A LOW COMPLEXITY WEIGHTED LEAST SQUARES NARROWBAND DOA ESTIMATOR FOR ARBITRARY ARRAY GEOMETRIES

Oliver Thiergart, Weilong Huang, and Emanuël A. P. Habets

International Audio Laboratories Erlangen*, Am Wolfsmantel 33, 91058 Erlangen, Germany

ABSTRACT

An increasing number of spatial filtering approaches requires narrowband direction-of-arrival (DOA) estimates. State-of-the-art (SOA) estimators such as root-MUSIC and ESPRIT are computationally complex and can be used only with specific array geometries. In this work, a low complexity DOA estimator is proposed that can be applied to arbitrary array geometries. The DOA is estimated by minimizing the weighted error between the observed and expected inter-microphone phase differences. The complexity of the proposed DOA estimator is significantly lower compared to that of the SOA estimators while providing a similar performance.

Index Terms— Direction-of-arrival estimation, narrowband, microphone arrays

1. INTRODUCTION

Many spatial filtering approaches (e. g., [1]) require narrowband direction-of-arrival (DOA) estimates which are updated for each time and frequency. Well-known narrowband DOA estimators such as ESPRIT [2] and root-MUSIC [3] suffer from two major drawbacks: First, they are computationally too expensive to be carried out for each time and frequency due to the involved eigenvalue decompositions. This is true also for their real-valued formulations unitary ESPRIT [4] and unitary root-MUSIC [5]. Secondly, they can be applied only to specific array geometries. For example, ESPRIT requires a microphone array which can be separated into two identical, rationally invariant subarrays. The original root-MUSIC [3] requires an uniform linear array (ULA). In [6], root MUSIC was derived for nonuniform linear arrays (NLAs), however, the microphones must be located on an equidistant grid.

Computationally less complex DOA estimators, which can be applied to almost any array geometry, usually consider the time difference-of-arrival (TDOA) or phase difference between microphone pairs of the array. The TDOA information is typically used when computing broadband DOA estimates. A recent example is discussed in [7], where the broadband DOAs are estimated by solving a weighted least-squares approach. Phase difference information is typically used to estimate narrowband DOAs. In [8], the phase differences between all pairs of a microphone array are compared to reference phase differences to obtain the desired DOA information. This approach requires no knowledge about the array geometry, but includes a prior array calibration step. In [9, 10], the calibration step is avoided by considering information on the array geometry. The phase difference vectors, which are closely located to the phase difference manifold, are used to estimate a probability density function, from which the DOA can be estimated using a grid search. An approach with lower computational complexity, which directly outputs a DOA estimate, was presented in [11]. Here, the DOA is found with a least-squares approach based on the phase differences between the first and other microphones. Due to the irregular array geometry, the approach can be applied even above the spatial aliasing frequency. However, the approach does not exploit all available phase difference information since not all possible microphone pairs are considered. The approach in [12] is similar to the one in [11] but considers all possible microphone pairs in the least squares formulation. Moreover, the authors propose a normalization of the estimated DOA vector to reduce the impact of errors. With this approach, the DOA can be estimated reliably in three dimensions with an arbitrary 3-dimensional array as long as no microphone pair violates the spatial sampling theorem. Note that neither [11] nor [12] compares the accuracy and computational complexity of the proposed least-squares based DOA estimator to the well-known DOA estimators such as ESPRIT or root-MUSIC.

In this paper, we propose a low-complexity narrowband DOA estimator similar to [12] which estimates the DOA based on the phase differences between all available microphone pairs using a least-squares approach. In contrast to [12], we avoid the normalization of the estimated DOA vector, which enables us to determine the DOA in three dimensions with an arbitrary *2*-dimensional microphone array. Moreover, we formulate a frequency-dependent weighted least-squares problem to exclude microphone pairs from the processing which violate the spatial sampling theorem. Therefore, in contrast to [12], the spatial aliasing frequency is determined by the spacing of the smallest microphone pairs. Throughout the paper we study the computational complexity of the proposed DOA estimator compared to ESPRIT and root-MUSIC and show that a similar estimation accuracy can be achieved.

