COMPARING MODELED AND MEASUREMENT-BASED SPHERICAL HARMONIC ENCODING FILTERS FOR SPHERICAL MICROPHONE ARRAYS

Archontis Politis

Dept. of Signal Processing and Acoustics
Aalto University, FI-02150, Finland
archontis.politis@aalto.fi

Hannes Gamper

Microsoft Research
Redmond, WA 98052, USA
hannes.gamper@microsoft.com

ABSTRACT

Spherical microphone array processing is commonly performed in a spatial transform domain, due to theoretical and practical advantages related to sound field capture and beamformer design and control. Multichannel encoding filters are required to implement a discrete spherical harmonic transform and extrapolate the captured sound field coefficients from the array radius to the far field. These spherical harmonic encoding filters can be designed based on a theoretical array model or on measured array responses. Various methods for both design approaches are presented and compared, and differences between modeled and measurement-based filters are investigated. Furthermore, a flexible filter design approach is presented that combines the benefits of previous methods and is suitable for deriving both modeled and measurement-based filters.

Index Terms— Spherical arrays, spherical harmonics, spatial sound, multichannel inversion

1. INTRODUCTION

Spherical microphone arrays (SMAs) allow capturing and analyzing sound fields spatially by evaluating the sound pressure recorded at various locations on the sphere. A convenient way to describe and process a sound field captured by an SMA is in the spherical harmonic domain (SHD) [1, 2, 3, 4, 5, 6, 7, 8]. Several approaches have been proposed previously to encode SMA signals into the SHD, based on modeled and measured SMA responses. Encoding filters aim to recover the spherical harmonic (SH) coefficients of the sound-field pressure, seen as an amplitude distribution of incident plane waves, from the finite radius of the measurement sphere. Two well-known issues occur in the encoding process. Firstly, spatial aliasing occurs due to the discretization of the SH transform (SHT) with a finite number of sensors, resulting in lower spatial frequencies being contaminated by higher ones. Secondly, high spatial frequencies vanish rapidly at low temporal frequencies, making it impossible to capture all spatial coefficients of interest with a frequency-independent performance without excessive amplification of sensor noise. Encoding filters aim to balance between achieving recovery of these sound field coefficients while maintaining microphone noise amplification below some application-dependent acceptable level.

Various approaches have been proposed for the design of the encoding filters. The most straightforward one focuses on the relation between the pressure distribution on the sensing sphere and the sound-field coefficients; ignoring spatial aliasing effects introduced by discrete sampling it reduces to a separable application of a frequency-independent discrete SHT matrix and frequency equalization of the output channels [3, 2, 9, 10, 11]. Alternatively, the whole array response can be modeled in the SHD, including aliasing, and the encoding can be formulated as a multiple-input multiple-output (MIMO) inversion problem [3, 4, 5]. Model-based encoding is convenient since it requires only basic knowledge of the array (microphone positions, sphere radius, baffled or open array). However, it is argued in [3, 5, 7] that better encoding performance can be achieved if directional array response measurements exist and are utilized to derive the filters. This is especially true for arrays that deviate from the model, e.g., due to mismatches of sensor positions or directivities, or cases that are more complex to model, such as dual-radius arrays [5]. Contrary to model-based filters, measurement-based encoding filters can be obtained only via MIMO inversion of the system response, either using directly the measurements [3] or by transforming them first to the SHD [5, 12].

In this work we present a review of the various approaches and evaluate their performance using a real SMA. Order-dependent performance metrics, previously introduced in [3] are revised, along with an additional metric that is order-independent based on the directivity index (DI) [5]. Finally, we introduce a variant of the MIMO inversion approaches proposed in [3, 5] that allows more flexibility on constraining the inversion.

Example code for simulation, encoding and evaluation of these methods can be found in [13].

2. ARRAY MODEL

It is assumed that the sound-field can be modeled as a continuous distribution of incident plane waves with amplitudes $a(\boldsymbol{\gamma})$, with $\boldsymbol{\gamma}$ denoting the direction vector of incidence. The Nth-order spherical harmonic representation of this distribution is given by the SHT as

$$\mathbf{a}_\mathrm{N}(k) = \mathcal{SHT}\left\{a(k,\boldsymbol{\gamma})\right\} = \int_{\boldsymbol{\gamma}} a(k,\boldsymbol{\gamma})\mathbf{y}_\mathrm{N}(\boldsymbol{\gamma})\,\mathrm{d}\boldsymbol{\gamma} \tag{1}$$

