}. \tag{18}$$

To cope with sources at any finite distance s_0, we have to perform the minimization (17) on the NFC-HOA signal $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PS})}$. However, we aim at constructing an upscaling matrix $\boldsymbol{U}_{N\to N'}$ to be applied on $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$. For this reason, we introduce a diagonal matrix $\boldsymbol{Z}_{\mathrm{PS}\to\mathrm{PW}}$ which describes the transformation between $\psi_{nm}^{(\mathrm{PW})}$ and $\psi_{nm}^{(\mathrm{PS})}$ according to (5). Accordingly, we define $\boldsymbol{Z}_{\mathrm{PW}\to\mathrm{PS}}$ as its inverse.

Once the matrix $\hat{\boldsymbol{\Omega}}$ is obtained, the upscaling matrix $\boldsymbol{U}_{N\to N'}$ can be constructed on the basis of the product $\boldsymbol{Y}_{N',\mathcal{Q}}^{\dagger}\,\hat{\boldsymbol{\Omega}}$ as

$$\boldsymbol{U}_{N\to N'} = \boldsymbol{Z}_{\mathrm{PS}\to\mathrm{PW}}\,\boldsymbol{Y}_{N',\mathcal{Q}}^{\dagger}\,\hat{\boldsymbol{\Omega}}\,\boldsymbol{Z}_{\mathrm{PW}\to\mathrm{PS}}. \tag{19}$$

The upscaled HOA coefficients can then be calculated as

$$\boldsymbol{\Psi}_{\boldsymbol{nm}}^{\prime(\mathrm{PW})} = \boldsymbol{U}_{N\to N'}\,\boldsymbol{\Psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}. \tag{20}$$

As upscaling requires prior knowledge on s_0, the impact of inaccurate estimates of it will be investigated in Sec. IV-B.

C. Proposed Architecture

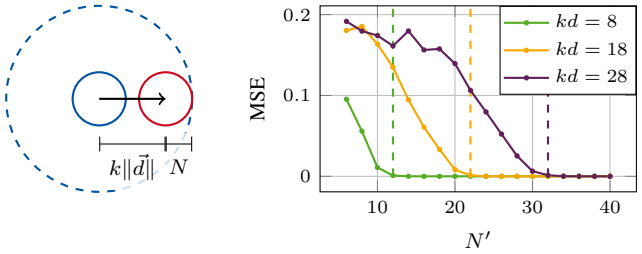
Instead of applying any sound field translation method on $\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$ as in (10), we start from its upscaled version $\boldsymbol{\psi}_{\boldsymbol{nm}}^{\prime(\mathrm{PW})}$ obtained via (20). The effect of applying the translation operation $\boldsymbol{T}_{N'}$ on $\boldsymbol{\psi}_{\boldsymbol{nm}}^{\prime}$ is a decreased displacement relative to the truncation order $\frac{ks_0}{N'} < \frac{ks_0}{N}$. The result of $\boldsymbol{T}_{N'}\,\boldsymbol{U}_{N\to N'}\,\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}$, however, is a signal of order N' which again has to be truncated to the desired output order N using a *truncation matrix*

$$\boldsymbol{C}_{N'\to N} = \underbrace{[\boldsymbol{I} \quad \boldsymbol{0}]}_{(N+1)^2 \times (N'+1)^2}, \tag{21}$$

i.e., the unity matrix $\boldsymbol{I}$ with $(N+1)^2$ diagonal elements is padded using the zero-matrix $\boldsymbol{0}$. In total, we yield an overall operation $\boldsymbol{T}_N^{\prime}$ of dimension $(N+1)^2 \times (N+1)^2$ which retains the signal order N at the output,

$$\begin{aligned}\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}}^{(\mathrm{PW})} &= \boldsymbol{C}_{N'\to N}\,\boldsymbol{T}_{N'}\,\boldsymbol{U}_{N\to N'}\,\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})} \\ &= \boldsymbol{T}_N^{\prime}\,\boldsymbol{\psi}_{\boldsymbol{nm}}^{(\mathrm{PW})}.\end{aligned} \tag{22}$$

The overall procedure (22) is illustrated in Fig. 1.



