On the Relation Between DOA-Vector Eigenbeam ESPRIT and Subspace Pseudointensity-Vector

Adrian Herzog
International Audio Laboratories Erlangen[*]
91058 Erlangen, Germany
adrian.herzog@audiolabs-erlangen.de

Emanuël A.P. Habets
International Audio Laboratories Erlangen[*]
91058 Erlangen, Germany
emanuel.habets@audiolabs-erlangen.de

***Abstract*—Eigenbeam ESPRIT (EB-ESPRIT) is a subspace based method to estimate directions-of-arrival (DOAs) of sound sources from a spherical microphone array recording in the spherical harmonics domain (SHD). In recent works, nonsingular EB-ESPRIT methods have been proposed which can estimate the source DOA-vectors without ambiguities. In another recent publication, a subspace based pseudointensity-vector method (SS-PIV) has been proposed for DOA estimation. In this work, we derive the mathematical relation between the DOA-vector EB-ESPRIT and the SS-PIV method. We show that the SS-PIV can be seen as a special case of the DOA-vector EB-ESPRIT. Using this relation, we propose a novel DOA-estimator denoted as extended pseudointensity-vector (PIV). In the evaluation, we compare the DOA-vector EB-ESPRIT and the extended PIV with the SS-PIV and PIV under noisy and reverberant conditions.**

***Index Terms*—Direction-of-arrival estimation, eigenbeam ESPRIT, subspace pseudointensity-vector**

I. Introduction

The direction-of-arrival (DOA) of a sound source captured with a microphone array is an important parameter for informed microphone array processing and acoustic scene analysis.

A commonly used approach is to estimate the DOA from the cross-correlations of different microphone signals [1]. If both sound-pressure and particle-velocity can be measured, one can estimate the source DOA from the sound-intensity vector [2] which, in case of a plane-wave source, is proportional to minus the DOA-vector [3]. Unwanted noise and reverberation can degrade the performance of DOA estimators. Therefore, subspace based methods such as multiple signal classification (MUSIC) [4] and estimation of signal parameters via rotational invariance technique (ESPRIT) [5] have been developed to reduce the effect of noise on the DOA estimation.

DOA estimation algorithms typically depend on the specific microphone array geometry. For 3D scenarios, spherical microphone arrays can be used which enable to efficiently represent the recorded sound-field in the spherical harmonics domain (SHD) [6]. In the SHD, the pseudointensity-vector (PIV) [7], [8] can be used to estimate the source DOA using SHD coefficients up to the first order. Another method splits a B-format signal into real and imaginary parts to estimate two DOAs per time and frequency [9], [10]. The mentioned subspace based methods can also be formulated in the SHD, yielding EB-MUSIC [11] and EB-ESPRIT [12]. In recent works [13]–[16], the problems of the original EB-ESPRIT [12] have been resolved. In [16], [17], the DOA-vectors are estimated by jointly combining three EB-ESPRIT type equations.

Recently, a subspace version of the PIV (SS-PIV) [18] was proposed. However, although higher order SHD coefficients are used to construct the signal subspace, only the zeroth- and first-order coefficients of the signal subspace are used. In contrary, EB-ESPRIT incorporates higher-order coefficients of the signal subspace. Although EB-ESPRIT and SS-PIV are subspace based DOA-estimation methods, the relation between these two approaches has not yet been described.

In this work, we derive the mathematical relation between the DOA-vector EB-ESPRIT [16], [17] and the SS-PIV method [18]. We find, that the DOA-vector EB-ESPRIT reduces to the SS-PIV if only one source is assumed and if only SHD coefficients up to the first order of the signal subspace eigenvector are considered. Therefore, the SS-PIV can be seen as a special case of the DOA-vector EB-ESPRIT. Using this relation and higher SHD orders of the signal, we propose a novel DOA-estimator denoted as extended PIV. In the evaluation, we show that the DOA estimation accuracy of the DOA-vector EB-ESPRIT and the extended PIV increases with increasing maximum SHD order.

II. Spherical Harmonics Transformation

Let $X_1, ..., X_P$ denote microphone signals recorded with a spherical microphone array with P microphones and radius r. The spherical coordinates of the microphones are denoted by $(r, \theta_p, \phi_p) \equiv (r, \Omega_p)$, where $\Omega_p \in \{(\theta, \phi) | \theta \in [0, \pi], \phi \in [-\pi, \pi)\}$ are the microphone directions.