where k is the wavenumber, $\mathbf{a}_\mathrm{N} = [a_{00}, ..., a_{nm}, ..., a_{\mathrm{NN}}]^\mathrm{T}$ are the sound-field coefficients, and $\int_{\boldsymbol{\gamma}} \cdot\,\mathrm{d}\boldsymbol{\gamma}$ denotes integration over all directions. The vector $\mathbf{y}_\mathrm{N}$ denotes a vector of $(\mathrm{N}+1)^2$ real SHs Y_{nm} of mode-number (n, m), up to a maximum order N, as

$$\mathbf{y}_\mathrm{N}(\boldsymbol{\gamma}) = [Y_{00}(\boldsymbol{\gamma}), \cdots, Y_{nm}(\boldsymbol{\gamma}), \cdots, Y_{\mathrm{NN}}(\boldsymbol{\gamma})]^\mathrm{T}, \tag{2}$$

where $n = 0, 1, \cdots, \mathrm{N}$ and $m = -n, \cdots, n$ and $(\cdot)^\mathrm{T}$ denotes the transpose. Spherical array processing in the SHD operates on the plane-wave coefficients $\mathbf{a}$.

The pressure coefficients $\mathbf{p}_\mathrm{N} = [p_{00}, \cdots, p_{nm}, \cdots, p_{\mathrm{NN}}]^\mathrm{T}$ due to this sound field, captured at a radius R from the origin, are

$$p_{nm}(kR) = b_n(kR)a_{nm}(k). \tag{3}$$

978-1-5386-1631-4/17/$31.00 ©2017 IEEE

The factors b_n depend on the radius, frequency, and sensor type. Common cases of practical interest are [14, 3]

$$b_n(kR) = \begin{cases} i^n \mathrm{j}_n(kR), & \textit{open} \\ \frac{-i^{n+1}}{(kR)^2 \mathrm{h}_n'(kR)}, & \textit{baffle} \end{cases} \tag{4}$$

where *baffle* refers to pressure sensing on the surface of a spherical rigid baffle of radius R. The functions $\mathrm{j}_n, \mathrm{h}_n$ correspond to spherical Bessel and spherical Hankel functions[1], and $(\cdot)'$ denotes the derivative with respect to the argument. The expansion order N is chosen according to the properties and size of the array [15, 16].

Alternatively, a general array response can be considered. Let us assume S sensors at directions $\mathbf{\Gamma}_\mathrm{S} = [\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_\mathrm{S}]$ and radius R. A matrix of SHs for the set of directions $\mathbf{\Gamma}_\mathrm{S}$ is denoted as $\mathbf{Y}_\mathrm{S} = [\mathbf{y}_\mathrm{N}(\boldsymbol{\gamma}_1), \cdots, \mathbf{y}_\mathrm{N}(\boldsymbol{\gamma}_\mathrm{S})]^\mathrm{T}$. Going forward, the wavenumber k is dropped for convenience of notation. Assuming that the response of the sth sensor is $h_s(\boldsymbol{\gamma})$, the array steering vector is $\mathbf{h}(\boldsymbol{\gamma}) = [h_1(\boldsymbol{\gamma}), ..., h_\mathrm{S}(\boldsymbol{\gamma})]^\mathrm{T}$. The array response can be expressed by its SHT coefficients, limited to order N, and resulting in the $\mathrm{S} \times (\mathrm{N}+1)^2$ matrix $\mathbf{G}_\mathrm{N} = [\mathbf{g}_1^{(\mathrm{N})}, ..., \mathbf{g}_\mathrm{S}^{(\mathrm{N})}]^\mathrm{T}$, with

$$\mathbf{g}_s^{(\mathrm{N})} = \mathcal{SHT}\left\{h_s(\boldsymbol{\gamma})\right\} = \int_{\boldsymbol{\gamma}} h_s(\boldsymbol{\gamma}) \mathbf{y}_\mathrm{N}(\boldsymbol{\gamma}) \, \mathrm{d}\boldsymbol{\gamma}. \tag{5}$$

The steering vector $\mathbf{h}$ is obtained from $\mathbf{G}_\mathrm{N}$ via an inverse SHT as

$$\mathbf{h}(\boldsymbol{\gamma}) = \mathbf{G}_\mathrm{N} \mathbf{y}_\mathrm{N}(\boldsymbol{\gamma}). \tag{6}$$

The relations (5) and (6) are valid for arrays with arbitrary geometry and directional characteristics. For the idealized models of (4), the array matrix $\mathbf{G}_\mathrm{N}$ reduces to

$$\mathbf{G}_\mathrm{N} = \mathbf{Y}_\mathrm{S} \mathbf{B}_\mathrm{N}, \tag{7}$$

where $\mathbf{B}_\mathrm{N} = \mathrm{diag}\left\{[b_0, ..., b_n, ..., b_\mathrm{N}]\right\}$ is the $(\mathrm{N}+1)^2 \times (\mathrm{N}+1)^2$ diagonal matrix of the coefficients in (4).