(a) Geometrical interpretation of (23)

(b) Experimental evaluation

Fig. 2. Choice of the upscaling order N'

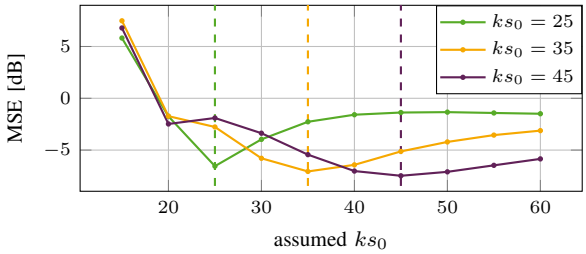


Fig. 3. Impact of estimation accuracy of s_0

D. Choice of Upscaling Order

Under the assumption that any sound field translation algorithm is able to well reconstruct the sound field within the signal's sweet spot $k\|\vec{x}\| < N'$ analog to (9), N' has to be chosen such that the target region of interest for the translated sound field $k\|\vec{x} - \vec{d}\| < N$ is fully covered[1]. This suggests to choose the upscaled translation order N' as

$$N' > k\|\vec{d}\| + N. \tag{23}$$

Its geometrical interpretation is shown in Fig. 2a. Eq. (23) will be verified next.

IV. Experimental Evaluation

In the following, we investigate the impact of the key parameters N' and the source distance estimate s_0. Moreover, we conduct a comprehensive evaluation of the overall system to draw a big picture of the system's potentials and limitations.

A. Choice of Upscaling Order

In order to validate (23), an experiment was conducted in which the order N' of a simulated oracle upscaled signal $\boldsymbol{\psi}'_{\boldsymbol{nm}}$, i.e., the result of an ideal upscaling stage, was varied. Plane wave translation (11) with three different displacements $k\|\vec{d}\| \in \{8, 18, 28\}$ was applied to obtain $\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}} = \boldsymbol{C}_{N'\to N}\boldsymbol{T}_{N'}\boldsymbol{\psi}'_{\boldsymbol{nm}}$ with $N = 4$. Fig. 2b shows the mean squared error MSE $= \frac{1}{L}\sum_l \|\tilde{\boldsymbol{\psi}}_{\boldsymbol{nm}}(l) - \mathring{\boldsymbol{\psi}}_{\boldsymbol{nm}}(l)\|_2^2$ of $L = 200$ observations in a completely diffuse sound field. Here, $\mathring{\boldsymbol{\psi}}_{\boldsymbol{nm}}(l)$ denotes the ground truth signal at the translated position. The error decreases strongly with increasing N' and is low for those N' which meet the condition (23), marked by the dashed lines.

B. Impact of Source Distance Estimate

As it is of great practical relevance, we investigate the impact of a model mismatch on the upscaling stage when the assumed value of s_0 does not match the actual source distances. In this experiment, the upscaled signal $\boldsymbol{\Psi}'_{\boldsymbol{nm}}$ is compared to the ground truth $\breve{\boldsymbol{\Psi}}'_{\boldsymbol{nm}}$, i.e., no translation is considered. Note that for translation methods which incorporate a geometrical model such as the warping-based translation method (12), the assumed value of s_0 has a double impact, which is not considered here.

[1]Remember that the area of interest $k\|\vec{x} - \vec{d}\| < N$ is furthermore only meaningful for $\|\vec{x}\| < s_0$ due to the interior problem definition of (2).