2. SIGNAL MODEL

We consider the time-frequency domain (frequency index k, time index n) and assume for each (k, n) a single plane wave. The sound is captured with M omnidirectional microphones located in $\mathbf{r}_{1\ldots M}$. The microphone signals are

$$\mathbf{x}(k,n) = \mathbf{x}_{\mathrm{s}}(k,n) + \mathbf{x}_{\mathrm{n}}(k,n), \tag{1}$$

where $\mathbf{x}_{\mathrm{s}}(k,n) = [X_{\mathrm{s},1}(k,n), \ldots, X_{\mathrm{s},M}(k,n)]^{\mathrm{T}}$ are the M microphone signals proportional to the plane wave and $\mathbf{x}_{\mathrm{n}}(k,n)$ models the microphone self-noise. The single-wave model in (1) is often assumed in acoustic DOA estimation approaches and holds even if multiple sources are active at the same time given that the source signals are sufficiently sparse. This assumption normally holds for speech signals in the time-frequency domain [13, 14]. The m-th microphone signal proportional to the plane wave can be written as

$$X_{\mathrm{s},m}(k,n) = \sqrt{\Psi_{\mathrm{s}}(k,n)}\,a(k,\mathbf{n},\mathbf{r}_m)e^{\jmath\phi_{\mathrm{s}}(k,n)}, \tag{2}$$

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

where $\Psi_{\mathrm{s}}(k,n)$ is the power of the wave and $\mathbf{n}(k,n)$ is a unit-norm vector corresponding to the DOA of the wave. The phase of the wave at the origin of the coordinate system is given by $\phi_{\mathrm{s}}(k,n)$. The function

$$a(k,\mathbf{n},\mathbf{r}) = e^{\jmath\kappa(k)\mathbf{r}^{\mathrm{T}}\mathbf{n}} \tag{3}$$

describes the phase shift of the wave along the displacement vector $\mathbf{r}$. Here, $\kappa(k)$ is the wavenumber for the given frequency index k. When considering a 3-dimensional coordinate system, the DOA vector $\mathbf{n}(k,n)$ can be expressed as

$$\mathbf{n}(k,n) = [\cos(\varphi)\cos(\vartheta),\ \sin(\varphi)\cos(\vartheta),\ \sin(\vartheta)]^{\mathrm{T}}, \tag{4}$$

where $\varphi(k,n)$ and $\vartheta(k,n)$ are the azimuth and elevation, respectively, of the wave. For a 2-dimensional coordinate system where the sound propagates in the horizontal plane, we have

$$\mathbf{n}(k,n) = [\cos(\varphi),\sin(\varphi)]^{T}. \tag{5}$$

Since the different components in (1) are assumed to be mutually uncorrelated, we can write the power spectral density (PSD) matrix of the microphone signals as

$$\boldsymbol{\Phi}_x(k,n) = \mathrm{E}\left\{\mathbf{x}(k,n)\mathbf{x}^{\mathrm{H}}(k,n)\right\} \tag{6a}$$

$$= \boldsymbol{\Phi}_{\mathrm{s}}(k,n) + \boldsymbol{\Phi}_{\mathrm{n}}(k), \tag{6b}$$

where $\boldsymbol{\Phi}_{\mathrm{s}}(k,n)$ and $\boldsymbol{\Phi}_{\mathrm{n}}(k)$ are the PSD matrix of the plane wave and noise, respectively. The $(m'm)$-th element of $\boldsymbol{\Phi}_{\mathrm{s}}(k,n)$ is the cross PSDs of the captured plane wave between the microphones m and m'. It can be written as

$$\Phi_{\mathrm{s},m'm}(k,n) = \mathrm{E}\left\{X_{\mathrm{s},m'}(k,n)X^{*}_{\mathrm{s},m}(k,n)\right\} \tag{7a}$$

$$= \Psi_{\mathrm{s}}(k,n)a(k,\mathbf{n},\mathbf{r}_{mm'}), \tag{7b}$$

where $\mathbf{r}_{mm'} = \mathbf{r}_{m'} - \mathbf{r}_{m}$. The noise of the different microphones is assumed to be independent and identically distributed (iid). Hence, the noise PSD matrix can be written as

$$\boldsymbol{\Phi}_{\mathrm{n}}(k) = \Phi_{\mathrm{n}}(k)\mathbf{I}_M, \tag{8}$$