Following the plane wave distribution formalism $a(\boldsymbol{\gamma})$, and including sensor noise $\mathbf{n}$ in the model, the array signals $\mathbf{x}$ are finally

$$\mathbf{x} = \int_{\boldsymbol{\gamma}} \mathbf{h}(\boldsymbol{\gamma}) a(\boldsymbol{\gamma}) \, \mathrm{d}\boldsymbol{\gamma} + \mathbf{n} = \mathbf{G}_\mathrm{N} \mathbf{a}_\mathrm{N} + \mathbf{n}. \tag{8}$$

Hence, to recover the sound field coefficients $\mathbf{a}$ from the signals $\mathbf{x}$, (8) should be inverted, with constraints on noise amplification. Spatial sampling conditions impose a limit on the maximum order $\mathrm{L} < \mathrm{N}$ of coefficients that can be recovered; if the microphones are close to uniformly distributed then $\mathrm{L} = \lfloor \sqrt{\mathrm{S}} - 1 \rfloor$. The problem can thus be specified as finding an optimal $(\mathrm{L}+1)^2 \times \mathrm{S}$ encoding matrix $\mathbf{E}$ that estimates $\mathbf{a}_\mathrm{L}$ from the microphone signals

$$\tilde{\mathbf{a}}_\mathrm{L} = \mathbf{E}\mathbf{x} = \mathbf{E}\mathbf{G}_\mathrm{N}\mathbf{a}_\mathrm{N} + \mathbf{E}\mathbf{n}. \tag{9}$$

We define the *SH White Noise Gain* (SH-WNG) as the ratio of the noise power at the recovered coefficients over the noise power at the microphones. Assuming spectrally and spatially white noise $\mathbf{n}$ with power σ_n^2 and $\mathcal{E}\left[\mathbf{n}\mathbf{n}^\mathrm{H}\right] = \sigma_n^2 \mathbf{I}_\mathrm{S}$,

$$\mathbf{g}_\mathrm{L} = \mathrm{diag}\left\{\mathcal{E}\left[\mathbf{E}\mathbf{n}\mathbf{n}^\mathrm{H}\mathbf{E}^\mathrm{H}\right]\right\}\left(\sigma_n^2\right)^{-1} = \mathrm{diag}\left\{\mathbf{E}\mathbf{E}^\mathrm{H}\right\}, \tag{10}$$

where $\mathbf{g}_\mathrm{L} = [G_1, \cdots, G_{(L+1)^2}]^\mathrm{T}$ is the vector of SH-WNG values for each SH channel and $\mathcal{E}\left[\cdot\right]$ denotes statistical expectation.

[1] Here, with plane wave directions defined by their angle of incidence rather than propagation, Hankel functions of the second type are used.

3. MODEL-BASED ENCODING FILTERS

3.1. Inversion of radial terms

The most common approach to designing the encoding matrix $\mathbf{E}$ is based on the ideal factorization solution of (3). However, this ignores spatial aliasing occurring at frequencies $f > c\mathrm{L}/(2\pi R)$, or equivalently $kR > \mathrm{L}$, where c denotes the speed of sound. Signals of order n, with $0 \leq n \leq \mathrm{L}$, are equalized as

$$a_{nm} = f(b_n)\mathcal{SHT}\left\{p(\boldsymbol{\gamma})\right\} = f(b_n) p_{nm}, \tag{11}$$

where $f(b_n)$ is a single-channel inversion of the b_n response, being $f(b_n) = 1/b_n$ in the unconstrained case. Hence, this radial inversion separates the encoding process into two stages, a frequency-independent matrix $\mathbf{E}_\mathrm{DSHT}$ performing a discrete SHT (DSHT), followed by $\mathrm{L}+1$ inversion filters applied to the transformed signals of each order [3, 2, 9, 10, 11]. In matrix form

$$\mathbf{E} = f(\mathbf{B}_\mathrm{L})\mathbf{E}_\mathrm{DSHT}. \tag{12}$$

The DSHT matrix can be of the form

$$\mathbf{E}_\mathrm{DSHT} = \begin{cases} \frac{1}{\mathrm{S}}\hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}, & \textit{uniform} \\ \hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}\mathbf{W}_\mathrm{S}, & \textit{quadrature} \\ (\hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}\hat{\mathbf{Y}}_\mathrm{S})^{-1}\hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}, & \textit{least-squares (LS)} \\ (\hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}\mathbf{W}_\mathrm{S}\hat{\mathbf{Y}}_\mathrm{S})^{-1}\hat{\mathbf{Y}}_\mathrm{S}^\mathrm{T}\mathbf{W}_\mathrm{S}, & \textit{weighted LS} \end{cases} \tag{13}$$