The discrete *spherical harmonics transform* (SHT) of $[X_1, ..., X_p]$ is defined as follows [19], [20]:

$$X_{lm} := \sum_{p=1}^{P} \mathfrak{q}_p X_p Y_{lm}^*(\Omega_p) \ , \tag{1}$$

for $l = 0, ..., L$ and $m = -l, ..., l$, where $(\cdot)^*$ denotes the complex conjugate, $Y_{lm}(\Omega)$ the spherical harmonic of order l and mode m [19], [20], L the maximum SHD order and $\mathfrak{q}_p$ are the sampling weights, which depend on the microphone

[*]A joint institution of the Friedrich Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

distribution [19], [20]. For uniform spatial sampling, one yields $\mathfrak{q}_p = \frac{4\pi}{P}$ [21].

For wavenumbers $\kappa < L/r$ [19] and appropriate sampling weights, the discrete SHT of a unit amplitude plane-wave X_{PW} takes the following form [19], [20]:

$$(X_{\text{PW}})_{lm} \approx b_l(\kappa r) Y_{lm}^*(\Omega_{\text{PW}}) \ , \tag{2}$$

where Ω_{PW} is the DOA of the plane-wave and $b_l(kr)$ the *mode-strength* of order l, which depends on the microphone-array properties only [19], [20]. To compensate for this dependency, one can divide the SHD coefficients by $b_l(kr)$ [22], yielding mode-strength compensated SHD (MC-SHD) signals.

In the following sections, we use a vector notation for spherical harmonics $\mathbf{y}(\Omega) = [Y_{00}(\Omega), Y_{1-1}(\Omega), ..., Y_{LL}(\Omega)]^T$ and MC-SHD signals $\mathbf{x} = [X_{00}, X_{1-1}, ..., X_{LL}]^T$, where $(\cdot)^T$ denotes the transpose. Without loss of generality, we assume that all signals have been transformed to the short-time Fourier transform (STFT) domain before the SHT, where time and frequency indices are omitted for brevity.

III. DOA-Vector EB-ESPRIT

Let $\mathbf{x} = [X_{00}, X_{1-1}, ..., X_{LL}]^T$ denote the signal vector in the MC-SHD, including J plane-wave sources and a diffuse noise component. The power spectral density (PSD) matrix $\mathbf{\Phi_x} := E\{\mathbf{x}\mathbf{x}^H\}$, where $E\{\cdot\}$ denotes the expectation operator and $(\cdot)^H$ the conjugate transpose, can be estimated from $\mathbf{x}$ using temporal and/or spectral averaging. The *signal subspace* is the span of the eigenvectors $[\mathbf{u}_1, ..., \mathbf{u}_J] =: \mathbf{U}_s$ corresponding to the J largest eigenvalues of $\mathbf{\Phi_x}$. For a non-reverberant scenario, $\mathbf{U}_s$ is related to the source DOAs $\Omega_1, ..., \Omega_J$ via

$$\mathbf{U}_s = [\mathbf{y}^*(\Omega_1), ..., \mathbf{y}^*(\Omega_J)]\mathbf{T} \ , \tag{3}$$

where $\mathbf{T}$ is an invertible complex valued $J \times J$ matrix [12].

EB-ESPRIT type methods use this relation and recurrence relations of spherical harmonics to estimate the J source DOAs. The DOA-vector EB-ESPRIT [16], [17] uses the following recurrence relations:

$$\begin{aligned}
\cos(\theta) Y_{lm}^*(\Omega) &= \sqrt{\tfrac{(l-m)(l+m)}{(2l-1)(2l+1)}}\, Y_{(l-1)m}^*(\Omega) \\
&\quad + \sqrt{\tfrac{(l+1-m)(l+1+m)}{(2l+1)(2l+3)}}\, Y_{(l+1)m}^*(\Omega) \\
\sin(\theta) \mathrm{e}^{\mp i\phi} Y_{lm}^*(\Omega) &= \pm \sqrt{\tfrac{(l-1\mp m)(l\mp m)}{(2l-1)(2l+1)}}\, Y_{(l-1)(m\pm 1)}^*(\Omega) \\
&\quad \mp \sqrt{\tfrac{(l+1\pm m)(l+2\pm m)}{(2l+1)(2l+3)}}\, Y_{(l+1)(m\pm 1)}^*(\Omega)
\end{aligned} \tag{4}$$

for $l = 0, ..., L_{\text{max}}$, $|m| \leq l$, $L_{\text{max}} \leq L-1$ and $Y_{lm}(\Omega) := 0$ if $|m| > l$. For the comparison with the SS-PIV method, we vary L_{max} from 0 to $L-1$.