where $\Phi_{\mathrm{n}}(k)$ is the noise power and $\mathbf{I}_M$ is the $M\times M$ identity matrix. Given this assumption, the off-diagonal elements of $\boldsymbol{\Phi}_x(k,n)$ are equal to the cross PSDs of the direct sound, i. e.,

$$\Phi_{x,m'm}(k,n) = \Phi_{\mathrm{s},m'm}(k,n), \quad m' \neq m. \tag{9}$$

The aim of the paper is to estimate the DOA vector $\mathbf{n}(k,n)$ (or φ and ϑ) from the noisy microphone signals $\mathbf{x}(k,n)$.

3. WEIGHTED LEAST SQUARES DOA ESTIMATOR

We assume a single narrowband plane wave and iid noise, i. e., the microphone input PSD matrix $\boldsymbol{\Phi}_x(k,n)$ is given in (6). From (7b) and (3) we can see that the phase of $\Phi_{\mathrm{s},m'm}(k,n)$ contains information on the DOA of the direct sound, i. e.,

$$\angle\Phi_{\mathrm{s},m'm}(k,n) = \kappa(k)\mathbf{r}^{\mathrm{T}}_{m'm}\mathbf{n}(k,n), \tag{10}$$

where $\mathbf{r}_{m'm} = \mathbf{r}_m - \mathbf{r}_{m'}$. Due to (9) the DOA can be determined from the input cross PSDs $\Phi_{x,m'm}(k,n)$. For this purpose, we collect all cross PSD $\Phi_{x,m'm}(k,n)$ of the upper triangle[1] of $\boldsymbol{\Phi}_x(k,n)$ in the vector

$$\boldsymbol{\phi}_x(k,n) = [\Phi_{x,12},\ldots,\Phi_{x,ij},\ldots,\Phi_{x,NM}]^{\mathrm{T}} \quad \forall i,j, \tag{11}$$

[1] Clearly, the lower triangle does not provide additional information as the matrix is Hermitian.

where $N = M-1$, $1\leq i\leq N$, and $i<j\leq M$. In practice, we select the elements of $\boldsymbol{\phi}_x(k,n)$ from the input PSD matrix $\boldsymbol{\Phi}_x(k,n)$, which can be estimated with (6a) when approximating the expectation by a temporal averaging. In this case, the vector in (11) is denoted by $\widehat{\boldsymbol{\phi}}_x(k,n)$. Since the estimated PSDs contain estimation errors, we determine the DOA using a least squares (LS) approach. Using (9) and (10), the phases of the estimated PSDs in $\widehat{\boldsymbol{\phi}}_x(k,n)$ can be written as

$$\widehat{\boldsymbol{\mu}}_{\mathrm{s}}(k,n) = \angle\widehat{\boldsymbol{\phi}}_x(k,n) \tag{12a}$$

$$= \mathbf{Q}(k)\mathbf{n}(k,n) + \boldsymbol{\Delta}(k,n), \tag{12b}$$

where

$$\mathbf{Q}(k) = \kappa(k)[\mathbf{r}_{12},\ldots,\mathbf{r}_{ij},\ldots,\mathbf{r}_{NM}]^{\mathrm{T}}, \tag{13}$$

is an $B\times D$ matrix $[B = \frac{M}{2}(M-1)]$ which describes the array geometry. Here, D is the number of dimensions of the considered Cartesian coordinate system, i. e., the number of elements of $\mathbf{n}(k,n)$ and $\mathbf{r}$, respectively. The vector $\boldsymbol{\Delta}(k,n)$ contains the errors of the elements of $\widehat{\boldsymbol{\mu}}_{\mathrm{s}}(k,n)$ due to estimation of the input PSD matrix. In the following, we consider a weighted least squares (WLS) approach, which minimizes the weighted ℓ^2-norm of $\boldsymbol{\Delta}(k,n)$, to find an estimate of $\mathbf{n}(k,n)$, i. e.,

$$\widehat{\mathbf{n}}(k,n) = \arg\min_{\mathbf{n}} \mathrm{tr}\left\{\boldsymbol{\Delta}^{\mathrm{T}}(k,n)\mathbf{W}(k,n)\boldsymbol{\Delta}(k,n)\right\}, \tag{14}$$