where $\hat{\mathbf{Y}}_\mathrm{S}$ is the SH matrix of the sensors up to order L and $\mathbf{W}_\mathrm{S}$ is a diagonal matrix of sampling weights $\sum_{s=1}^{\mathrm{S}} w_s = 1$ that restore orthogonality of the DSHT in the case that the microphones are not uniformly arranged. In the case of uniformity, all cases reduce to the first case. Application of the DSHT matrix already results in a decrease of noise power at the output by a factor of $1/\mathrm{S}$, or equivalently $-10\log_{10}\mathrm{S}$ dB. Therefore, for a SH-WNG constraint of G_max, the maximum gain of the encoding filters should not exceed $\sqrt{\mathrm{S}}G_\mathrm{max}$.

In order to constrain the inversion to such a level, Tikhonov regularisation can be applied, with the filter responses given by

$$f(b_n) = \frac{b_n^*}{|b_n|^2 + \beta^2} \tag{14}$$

where $(\cdot)^*$ denotes the complex conjugate, and β is the regularisation term given as [3]

$$\beta^2 = \frac{1 - \sqrt{1 - (\mathrm{S}G_\mathrm{max}^2)^{-1}}}{1 + \sqrt{1 - (\mathrm{S}G_\mathrm{max}^2)^{-1}}}. \tag{15}$$

The regularisation term forces the inverse filter response to decay rapidly at frequencies below the noise threshold. Alternatively, a maximum threshold for the filter gain can be enforced directly during the inversion [9, 17]

$$f(b_n) = \begin{cases} \frac{1}{b_n}, & \text{for } 1/|b_n| \leq \sqrt{\mathrm{S}}G_\mathrm{max} \\ \frac{b_n^*}{|b_n|}\sqrt{\mathrm{S}}G_\mathrm{max}, & \text{for } 1/|b_n| > \sqrt{\mathrm{S}}G_\mathrm{max} \end{cases} \tag{16}$$

The above limiting approach induces a discontinuity in the response of the filter. It has been shown that it is advantageous to apply a soft-limiting approach as [9]

$$f(b_n) = \frac{2\sqrt{\mathrm{S}}G}{\pi}\frac{b_n^*(f)}{|b_n|}\arctan\frac{\pi}{2\sqrt{\mathrm{S}}G}\frac{1}{|b_n|}. \tag{17}$$

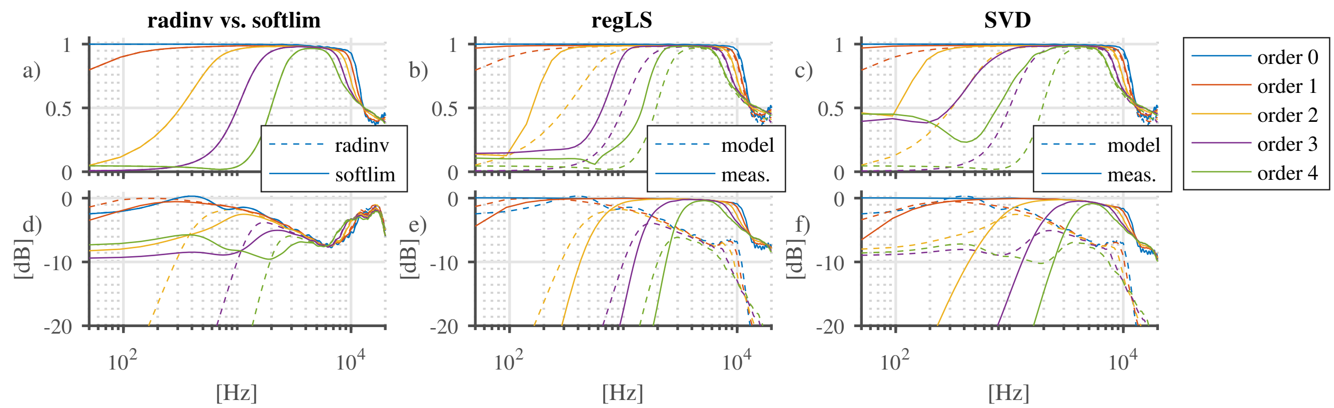

Figure 1: (a–c) Spatial correlation C_n and (d–f) level difference L_n for the Eigenmike and $n = 0, \cdots, 4$; (a,d) model-based **radinv** (14) vs. **softlim** (17), and model-based vs. measurement-based LS-solution for (b, e) **regLS** (19) and (c, f) **SVD** (24).