As discussed in [16], [17], using (3) and (4), one can derive the following three EB-ESPRIT equations:

$$(\mathbf{D}_0\mathbf{U}_s)\mathbf{\Psi}_z = \mathbf{D}_z\mathbf{U}_s \quad \text{and} \quad (\mathbf{D}_0\mathbf{U}_s)\mathbf{\Psi}_\pm = \mathbf{D}_\pm\mathbf{U}_s \tag{5}$$

with

$$\begin{aligned}
[\mathbf{D}_z(\cdot)]_{lm} &= \sqrt{(l-m)(l+m)/(2l-1)(2l+1)}\, [(\cdot)]_{(l-1)m} \\
&\quad + \sqrt{(l+1-m)(l+1+m)/(2l+1)(2l+3)}\, [(\cdot)]_{(l+1)m} \\
[\mathbf{D}_\pm(\cdot)]_{lm} &= \pm\sqrt{(l-1\mp m)(l\mp m)/(2l-1)(2l+1)}\, [(\cdot)]_{(l-1)(m\pm 1)} \\
&\quad \mp \sqrt{(l+1\pm m)(l+2\pm m)/(2l+1)(2l+3)}\, [(\cdot)]_{(l+1)(m\pm 1)} \\
\mathbf{D}_0(\cdot) &= [[(\cdot)]_{00}, ..., [(\cdot)]_{L_{\text{max}}L_{\text{max}}}]^T
\end{aligned} \tag{6}$$

for $l = 0, ..., L_{\text{max}}$ and $|m| \leq l$, where $[(\cdot)]_{lm} := 0$ if $|m| > l$ and the $J \times J$ matrices

$$\begin{aligned}
\mathbf{\Psi}_z &= \mathbf{T}^{-1}\text{diag}\{\cos(\theta_1), ..., \cos(\theta_J)\}\mathbf{T} \\
\mathbf{\Psi}_\pm &= \mathbf{T}^{-1}\text{diag}\{\sin(\theta_1)\mathrm{e}^{\mp i\phi_1}, ..., \sin(\theta_J)\mathrm{e}^{\mp i\phi_J}\}\mathbf{T} \ .
\end{aligned} \tag{7}$$

Optimal $\mathbf{\Psi}_z$ and $\mathbf{\Psi}_\pm$ can be determined from (5) using the pseudoinverse of $\mathbf{D}_0\mathbf{U}_s$. Estimates of $\cos(\theta_j)$ and $\sin(\theta_j)\mathrm{e}^{\pm i\phi_j}$ for $j = 1, ..., J$ can be derived by jointly diagonalizing $\mathbf{\Psi}_z$, $\mathbf{\Psi}_-$ and $\mathbf{\Psi}_+$ [17]. The DOA-vector of the j-th source $\mathbf{n}_j = [\sin(\theta_j)\cos(\phi_j), \sin(\theta_j)\sin(\phi_j), \cos(\theta_j)]$ can then be estimated using the resulting eigenvalues via [17]

$$\hat{\mathbf{n}}_j = \text{Re}\left\{ \left[\ \tfrac{1}{2}(\mu_j + \nu_j), \tfrac{1}{2i}(\mu_j - \nu_j), \lambda_j \ \right]^T \right\} \tag{8}$$

for $j = 1, ..., J$, where $\text{Re}\{\cdot\}$ denotes the real operator and λ_j, μ_j, ν_j the j'th eigenvalues of $\mathbf{\Psi}_z$, $\mathbf{\Psi}_-$ and $\mathbf{\Psi}_+$, respectively.

IV. Subspace Pseudointensity Vector

The pseudointensity-vector (PIV) method [7] makes use of the fact that for a plane-wave source the corresponding sound intensity-vector points in the direction opposite to the source DOA. The PIV is an estimate of the intensity-vector using zeroth- and first-order SHD coefficients of the signal [7]. In [18] a modified PIV $\mathbf{m}$ is defined which, for a plane-wave source, points towards the source DOA as follows:

$$\mathbf{m} := \frac{4\pi\sqrt{4\pi}}{3}\text{Re}\left\{ X_{00}^*\mathbf{M} \begin{bmatrix} X_{1-1} \\ X_{10} \\ X_{11} \end{bmatrix} \right\} \ , \tag{9}$$