where $\mathbf{W}(k,n) = \mathrm{diag}\left\{W_{11}(k,n), W_{22}(k,n),\ldots,W_{BB}(k,n)\right\}$ is a time and frequency dependent $B\times B$ diagonal matrix containing the weights for each element of $\boldsymbol{\Delta}$ that are discussed later. The LS solution is given by

$$\widehat{\mathbf{n}}(k,n) = [\mathbf{B}(k,n)\mathbf{Q}(k)]^{-1}\,\mathbf{B}(k,n)\widehat{\boldsymbol{\mu}}_{\mathrm{s}}(k,n). \tag{15}$$

where $\mathbf{B}(k,n) = \mathbf{Q}^{\mathrm{T}}(k)\mathbf{W}(k,n)$. Clearly, for computing $\widehat{\mathbf{n}}(k,n)$ the matrix $\mathbf{Q}^{\mathrm{T}}(k)\mathbf{W}(k,n)\mathbf{Q}(k)$ must have full rank. This requires that the array microphones span a D-dimensional coordinate system. In other words, estimating $\mathbf{n}(k,n)$ in D dimensions requires a D-dimensional microphone array, which represents the only restriction on the microphone configuration. For example, when a 3-dimensional microphone array is used, we can estimate the 3-dimensional DOA vector $\mathbf{n}(k,n)$ in (4). This allows us to compute both the azimuth $\varphi(k,n)$ and elevation $\vartheta(k,n)$ of the DOA. With a planar microphone array ($D=2$), we can estimate the first two dimensions of $\mathbf{n}(k,n)$ in (4). In this case, we can compute the azimuth $\varphi(k,n)$ as well as the absolute value of the elevation $\vartheta(k,n)$, i. e., an up-down ambiguity remains. Note that the approach in [12] can estimate only $\varphi(k,n)$ when using a planar array ($D=2$) as the authors normalize the estimated vector $\widehat{\mathbf{n}}(k,n)$. If a linear array is used ($D=1$), the matrix $\mathbf{Q}(k)$ becomes a vector, i. e.,

$$\mathbf{q}(k) = \kappa(k)\ [r_{12},\ldots,r_{ij},\ldots,r_{NM}]^{\mathrm{T}}, \tag{16}$$

where r_{ij} is the spacing between microphone i and j. In this case, we can estimate the first dimension of the DOA vector $\mathbf{n}(k,n)$. Assuming that the microphone array defines the x-axis of the coordinate system and that the sources are located in the horizontal plane, then the first dimension of $\mathbf{n}(k,n)$ is the cosine of the azimuth of the DOA in (5). Thus,

$$\cos\widehat{\varphi}(k,n) = \mathbf{v}^{\mathrm{T}}(k,n)\,\widehat{\boldsymbol{\mu}}_{\mathrm{s}}(k), \tag{17}$$

where $\mathbf{v}(k,n) = \mathbf{W}(k,n)\mathbf{q}(k)[\mathbf{q}^{\mathrm{T}}(k)\mathbf{W}(k,n)\mathbf{q}(k)]^{-1}$. In this case, we can compute $\widehat{\varphi}(k,n)$ in the range $[0,\pi]$, i. e., we cannot distinguish if the sound arrives from the front or back of the array.

Typically, the weights $\mathbf{W}(k,n)$ are chosen depending on the error variance [15], in our case, the variance of the elements of the vector $\boldsymbol{\Delta}(k,n)$. These variances depend on the signal-to-noise ratio (SNR), the frequency, the microphones spacing r_{ij}, as well as on the DOA of the plane wave. However, since the latter is unknown, we assume equal error variances in this work, which would lead to unit weights, i. e., $\mathbf{W}(k,n) = \mathbf{I}_B\ \forall n$. Instead of compensating for different error variances, the weights can also be used to exclude outliers. This is necessary because the LS approach is known to be very sensitive to outliers. In our case, outliers can appear when a microphone pair is used above the spatial aliasing frequency, i. e., when using a cross PSD $\widehat{\Phi}_{x,ij}(k,n)$ for which $\|\kappa(k)\mathbf{r}_{ij}\| \geq \pi$. To exclude these outliers, we set the corresponding weights to zero. This means, the weights are given by