3.2. Least-squares solution

The single-channel encoding filters in Section 3.1 may result in over-amplification of the aliased components at high frequencies. An alternative is to consider the general array response model of (9) and solve it as a MIMO least-squares inversion that takes into account the effects of spatial aliasing. This approach minimizes

$$\arg\min_{\mathbf{E}} ||\mathbf{E}\mathbf{G}_{\mathrm{N}}\mathbf{a}_{\mathrm{N}} - \mathbf{a}_{\mathrm{L}}||_2^2 \equiv \arg\min_{\mathbf{E}} ||\mathbf{E}\mathbf{G}_{\mathrm{N}} - \mathbf{I}_{\mathrm{LN}}||_F^2 \quad \text{s.t.}\ \max\{\mathbf{g}_{\mathrm{L}}\} < G_{\max} \tag{18}$$

where $\mathbf{I}_{\mathrm{LN}} = [\mathbf{I}_{\mathrm{L}}\mathbf{0}]$ is a $(\mathrm{L}+1)^2 \times (\mathrm{N}+1)^2$ identity matrix padded with zeros that selects SH coefficients up to order L. The $||\cdot||_2$ and $||\cdot||_F$ denote the $L2$ vector norm and Frobenius matrix norm respectively. The regularized least-squares solution to (18) is given by Jin et al. [5] as

$$\mathbf{E} = \mathbf{I}_{\mathrm{LN}}\mathbf{G}_{\mathrm{N}}^{\mathrm{H}}(\mathbf{G}_{\mathrm{N}}\mathbf{G}_{\mathrm{N}}^{\mathrm{H}} + \beta^2\mathbf{I}_{\mathrm{S}})^{-1}. \tag{19}$$

with the regularization value $\beta = 1/(2G_{\max})$.

4. MEASUREMENT-BASED ENCODING FILTERS

Model-based encoding filters are suitable for microphone arrays conforming with the theoretical response of (3) and (4) or (7). However, in practical applications the array may violate the assumptions of ideal behavior due to inaccuracies in sensor positioning, variability in sensor directivity and diaphragm size, etc. In this case the model-based filters may result in suboptimal capture of the SH signals compared to the theoretical array model. By deriving encoding filters from array response measurements the actual array properties can be taken into account. This approach is also useful when the array properties are unknown.

Assuming that the measured responses for D directions $\mathbf{\Gamma}_{\mathrm{D}}$ have been collected in the $\mathrm{S}\times\mathrm{D}$ matrix $\mathbf{H}_{\mathrm{D}} = [\mathbf{h}(\boldsymbol{\gamma}_1), ..., \mathbf{h}(\boldsymbol{\gamma}_{\mathrm{D}})]^{\mathrm{T}}$, and $\mathbf{Y}_{\mathrm{D}}$ is the SH matrix for the same directions up to order L, the weighted least-squares solution to

$$\arg\min_{\mathbf{E}} ||\mathbf{E}\mathbf{H}_{\mathrm{D}} - \mathbf{Y}_{\mathrm{D}}^{\mathrm{T}}||_F^2 \quad \text{s.t.}\ \max\{\mathbf{g}_{\mathrm{L}}\} < G_{\max} \tag{20}$$

is given by Moreau et al. [3]

$$\mathbf{E} = \mathbf{Y}_{\mathrm{D}}^{\mathrm{T}}\mathbf{W}_{\mathrm{D}}\mathbf{H}_{\mathrm{D}}^{\mathrm{H}}(\mathbf{H}_{\mathrm{D}}\mathbf{W}_{\mathrm{D}}\mathbf{H}_{\mathrm{D}}^{\mathrm{H}} + \beta^2\mathbf{I}_{\mathrm{S}})^{-1}. \tag{21}$$

Similar to $\mathbf{W}_{\mathrm{S}}$ in (13), the $\mathrm{D}\times\mathrm{D}$ matrix $\mathbf{W}_{\mathrm{D}}$ contains appropriate sampling weights for the measurement grid directions.