where the matrix $\mathbf{M}$ is defined as

$$\mathbf{M} := \begin{bmatrix} Y_{1-1}(\Omega_x) & Y_{10}(\Omega_x) & Y_{11}(\Omega_x) \\ Y_{1-1}(\Omega_y) & Y_{10}(\Omega_y) & Y_{11}(\Omega_y) \\ Y_{1-1}(\Omega_z) & Y_{10}(\Omega_z) & Y_{11}(\Omega_z) \end{bmatrix} \tag{10}$$

with $\Omega_x = (\pi/2, 0)$, $\Omega_y = (\pi/2, \pi/2)$ and $\Omega_z = (0, 0)$. Evaluating the spherical harmonics in $\mathbf{M}$ and inserting the result in (9) one yields:

$$\mathbf{m} := \frac{4\pi}{\sqrt{6}}\text{Re}\left\{ X_{00}^* \begin{bmatrix} 1 & 0 & -1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} X_{1-1} \\ X_{10} \\ X_{11} \end{bmatrix} \right\} \ . \tag{11}$$

The subspace PIV (SS-PIV) $\mathbf{m}_{ss}$ is obtained by replacing the coefficients of $\mathbf{x}$ in (11) by the respective coefficients of the dominant PSD matrix eigenvector $\mathbf{u}_s$ as follows [18]:

$$\mathbf{m}_{ss} := \frac{4\pi}{\sqrt{6}}\text{Re}\left\{ [\mathbf{u}_s]_{00}^* \begin{bmatrix} 1 & 0 & -1 \\ -i & 0 & -i \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} [\mathbf{u}_s]_{1-1} \\ [\mathbf{u}_s]_{10} \\ [\mathbf{u}_s]_{11} \end{bmatrix} \right\} \ . \tag{12}$$

The source DOA can be estimated from the direction of $\mathbf{m}_{ss}$.

V. Relation of DOA-Vector EB-ESPRIT to SS-PIV

To compare the DOA-vector EB-ESPRIT to the SS-PIV, we assume that there is only one plane-wave source ($J = 1$) present in the sound field at DOA $\Omega = (\theta, \phi)$. Note that, if several plane-wave sources exist, we can still use $J = 1$ if we assume that only one source is dominant per time-frequency bin (sparseness assumption). In practice, this is not always the case, which can yield wrong DOA estimates at time-frequency bins where more than one source is active. Using the sparseness assumption, (5) simplifies to

$$(\mathbf{D}_0\mathbf{u}_s)\Psi_z = \mathbf{D}_z\mathbf{u}_s \quad \text{and} \quad (\mathbf{D}_0\mathbf{u}_s)\Psi_\pm = \mathbf{D}_\pm\mathbf{u}_s \tag{13}$$

with $\mathbf{D}_0$, $\mathbf{D}_z$ and $\mathbf{D}_\pm$ defined in (6). The complex scalars Ψ_z, Ψ_- and Ψ_+ correspond to $\cos(\theta)$, $\sin(\theta)\mathrm{e}^{i\phi}$ and $\sin(\theta)\mathrm{e}^{-i\phi}$, respectively. The least squares solutions are given by:

$$\Psi_z = \frac{(\mathbf{D}_0\mathbf{u}_s)^H\mathbf{D}_z\mathbf{u}_s}{\|(\mathbf{D}_0\mathbf{u}_s)\|^2} \quad \text{and} \quad \Psi_\pm = \frac{(\mathbf{D}_0\mathbf{u}_s)^H\mathbf{D}_\pm\mathbf{u}_s}{\|(\mathbf{D}_0\mathbf{u}_s)\|^2} \ . \tag{14}$$

The DOA-vector $\hat{\mathbf{n}}$ can be estimated from Ψ_z and $\Psi_\pm$ analogously to (8):

$$\begin{aligned}\hat{\mathbf{n}} &= \mathrm{Re}\left\{\begin{bmatrix}\frac{1}{2}(\Psi_- + \Psi_+)\\ \frac{1}{2i}(\Psi_- - \Psi_+)\\ \Psi_z\end{bmatrix}\right\}\\ &= \frac{1}{\|(\mathbf{D}_0\mathbf{u}_s)\|^2}\mathrm{Re}\left\{\begin{bmatrix}(\mathbf{D}_0\mathbf{u}_s)^H\mathbf{D}_x\mathbf{u}_s\\ (\mathbf{D}_0\mathbf{u}_s)^H\mathbf{D}_y\mathbf{u}_s\\ (\mathbf{D}_0\mathbf{u}_s)^H\mathbf{D}_z\mathbf{u}_s\end{bmatrix}\right\} \ ,\end{aligned} \tag{15}$$

where we defined:

$$\mathbf{D}_x := \tfrac{1}{2}(\mathbf{D}_- + \mathbf{D}_+) \quad \text{and} \quad \mathbf{D}_y := \tfrac{1}{2i}(\mathbf{D}_- - \mathbf{D}_+) \ . \tag{16}$$