$$W_{bb}(k) = \begin{cases} 1, & q_b < \pi \\ 0, & \text{otherwise}, \end{cases} \tag{18}$$

where $b \in [1, B]$ and q_b is the ℓ^2-norm of the b-th row of $\mathbf{Q}$. Note that the b-th microphone pair ij can be excluded (i. e., the corresponding weight can be set to zero) if the resulting matrix $\mathbf{Q}^{\mathrm{T}}(k)\mathbf{W}(k)\mathbf{Q}(k)$ remains full rank. Otherwise, the inverse in (15) cannot be computed. When defining the elements of $\mathbf{W}(k,n)$ time-invariant as in (18), we can compute $\mathbf{v}(k)$ in (17) in advance for the given microphone array. This yields a DOA estimator with very low computational complexity.

4. COMPUTATIONAL COMPLEXITY

This section discusses the computational complexity of the DOA estimator proposed in Sec. 3 and two state-of-the-art (SOA) DOA estimators, namely estimation of signal parameters via rotational invariance techniques (ESPRIT) and root-MUSIC. For this purpose, we assume a linear microphone array with M microphones. All considered DOA estimators require an estimate of the input PSD matrix $\boldsymbol{\Phi}_x(k,n)$ and provide an estimate of the cosine of the DOA $\varphi(k,n)$.

To estimate the DOA of a single plane wave with ESPRIT, the estimated $\boldsymbol{\Phi}_x(k,n)$ needs to be decomposed using an eigenvalue decomposition (EVD). The eigenvector corresponding to the largest eigenvalue represents the so-called signal subspace and the DOA is estimated by considering the phase differences between signal subspace elements [2]. The EVD is the most expensive step in ESPRIT resulting in the computational complexity of $\mathcal{O}(M^3)$ [16].

The estimation of the DOA of a single plane wave with root-MUSIC also requires an EVD of $\boldsymbol{\Phi}_x(k,n)$. The eigenvectors of $\boldsymbol{\Phi}_x(k,n)$ corresponding to the $M-1$ smallest eigenvalues represent the so-called noise subspace $\mathbf{Q}(k,n)$. For the noise subspace $\mathbf{Q}(k,n)$ we have

$$\mathbf{a}^{\mathrm{H}}(k,n)\mathbf{Q}(k,n)\mathbf{Q}^{\mathrm{H}}(k,n)\mathbf{a}(k,n) = 0, \tag{19}$$

where $\mathbf{a}(k,n)$ is the propagation vector for the plane wave, which depends on the DOA of the wave. The m-th element of $\mathbf{a}(k,n)$ is given by (3). The original root-MUSIC [3] is applied to ULAs. For such an array, we have $\mathbf{a} = \left[1, z^1, z^2, \ldots, z^{M-1}\right]^{\mathrm{T}}$, where $z = e^{-j\kappa(k)r\cos\varphi(k,n)}$ with r being the inter-microphone spacing. From this we can see that the left-hand side of (19) represents a complex polynomial of order $P = 2(M-1)$. Estimating the DOA requires to compute the roots of the polynomial [3], for which the complexity is $\Theta(P^2\log(P))$ [17]. Note that root-MUSIC can also be applied to a NLA array where the microphones are located on an equidistant grid with spacing r [6]. In this case, the propagation vector becomes $\mathbf{a} = \left[1, \ldots, z^{M'-1}\right]^{\mathrm{T}}$, where $M' = l/r$. Here, l is the array size and M' is the number of grid points used by the NLA. The order of the polynomial is now given by $P = 2(M'-1)$, which can become large when the inter-microphone spacings of the NLA are very different. Thus, when r is much smaller compared to l, the computation complexity of the root finding for a NLA can become significantly larger than the computational complexity for a ULA.