Alternatively, measurement-based filters can be computed with (19) by estimating the SH array matrix $\tilde{\mathbf{G}}$ in the least-squares sense from the array response measurements via:

$$\arg\min_{\tilde{\mathbf{G}}_{\mathrm{N}}} ||\tilde{\mathbf{G}}_{\mathrm{N}}\mathbf{Y}_{\mathrm{D}}^{\mathrm{T}} - \mathbf{H}_{\mathrm{D}}||_F^2 \tag{22}$$

with the solution

$$\tilde{\mathbf{G}}_{\mathrm{N}} = \mathbf{H}_{\mathrm{D}}\mathbf{W}_{\mathrm{D}}\mathbf{Y}_{\mathrm{D}}(\mathbf{Y}_{\mathrm{D}}^{\mathrm{T}}\mathbf{W}_{\mathrm{D}}\mathbf{Y}_{\mathrm{D}})^{-1}. \tag{23}$$

Note that the maximum expansion order N of the array response is limited by the number D and arrangement of measurements. For equiangular measurement grids in azimuth and elevation that order is $N \approx \lceil\sqrt{D}/2 - 1\rceil$ [16]. Since (21) and (19) give equivalent results, we evaluate only the SH domain-based inversion of (19), which applies to both the model- and measurement-based approach.

5. FLEXIBLE LS-INVERSION BASED ON SVD

The approaches based on the inversion of the theoretical radial terms permit flexible single-channel inversion rules suitable for the target responses, such as the soft-limiting approach of (17). However, these filters ignore the effect of spatial aliasing, and they cannot take into account the true measured directional response of the array, if available. The least-squares solutions of (19) and (21) are more general, but they lack flexibility other than not exceeding the desired SH-WNG level, due to the global Tikhonov regularization parameter. It is more practical to control the inversion with respect to the output SH channel response, as in the single-channel radial inversion cases. We propose a simple least-squares solution that can be applied to both modeled and measured responses, while permitting the design flexibility of the single-channel inversion.

Given the least-squares inversion of (19) and the singular value decomposition (SVD) of the SH array matrix, $\mathbf{G}_{\mathrm{N}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$,

$$\begin{aligned}\mathbf{E} &= \mathbf{I}_{\mathrm{LN}}\mathbf{G}_{\mathrm{N}}^{\mathrm{H}}(\mathbf{G}_{\mathrm{N}}\mathbf{G}_{\mathrm{N}}^{\mathrm{H}})^{-1}\\ &= \mathbf{I}_{\mathrm{LN}}\mathbf{V}\mathbf{\Sigma}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}(\mathbf{U}\hat{\mathbf{\Sigma}}^2\mathbf{U}^{\mathrm{H}})^{-1}\\ &= \hat{\mathbf{V}}\mathbf{A}\mathbf{U}^{\mathrm{H}},\end{aligned} \tag{24}$$

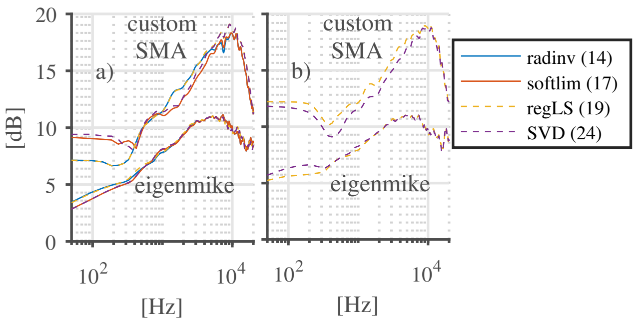

Figure 2: DI of Eigenmike and custom 64-channel SMA, for a) model-based and b) measurement-based filters.

where $\hat{\mathbf{\Sigma}}^2$ is the $\mathrm{S} \times \mathrm{S}$ matrix of squared singular values and $\hat{\mathbf{V}}$ is the $(\mathrm{L}+1)^2 \times \mathrm{S}$ sub-matrix of $\mathbf{V}$ containing the right singular vectors contributing to the recovery of the target coefficients up to order L. $\mathbf{A}$ is a diagonal matrix with entries $[\mathbf{A}]_{ss} = \sigma_s^{-1}$ in the unconstrained case. Constraining the encoding now reduces to constraining the inversion of the singular values in $\mathbf{A}$:

$$\mathbf{E} = \hat{\mathbf{V}} f(\mathbf{A}) \mathbf{U}^{\mathrm{H}}. \tag{25}$$

For example, if it is desired that the soft-limiting approach of (17) is applied, then the modified singular values are given by

$$f([\mathbf{A}]_{ss}) = \frac{2G_{\max}}{\pi} \frac{\sigma_s^*}{|\sigma_s|} \arctan \frac{\pi}{2G_{\max}} \frac{1}{|\sigma_s|}. \tag{26}$$