Let us now examine the special case $L_{\max} = 0$. Evaluating (6) for $(l, m) = (0, 0)$, we find:

$$\begin{aligned}&(\mathbf{D}_0\mathbf{u}_s) \xrightarrow{L_{\max}=0} [\mathbf{u}_s]_{00} \ , \quad (\mathbf{D}_z\mathbf{u}_s) \xrightarrow{L_{\max}=0} \tfrac{1}{\sqrt{3}}[\mathbf{u}_s]_{10} \ ,\\ &(\mathbf{D}_x\mathbf{u}_s) \xrightarrow{L_{\max}=0} \tfrac{1}{\sqrt{6}}([\mathbf{u}_s]_{1-1} - [\mathbf{u}_s]_{11}) \text{ and}\\ &(\mathbf{D}_y\mathbf{u}_s) \xrightarrow{L_{\max}=0} \tfrac{1}{i\sqrt{6}}([\mathbf{u}_s]_{1-1} + [\mathbf{u}_s]_{11}) \ .\end{aligned} \tag{17}$$

Substituting (17) in (15), one yields the following expression of the DOA-vector estimated using the DOA-vector EB-ESPRIT for $J = 1$ and $L_{\max} = 0$:

$$\hat{\mathbf{n}} = \frac{1}{\sqrt{6}|[\mathbf{u}_s]_{00}|^2}\mathrm{Re}\left\{[\mathbf{u}_s]_{00}^* \begin{bmatrix}1 & 0 & -1\\ -i & 0 & -i\\ 0 & \sqrt{2} & 0\end{bmatrix}\begin{bmatrix}[\mathbf{u}_s]_{1-1}\\ [\mathbf{u}_s]_{10}\\ [\mathbf{u}_s]_{11}\end{bmatrix}\right\} \ . \tag{18}$$

Comparing (18) to (12), one can see that the expressions coincide up to a different normalization factor, which, however, is not relevant for the DOA estimation. Hence, the DOA-vector EB-ESPRIT reduces to the SS-PIV method for $J = 1$ and $L_{\max} = 0$. For $L_{\max} > 0$, the DOA-vector EB-ESPRIT can be seen as an extension of the SS-PIV incorporating higher order coefficients of $\mathbf{u}_s$.

VI. Extended PIV

As discussed in Sec. IV, the SS-PIV is derived from the PIV, defined in (11), by replacing the coefficients of the MC-SHD signal $\mathbf{x}$ by the respective coefficients of the signal subspace eigenvector $\mathbf{u}_s$. Viewing (15) as an extension of the SS-PIV, one can analogously derive an extension of the PIV by replacing $\mathbf{u}_s$ in (15) with $\mathbf{x}$. Choosing a normalization compatible with (11), one can therefore define an *extended PIV*, including higher-order SHD coefficients of $\mathbf{x}$, as follows:

$$\mathbf{m}_{\text{ext}} := 4\pi\ \mathrm{Re}\left\{\begin{bmatrix}(\mathbf{D}_0\mathbf{x})^H\mathbf{D}_x\mathbf{x}\\ (\mathbf{D}_0\mathbf{x})^H\mathbf{D}_y\mathbf{x}\\ (\mathbf{D}_0\mathbf{x})^H\mathbf{D}_z\mathbf{x}\end{bmatrix}\right\} \ . \tag{19}$$

As no PSD matrix estimation and eigen-decomposition is needed, this method has a lower computational complexity and, therefore, is more suitable for real-time implementations.

In the following section, we compare the proposed extended PIV and the DOA-vector EB-ESPRIT to the PIV and SS-PIV methods for different L and $L_{\max}$.

VII. Evaluation

A. Setup

To compare the discussed DOA estimation methods, scenarios with one plane-wave source, diffuse noise and reverberation were simulated. Four speech signals, each 4 seconds long and sampled at 16 kHz, and 24 DOAs, uniformly distributed over the sphere, yielding $4 \times 24 = 96$ configurations. The speech signals were taken from [23].