Approach	Complexity
ESPRIT	$\mathcal{O}(M^3)$
Root-MUSIC (ULA)	$\mathcal{O}(M^3) + \Theta(P^2\log(P))$, $P = 2(M-1)$
Root-MUSIC (NLA[1])	$\mathcal{O}(M^3) + \Theta(P^2\log(P))$, $P = 2(l/r-1)$
Proposed WLS-based	$\mathcal{O}(M)$

Table 1. Computational complexity of the studied DOA estimators ([1]where all microphones lie on an equidistant grid with spacing r)

The WLS-based DOA estimator proposed in Sec. 3 uses the off-diagonal elements of $\boldsymbol{\Phi}_x(k,n)$, i. e., the cross PSDs $\Phi_{x,m'm}(k,n)$, in (12). The cosine of the DOA $\varphi(k,n)$ is then estimated using (17). The dot product of the two vectors in (17) results in a computational complexity of $\mathcal{O}(M)$. Since the vector $\mathbf{v}(k)$ can be computed in advance as explained in Sec. 3, $\mathcal{O}(M)$ is the computational complexity for the WLS-based estimator, which is significantly smaller than the computational complexity of ESPRIT or root-MUSIC. Note that the computational complexity can be further reduced by taking into account the fact that for some frequencies $\mathbf{v}(k) = 0$.

Table 1 summarizes the computational complexity of the different DOA estimators. For ESPRIT and root-MUSIC, the specified values represents the lower bound as only the complexity of the EVD and root finding, respectively, are taken into account.

5. SIMULATION RESULTS

We have carried out simulations to study the performance of the proposed DOA estimator for a linear array. For this purpose, we have simulated a single plane wave with specific frequency, DOA, and random phase using (2). We were considering different microphone array geometries which are explained later. The microphone signals were computed with (1) where the noise $\mathbf{x}_{\mathrm{n}}(k,n)$ was modeled as spatially white Gaussian noise with specific SNR. From the microphone signals $\mathbf{x}(k,n)$ we have estimated the input PSD $\boldsymbol{\Phi}_x(k,n)$ with (6a), where the expectation was approximated by a temporal averaging over 10 realizations of the microphone signals. This corresponds to a typical averaging length in practice. Finally, the DOA $\varphi(k,n)$ of the plane wave was estimated from $\boldsymbol{\Phi}_x(k,n)$ with (depending on the array geometry)

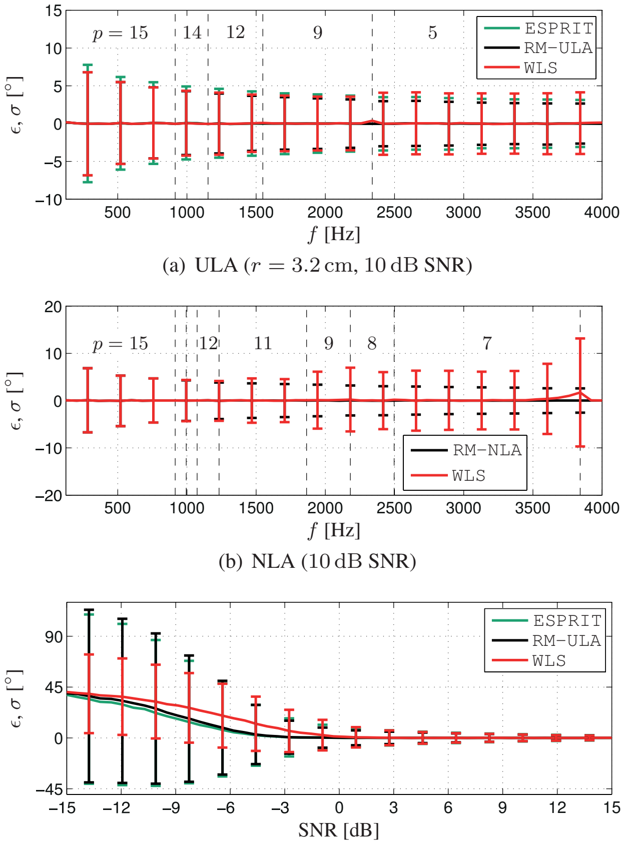
- `ESPRIT`: ESPRIT [2] (subarrays with maximum overlap)
- `RM-ULA`: originally proposed Root-MUSIC for ULAs [3]
- `RM-NLA`: Root-MUSIC for NLAs proposed in [6]
- `WLS`: WLS estimator proposed in Sec. 3 computed with (17).