6. EVALUATION

The performance of the filters recovering the sound field coefficients $\mathbf{a}_{\mathrm{L}}$ can be evaluated via the transformed array response

$$\hat{\mathbf{y}}_{\mathrm{L}}(f, \boldsymbol{\gamma}) = \mathbf{E}(f) \tilde{\mathbf{G}}_{\mathrm{N}}(f) \mathbf{y}_{\mathrm{N}}(\boldsymbol{\gamma}), \tag{27}$$

with $\hat{\mathbf{y}}_{\mathrm{L}} = [\hat{Y}_{00}, \hat{Y}_{1(-1)}, \cdots, \hat{Y}_{\mathrm{LL}}]^{\mathrm{T}}$. Ideally, $\hat{\mathbf{y}}_{\mathrm{L}}$ is frequency-independent and equal to the theoretical SH vector, $\hat{\mathbf{y}}_{\mathrm{L}} \approx \mathbf{y}_{\mathrm{L}}$. A measure of similarity between these two vectors is the normalized spatial correlation per order $C_n \in [0, 1]$ as [3]

$$C_n(f) = \frac{\sum_{m=-n}^{\mathrm{n}} \hat{\mathbf{y}}_{nm}^{\mathrm{H}}(f) \mathbf{W}_{\mathrm{D}} \mathbf{y}_{nm}}{\sqrt{A_n(f) B_n}} \tag{28}$$

where $\hat{\mathbf{y}}_{nm} = [\hat{Y}_{nm}(\boldsymbol{\gamma}_1), \cdots, \hat{Y}_{nm}(\boldsymbol{\gamma}_{\mathrm{D}})]^{\mathrm{T}}$, and $\mathbf{y}_{nm} = [Y_{nm}(\boldsymbol{\gamma}_1), \cdots, Y_{nm}(\boldsymbol{\gamma}_{\mathrm{D}})]^{\mathrm{T}}$ and A_n, B_n are the energies

$$A_n(f) = \sum_{m=-n}^{\mathrm{n}} \hat{\mathbf{y}}_{nm}^{\mathrm{H}}(f) \mathbf{W}_{\mathrm{D}} \hat{\mathbf{y}}_{nm}(f), \tag{29}$$

$$B_n = \sum_{m=-n}^{\mathrm{n}} \mathbf{y}_{nm}^{\mathrm{H}} \mathbf{W}_{\mathrm{D}} \mathbf{y}_{nm} \approx 1. \tag{30}$$

The mean diffuse level difference between the ideal and reconstructed SHs can be defined as [3]

$$L_n(f) = \frac{A_n(f)}{B_n}. \tag{31}$$

Note that for appropriate sampling weights $\mathbf{W}_{\mathrm{D}}$ of the measurement points that approximate integration over the sphere, the ideal level of SHs should be unity, $B_n \approx 1$.

A common order-independent performance metric is the directivity index (DI) of the maximum-directivity beamformer in the SH domain. If the array can recover coefficients up to order L, then the maximum-directivity beamformer weights for any reference direction are given by the vector $\mathbf{y}_{\mathrm{L}}$, with a maximum theoretical DI of $10 \log_{10}(\mathrm{L}+1)^2$ dB. The actual DI is computed as

$$DI(f) = 10 \log_{10} \frac{\mathrm{maxabs}\left[\mathbf{y}_{\mathrm{L}}^{\mathrm{T}} \mathbf{E}(f) \mathbf{H}_{\mathrm{D}}(f)\right]^2}{\mathbf{y}_{\mathrm{L}}^{\mathrm{T}} \mathbf{E}(f) \mathbf{H}_{\mathrm{D}}(f) \mathbf{W}_{\mathrm{D}} \mathbf{H}_{\mathrm{D}}^{\mathrm{H}}(f) \mathbf{E}^{\mathrm{H}}(f) \mathbf{y}_{\mathrm{L}}}, \tag{32}$$

where $\mathrm{maxabs}\,[\cdot]$ denotes the maximum absolute vector element.

The performance of the various approaches is tested on two real SMAs based on directional measurements $\mathbf{H}_{\mathrm{D}}$, and the computation of the SHD array matrix $\tilde{\mathbf{G}}_{\mathrm{N}}$ using (23). A commercial 4th-order, 32-channel SMA (*Eigenmike*) [1, 18] was measured in an anechoic chamber using an equiangular grid of $(5, 10)$ degrees spacing in azimuth and elevation, respectively. A custom 7th-order, 64-channel SMA described in [19] was measured at 512 directions.