Two point sources with varying source distances $ks_0 \in \{25, 35, 45\}$ and amplitudes $s_1 = 0.9$ and $s_2 = 0.7\mathrm{e}^{-\mathrm{i}0.85\pi}$ from directions $(\theta, \phi) = (40°, 25°)$ and $(150°, -45°)$ were simulated with additive diffuse noise at an SNR of 20 dB yielding a signal $\boldsymbol{\psi}_{\boldsymbol{nm}}$ with $N = 4$. For the upscaling stage, $L = 20$ observations were taken into account. The algorithm was parametrized with $Q = 900$ and $\lambda = 10^{-4}$. The assumed source distance passed to the upscaling stages was varied independent of the actual s_0. The MSE between $\boldsymbol{\Psi}'_{\boldsymbol{nm}}$ and $\breve{\boldsymbol{\Psi}}'_{\boldsymbol{nm}}$ was calculated over 100 repetitions of the experiment.

The results are shown in Fig. 3. The best performance is achieved when the assumed source distance matches the actual one. For inaccurate assumptions, $\breve{\psi}$ becomes less sparse resulting in a worse upscaling performance. Inaccuracies of the assumed source distance have a large impact at close distances and are less critical for greater distances. Moreover, the translation is more robust to overestimation than underestimation.

C. Overall System Evaluation

Aiming at a comprehensive evaluation of the overall system, we repeatedly conducted a simulation with stochastic variations of the acoustical scene. The number of point sources and their source distances were randomly picked from $I \in \{1, 2, 3, 4\}$ and $ks_0 \in \{25, 50, 100\}$, respectively, with random amplitudes and phases. Additive diffuse noise was added to the signal. In each experiment, plane wave and warping-based translation both with and without upscaling was applied on $L = 20$ observations of the plane-wave expansion $\boldsymbol{\psi}_{\boldsymbol{nm}}$ with $N = 4$. The translation displacement was $k\|\vec{d}\| = 9.6$. Each experiment was repeated 300 times at different diffuse noise levels in order to capture the SNR dependency. The upscaling hyper-parameters were chosen as $Q = 900$ and λ was empirically adjusted to the SNR.

Since the MSE is not meaningful for those translation methods which do not aim at a physically correct reconstruction of the sound field, we also employ the phase-neglecting, psychoacoustically motivated spatial fidelity measure [8]

$$\sigma = 1 - \frac{2}{\pi}\cos^{-1}\left(\frac{\int_{\mathcal{S}^2} |\tilde{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})|^2 \cdot |\mathring{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})|^2 \mathrm{d}\vec{s}_\mathrm{u}}{\sqrt{\int_{\mathcal{S}^2} |\tilde{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})|^4 \mathrm{d}\vec{s}_\mathrm{u}} \sqrt{\int_{\mathcal{S}^2} |\mathring{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})|^4 \mathrm{d}\vec{s}_\mathrm{u}}}\right)$$

which compares $\tilde{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})$ to the ground truth $\mathring{\psi}^{(\mathrm{PW})}(\vec{s}_\mathrm{u})$. A value of $\sigma = 1$ indicates high spatial fidelity.

Fig. 4 shows the mean, median and the 50 % confidence interval over all observations of the 300 repetitions. As expected, the space-warping-based translation performs poorly

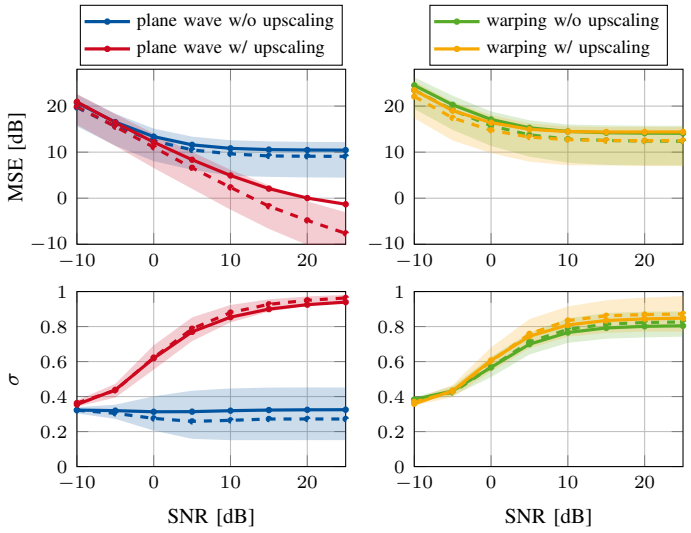


Fig. 4. Results of overall evaluation: mean (—●—), median (- -●- -), and 25 % to 75 % percentile range (▭) of MSE and spatial similarity over all observations of the 300 repetitions of the experiment.