For the non-reverberant scenario, the source signals were first transformed to the STFT domain and then multiplied with $\mathbf{y}^*(\Omega)$, where Ω denotes the DOA. For the reverberant scenario, a virtual rigid spherical microphone array with 32 almost uniformly distributed microphones and a radius of 7 cm was placed at $[4.103\ \text{m}, 3.471\ \text{m}, 2.912\ \text{m}]$ in a virtual $8\times7\times6$ m^3 shoebox room. The sources were placed at a distance of 2 m from the array center and the spherical microphone array impulse responses (SMIRs) were generated with [24], which is based on the image method [25]. Reverberation times T_{60} of 0.3 s and 0.6 s were used. The source signals were convolved with the SMIRs in the time domain, then transformed to the STFT domain and finally transformed to the MC-SHD using (1) and a regularized mode-strength compensation $X_{lm}(k) \to [b_l^*(kr)/(|b_l(kr)|^2 + 10^{-6})]X_{lm}(k)$.

Diffuse stationary noise with an adjustable input signal plus reverberation-to-noise ratio of

$$\text{iSNR} = 10\log_{10}(\sigma_s^2/\sigma_v^2)\ \text{dB} \tag{20}$$

was added, where σ_v^2 denotes the noise variance and σ_s^2 is the mean energy of the noiseless signal, excluding time-frames with energies less than 1% of the maximum frame-energy.

For the STFT, a square-root-Hann window of 128 samples (8 ms) length, a hopsize of 64 samples (4 ms) and a DFT size of 256 was chosen. For the MC-SHD the maximum SHD order was set to $L = 3$.

The PSD matrix $\mathbf{\Phi}_\mathbf{x}$ was estimated per time-frame and frequency bin within the range $[100, 2313]$ Hz by averaging

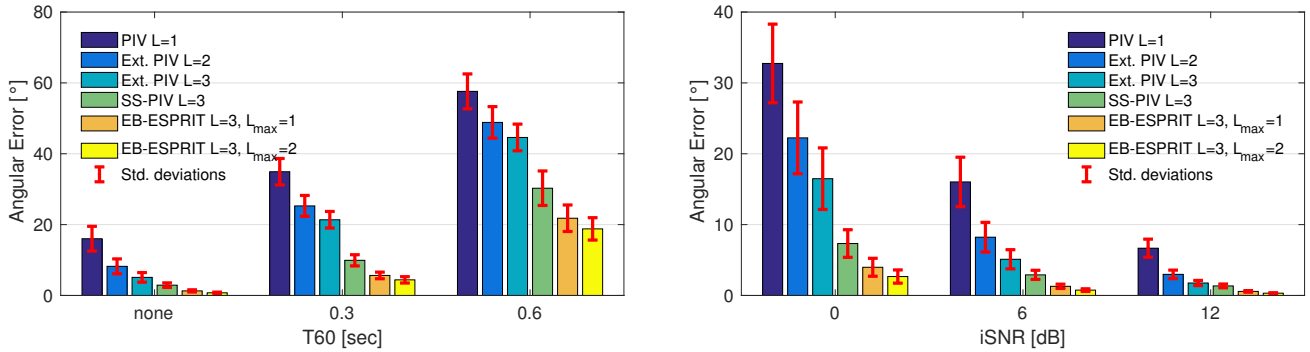

Fig. 1. Mean angular estimation errors. Left: different reverberation and iSNR = 6 dB. Right: different iSNR and no reverberation.

$\mathbf{x}\mathbf{x}^H$ over 25 time-frames (100 ms) centred around the current frame. For the evaluation, angular distances $d_\angle$ of the estimated DOA-vectors/PIVs w.r.t. the true DOA-vector were calculated using:

$$d_\angle(\mathbf{v}_1, \mathbf{v}_2) := \operatorname{acos}\left(\mathbf{v}_1^T \mathbf{v}_2 / (\|\mathbf{v}_1\| \|\mathbf{v}_2\|)\right) \ , \qquad (21)$$

where $\mathbf{v}_1$, $\mathbf{v}_2$ are two real vectors and $\operatorname{acos}(\cdot)$ denotes the inverse cosine. As the DOA estimation at time-frequency indices where only noise is present is expected to yield random results, we neglect time-frequency indices where the signal plus reverberant energy is smaller then -30 dB of the maximum energy for the computation of the mean angular errors in our evaluation. Hence, only the time-frequency indices with