Figure 1 shows the spatial correlation and mean level difference between ideal SHs and SHs reconstructed by applying the encoding filters to the measured responses of the Eigenmike array. There are no significant differences between the two model-based radial inversion filters, **radinv** (14) and **softlim** (17), apart from the smoother low-frequency level decay of the **softlim** filters. The model-based LS filters, **regLS** (19), exhibit similar performance as the **radinv** ones, apart from stronger suppression of components above the aliasing limit, an indication that **radinv** and **softlim** over-amplify them. The measurement-based equivalent filters, **regLS** (19) and the proposed **SVD** (25), offer significant improvements over their model-based versions. The spatial correlation is extended significantly towards lower frequencies, while the level of the captured SH coefficients is better equalized, especially towards higher frequencies. In general, measurement-based filters achieve performance much closer to the ideal frequency-independent capture of sound-field coefficients for wider frequency ranges than the theoretical filters indicate. The proposed SVD-based filters are similar to the regularization-based ones, with a somewhat extended performance towards lower frequencies due to the application of a soft-limiting rule rather than a global regularization value.

Figure 2 shows the DI for the two measured SMAs, evaluated with (32). The theoretical maximum DIs are 14 dB for the Eigenmike and 18 dB for the custom 64-channel SMA. While model-based (a,c) and measurement-based (b,d) filters achieve similar performance at mid-to-high frequencies up to the aliasing limit, measurement-based filters yield a consistently higher DI at low frequencies, effectively extending the frequency range.

7. CONCLUSIONS

This study summarizes model- and measurement-based methods for deriving SH encoding filters for SMAs, and evaluates them on two real arrays. The results indicate that least-squares encoding is more general than simple radial inversion. In line with earlier works, measurement-based filters capture the sound field coefficients at a wider frequency range than model-based ones. A least-squares encoding is proposed that allows finer control of the inversion process and improves low-frequency performance compared to prior approaches. Future work will study the method's benefit for complex arrays with more challenging inversion cases.

8. REFERENCES

[1] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 2002, pp. II–1781.

[2] B. Rafaely, “Plane-wave decomposition of the sound field on a sphere by spherical convolution,” *Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, 2004.

[3] S. Moreau, J. Daniel, and S. Bertet, “3D sound field recording with higher order ambisonics - objective measurements and validation of spherical microphone,” in *Proc. 120th Int. Conv. of AES*, Paris, France, 2006.

[4] Z. Li and R. Duraiswami, “Flexible and optimal design of spherical microphone arrays for beamforming,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 702–714, 2007.

[5] C. T. Jin, N. Epain, and A. Parthy, “Design, optimization and evaluation of a dual-radius spherical microphone array,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 193–204, 2014.

[6] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015.

[7] A. Politis, J. Vilkamo, and V. Pulkki, “Sector-based parametric sound field reproduction in the spherical harmonic domain,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015.

[8] D. P. Jarrett, E. A. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. Springer, 2017.

[9] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, “Soft-Limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren,” in *Proc. 37th Deutsche Jahrestagung für Akustik (DAGA)*, 2011, pp. 661–662.

[10] R. Baumgartner, H. Pomberger, and M. Frank, “Practical implementation of radial filters for ambisonic recordings,” in *Proc. Int. Conf. Spatial Audio (ICSA)*, Detmold, Germany, 2011.

[11] S. Lösler and F. Zotter, “Comprehensive radial filter design for practical higher-order ambisonic recording,” in *Fortschritte der Akustik (DAGA)*, 2015, pp. 452–455.

[12] S. Delikaris-Manias, C. A. Valagiannopoulos, and V. Pulkki, “Optimal directional pattern design utilizing arbitrary microphone arrays: A continuous-wave approach,” in *Proc. 134th Int. Conv. AES*, Rome, Italy, 2013.

[13] A. Politis, “Spherical Array Processing library,” http://research.spa.aalto.fi/projects/spharrayproc-lib/spharrayproc.html, 2016.

[14] H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Springer, 2007.

[15] R. A. Kennedy, P. Sadeghi, T. D. Abhayapala, and H. M. Jones, “Intrinsic limits of dimensionality and richness in random multipath fields,” *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2542–2556, 2007.

[16] M. A. Doron and E. Doron, “Wavefield modeling and array processing. I. Spatial sampling,” *IEEE Transactions on signal Processing*, vol. 42, no. 10, pp. 2549–2559, 1994.

[17] J. Meyer and G. Elko, “A qualitative analysis of frequency dependencies in ambisonics decoding related to spherical microphone array recording,” in *Proc. AES Int. Conf. Sound Field Control*, Guildfor, UK, 2016.

[18] mh acoustics, “Eigenmike em32,” https://mhacoustics.com/products#eigenmike1, 2017.

[19] H. Gamper, M. R. P. Thomas, L. Corbin, and I. J. Tashev, “Synthesis of device-independent noise corpora for realistic ASR evaluation,” in *Proc. Interspeech*, San Francisco, CA, USA, 2016.