in terms of MSE due to the fact that it neglects the phase. At a sufficiently high SNR, the method yields high spatial fidelity values σ which are only marginally increased when upscaling is used, i.e., this translation method does not benefit from an increased spatial resolution. In contrast, the upscaling shows a significant effect on the plane wave translation both in terms of the MSE and σ. While the vanilla plane wave translation shows deficient performance, the upscaled variant outperforms the space warping in σ and especially in terms of the MSE as it accurately reconstructs the phase. The strong dependency on SNR can be explained with the fact that the sparsity assumptions on the signal are violated for low SNRs.

V. Conclusions

In this paper, we proposed a framework for enhanced sound field translation using sparse recovery of a plane wave decomposition. The underlying idea is to upscale a given HOA signal to a higher truncation order in order to provide spatial information of increased resolution to the downstream translation mechanism. We introduced the mathematical fundamentals, discussed the choice of upscaling order, and investigated the impact of inaccuracies of the source distance estimate on the upscaled signal. In an extensive simulation, we demonstrated the effect of upscaling on different translation methods for different SNRs. The performance of the plane wave translation could be improved significantly.

References

[1] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2002.

[2] T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, May 2002.

[3] J. Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris 6, 2001.

[4] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[5] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, 2005.

[6] F. Schultz and S. Spors, "Data-based binaural synthesis including rotational and translatory head-movements," in *AES 52nd International Conference on Sound Field Control*, 2013.

[7] N. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Elsevier, 2004.

[8] M. Kentgens and P. Jax, "Translation of a higher-order ambisonics sound scene by space warping," in *AES International Conference on Audio for Virtual and Augmented Reality (AVAR)*, 2020.

[9] J. G. Tylka, "Virtual navigation of ambisonics-encoded sound fields containing near-field sources," Ph.D. dissertation, Princeton University, Princeton, NJ, USA, 2019.

[10] A. Allen and W. B. Kleijn, "Ambisonics soundfield navigation using directional decomposition and path distance estimation," in *4th International Conference on Spatial Audio (ICSA)*, Graz, Austria, Sep. 2017, pp. 117–122.

[11] M. Kentgens, A. Behler, and P. Jax, "Translation of a higher order ambisonics sound scene based on parametric decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[12] S. Koyama, S. Shimauchi, and H. Ohmuro, "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 4443–4447.

[13] G. Routray and R. M. Hegde, "Sparsity based framework for spatial sound reproduction in spherical harmonic domain," in *26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018.

[14] P. K. T. Wu, N. Epain, and C. Jin, "A super-resolution beamforming algorithm for spherical microphone arrays using a compressed sensing approach," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, May 2013.

[15] ——, "A dereverberation algorithm for spherical microphone arrays using compressed sensing techniques," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[16] L. Birnie, T. Abhayapala, P. Samarasinghe, and V. Tourbabin, "Sound field translation methods for binaural reproduction," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2019.

[17] A. Wabnitz, N. Epain, A. McEwan, and C. Jin, "Upscaling ambisonic sound scenes using compressed sensing techniques," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011.

[18] A. Wabnitz, N. Epain, and C. T. Jin, "A frequency-domain algorithm to upscale ambisonic sound scenes," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Mar. 2012.

[19] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[20] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *AES 23rd International Conference*, 2003.

[21] M. Kentgens, S. Kühl, C. Antweiler, and P. Jax, "From spatial recording to immersive reproduction - design & implementation of a 3DOF audio-visual VR system," in *AES 145th Convention*, 2018.

[22] J. G. Tylka and E. Choueiri, "Comparison of techniques for binaural navigation of higher-order ambisonic soundfields," in *AES 139th Convention*, 2015.

[23] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.