$$|S_{00}^{\text{rev}}(m,k)|^2 > 10^{-3} \cdot \max_{m,k} |S_{00}^{\text{rev}}(m,k)|^2 \qquad (22)$$

are considered, where (m, k) denotes the time-frequency index and $S_{00}^{\text{rev}}(m,k)$ the omnidirectional component of the reverberant signal without noise. The instantaneous signal plus reverberation-to-noise ratios (SNRs) for Fig. 3 are computed as follows:

$$\text{SNR}_{\text{inst.}}(m,k) = 10 \log\left(4\pi |S_{00}^{\text{rev}}(m,k)|^2 / \sigma_v^2\right) \ . \qquad (23)$$

The distributions of the angular errors over instantaneous SNR are derived by averaging all angular errors at the analysed time-frequency bins within 1 dB intervals of the corresponding instantaneous SNRs.

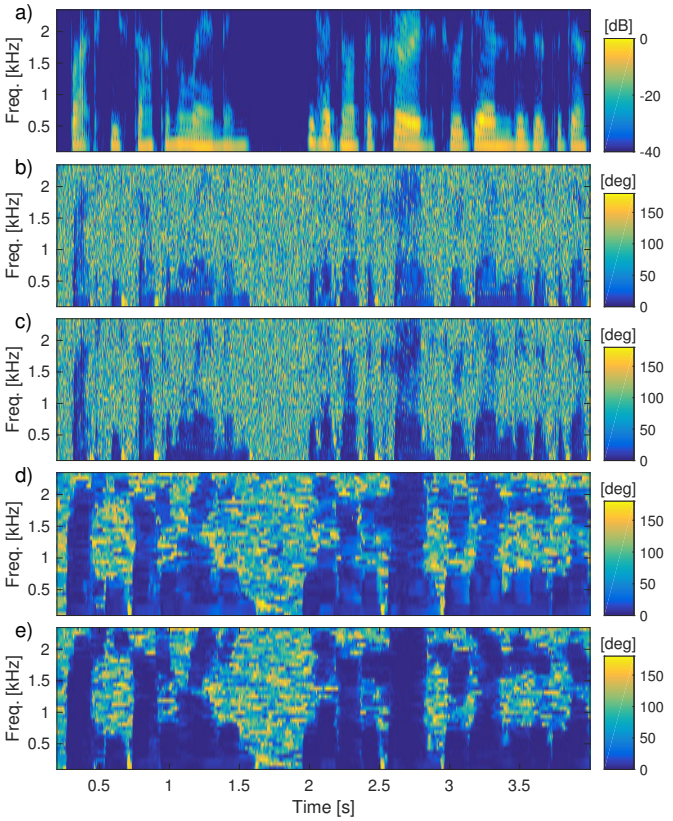

Fig. 2. Angular errors for source at $[98°, 4.2°]$, $T_{60} = 0.3$ s, iSNR = 6 dB. a) Spectrogram of clean speech, b) PIV, c) Extended PIV with $L_{\max} = 2$, d) SS-PIV, e) DOA-vector EB-ESPRIT with $L_{\max} = 2$.

B. Results

In Fig. 1, the mean angular errors for the different DOA estimators are shown for different T_{60} and iSNR. The results were averaged over the 96 configurations, where the standard deviation is represented with red errorbars in Fig. 1. One can see, that the inclusion of higher order SHD coefficients increases the estimation accuracy of both the PIV and the SS-PIV. In particular, even the extension of PIV and SS-PIV by one order ($L_{\max} = 1$), significantly improves the DOA estimation. The standard deviations of the subspace based methods are small for the non-reverberant case, suggesting that the estimation accuracy of these methods does not depend on the specific source signal or DOA unless reverberation is introduced. For the considered maximum SHD order, the subspace methods are significantly more accurate than the other methods.

In Fig. 2, spectrograms of the angular errors are shown. One can see, that the DOA estimation accuracy scales with the source signal energy, shown in Fig. 2 a), at the respective time-frequency bins. The subspace methods yield low estimation errors even at time-frequency bins where the source signal energy is low. Moreover, the extended PIV yields less estimation errors than the PIV at time-frequency bins where the source is less dominant and analogously for the DOA-vector EB-ESPRIT w.r.t. the SS-PIV.