To evaluate the performance of the different DOA estimators, we are considering the mean error $\epsilon(k,n)$ and error variance $\sigma^2(k,n)$, which are computed as

$$\epsilon(k,n) = \mathrm{E}\left\{\widehat{\varphi}(k,n) - \varphi(k,n)\right\}, \tag{20a}$$

$$\sigma^2(k,n) = \mathrm{E}\left\{\left(\widehat{\varphi}(k,n) - \varphi(k,n) - \epsilon(k,n)\right)^2\right\}. \tag{20b}$$

The expectation was approximated by averaging over 20000 realizations of the experiment.



(a) ULA ($r = 3.2$ cm, 10 dB SNR)

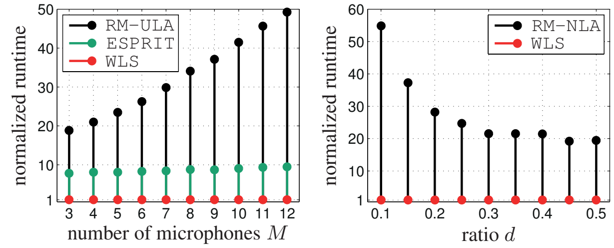
(b) NLA (10 dB SNR)

(c) ULA ($r = 3.2$ cm, $f = 1.7$ kHz)

Fig. 1. DOA estimation performance for a linear array with $M = 6$ microphones

The mean error $\epsilon(k, n)$ and standard deviation $\sigma(k, n)$ are depicted in Fig. 1(a) as a function of frequency. We were considering a ULA with $M = 6$ microphones and microphone spacing $r = 3.2\,\text{cm}$. The noise power was chosen such that the SNR was indirectly proportional to the frequency with 10 dB SNR at $f = 1\,\text{kHz}$. The DOA of the plane wave was chosen randomly in the range $\varphi(k, n) \in [0, 90°]$ for each realization (uniform distribution). The DOA was estimated with `RM-ULA`, `ESPRIT`, and `WLS`. The number $p(k) = \text{tr}\,\{\mathbf{W}(k)\}$ depicted in the plot indicates how many microphone pairs were considered effectively by WLS for the given frequency range. The results in Fig. 1(a) show that all DOA estimators provided almost unbiased results when averaging the error over random DOAs. For `RM-ULA` and `ESPRIT` the error variance became slightly smaller for higher frequencies. At lower frequencies, `WLS` yielded the same variance as `RM-ULA` and a smaller variance than `ESPRIT`. For higher frequencies, however, `WLS` yielded the highest estimation variance. The reason is that at higher frequencies, the number p of microphone pairs considered by `WLS` for estimating the DOAs decreases to avoid spatial aliasing.

Figure 1(b) shows the same results as Fig. 1(a) but for an NLA where the $M = 6$ microphones were located at the grid points $[0, 1, 2, 4, 8, 16] \times 1\,\text{cm}$. We can see the same trend as in Fig. 1(a), i. e., the results were unbiased and the estimation variance of the proposed `WLS` estimator was comparatively high at higher frequencies. As in Fig. 1(a), the reason for the higher variance is the reduced number of considered microphone pairs at higher frequencies.



(a) ULA with $r = 3.2$ cm

(b) NLA with $M = 3$

Fig. 2. Runtime of the different DOA estimators measured in MATLAB (values normalized w. r. t. the runtime of `WLS`)

The performance of the different DOA estimators is studied in Fig. 1(c) for different SNRs. We were considering a ULA with $M = 6$ microphones with spacing $r = 3.2\,\text{cm}$, i. e., the same array as in Fig. 1(a). The DOA of the plane wave was chosen randomly as in the experiment before. The frequency was $f = 1.7\,\text{kHz}$ for which the `WLS` estimator was effectively using $p = 9$ microphone pairs (out of $B = 15$ possible pairs). We can see that for all DOA estimators, the mean error $\epsilon(k, n)$ was increasing towards low SNRs. The `WLS` estimator was slightly less accurate than `ESPRIT` and `RM-ULA` at low SNRs. In terms of estimation variance, the proposed `WLS` estimator performed similar compared to the other estimators for medium and higher SNRs.