In Fig. 3, the distributions of the angular errors over instantaneous SNRs are shown. The results were averaged over all 96

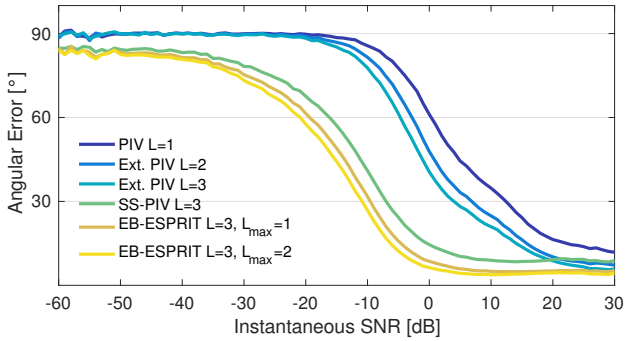

Fig. 3. Distribution of angular errors over instantaneous SNR.

configurations. For low instantaneous SNRs, the non-subspace based methods yield mean angular estimation errors around $90°$, which is expected as the mean angular estimation error of a random estimate is $90°$. For instantaneous SNRs greater than -10 dB the angular errors decrease. For instantaneous SNRs greater than 20 dB the extended PIV methods have lower estimation errors than the SS-PIV method. The subspace based methods yield estimation errors smaller than $90°$ even at low instantaneous SNRs. The accuracy increases until $\sim$ 10 dB and remains constant for higher instantaneous SNRs, suggesting that the noise has negligible influence on the signal subspace for SNRs $>$ 10 dB.

VIII. Conclusion

We showed that the DOA-vector EB-ESPRIT [16], [17] reduces to the SS-PIV method [18] when only one source DOA is estimated and the recurrence formulas involved are only used up to $L_{\max} = 0$. Using this relation between the DOA-vector EB-ESPRIT and the SS-PIV as well as the relation between the PIV and the SS-PIV, we defined an extended PIV (19) which incorporates higher-order SHD coefficients of the signal vector. In the evaluation, we showed that the extended PIV and DOA-vector EB-ESPRIT can estimate the source DOA with higher accuracy than the PIV and SS-PIV, respectively, and the accuracy increases with increasing $L_{\max}$.

References

[1] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[2] A. Nehorai and E. Paldi, "Acoustic vector-sensor array processing," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2481–2491, Sep. 1994.

[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*, ser. Intl. Series in Pure and Applied Physics. New York: McGraw Hill, 1968.

[4] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford University, Stanford, CA., 1981.

[5] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, 1989.

[6] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, May 2002, pp. 1781–1784.

[7] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 442–446.

[8] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation using pseudo-intensity vectors with direct-path dominance test," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Nice, France, Aug. 2015, pp. 2296–2300.

[9] S. Berge and N. Barrett, "A new method for B-format to binaural transcoding," in *Proc. AES 40th Conference*, Tokyo, Japan, Oct. 2010.

[10] O. Thiergart and E. A. P. Habets, "Robust direction-of-arrival estimation of two simultaneous plane waves from a B-format signal," in *Proc. IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Nov. 2012.

[11] D. Khaykin and B. Rafaely, "Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2009, pp. 221–224.

[12] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 5276–5279.

[13] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, Oct. 2011.

[14] B. Jo and J. W. Choi, "Direction of arrival estimation using nonsingular spherical ESPRIT," *J. Acoust. Soc. Am.*, vol. 143, no. 3, pp. EL181–EL187, Mar. 2018.

[15] Q. Huang, L. Zhang, and Y. Fang, "Two-step spherical harmonics ESPRIT-type algorithms and performance analysis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 9, Sep. 2018.

[16] B. Jo and J. W. Choi, "Nonsingular EB-ESPRIT for the localization of early reflections in a room," *J. Acoust. Soc. Am.*, vol. 144, no. 3, Sep. 2018.

[17] A. Herzog and E. A. P. Habets, "Eigenbeam-ESPRIT for DOA-vector estimation," *IEEE Signal Process. Lett.*, 2019.

[18] A. H. Moore, C. Evers, and P. A. Naylor, "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 1, pp. 178–192, 2017.

[19] B. Rafaely, *Fundamentals of Spherical Array Processing*. Springer, 2015, vol. 8.

[20] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. Springer, 2017, vol. 9.

[21] J. Meyer and G. W. Elko, "Spherical microphone arrays for 3d sound recordings," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Norwell, MA, USA: Kluwer Academic Publishers, 2004, pp. 67–89.

[22] B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, Oct. 2004.

[23] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Apr. 2015, pp. 5206–5210.

[24] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.

[25] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.