Finally, Fig. 2(a) and 2(b) visualize the computational complexity of the different DOA estimators. The plots show the runtime of the estimators in MATLAB. All runtimes were normalized w. r. t. the runtime of the `WLS` estimator. Note that the runtime was measured after estimating the signal PSD matrix $\mathbf{\Phi}_x(k, n)$ (which is used by all studied DOA estimators) until $\cos\widehat{\varphi}(k, n)$ was found. Figure 2(a) shows the runtime for a ULA as a function of the microphone number M. The microphone spacing was $r = 3.2\,\text{cm}$ and we were considering the frequency $f = 500\,\text{Hz}$ where the `WLS` estimator made use of all microphone pairs. The plot shows that both `ESPRIT` and `RM-ULA` were significantly more complex than the proposed `WLS` estimator. For example for $M = 12$, the runtime of `ESPRIT` was 10 times larger compared to `WLS`, while the runtime for `RM-ULA` was 50 times as large. Figure 2(b) shows the runtimes for an NLA with $M = 3$ microphones. Here, the position of the center microphone was varied and the x-axis of the plot shows the ratio d between the smallest microphone spacing an the array size. For $d = 0.5$, we obtain a ULA. For $d < 0.5$ we have an NLA. In this case, the effective number of microphones considered by `RM-NLA` significantly increases leading to a higher runtime. In contrast, for `WLS` the runtime is independent of d.

6. CONCLUSIONS

We have proposed a narrowband DOA estimator with significantly lower computational complexity compared to SOA approaches such as ESPRIT or root-MUSIC. The estimator can be applied to almost any array geometry and determines the DOA based on the phase differences between the available microphone pairs using a weighted least-squares approach. The spatial aliasing frequency is determined by the smallest microphone pairs. At low and medium frequency, the proposed estimator provides a similar performance as the SOA approaches. At higher frequencies, the estimation variance of the proposed approach is comparatively higher since specific microphone pairs are excluded from the estimation to avoid spatial aliasing.

7. REFERENCES

[1] O. Thiergart, M. Taseska, and E. A. P. Habets, "An informed parametric spatial filter based on instantaneous direction-of-arrival estimates," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 12, pp. 2182–2196, December 2014.

[2] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, July 1989.

[3] B. Rao and K. Hari, "Performance analysis of Root-MUSIC," in *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*, vol. 2, 1988, pp. 578–582.

[4] M. Haardt and J. Nossek, "Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden," *Signal Processing, IEEE Transactions on*, vol. 43, no. 5, pp. 1232–1242, May 1995.

[5] M. Zoltowski, G. Kautz, and S. Silverstein, "Beamspace Root-MUSIC," *Signal Processing, IEEE Transactions on*, vol. 41, no. 1, pp. 344–364, January 1993.

[6] A. Mhamdi and A. Samet, "Direction of arrival estimation for nonuniform linear antenna," in *Communications, Computing and Control Applications (CCCA), 2011 International Conference on*, March 2011, pp. 1–5.

[7] D. Ying, J. Li, Y. Feng, and Y. Yan, "Direction of arrival estimation based on weighted minimum mean square error," in *Signal and Information Processing (ChinaSIP), 2013 IEEE China Summit International Conference on*, July 2013, pp. 318–321.

[8] Y.-W. Wu, S. Rhodes, and E. Satorius, "Direction of arrival estimation via extended phase interferometry," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 31, no. 1, pp. 375–381, January 1995.

[9] K. Fujimoto, N. Ding, and N. Hamada, "Multiple sources' direction finding by using reliable component on phase difference manifold and kernel density estimator," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 2601–2604.

[10] A. Kijima, Y. Mitsukura, and N. Hamada, "DOA estimation by 3-D microphone array in the presence of spatial aliasing," in *Systems, Process Control (ICSPC), 2013 IEEE Conference on*, December 2013, pp. 209–213.

[11] A. Im, M. Gialich, Z. Aliyazicioglu, and H. K. Hwang, "DOA estimation via phase measurement," in *Progress in Electromagnetics Research Symposium (PIERS) Proceedings*, Taipei, Taiwan, March 2013, pp. 453–457.

[12] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA estimation for multiple sparse sources with arbitrarily arranged multiple sensors," *Journal of Signal Processing Systems*, vol. 63, no. 3, pp. 265–275, 2011.

[13] S. Rickard and Z. Yilmaz, "On the approximate W-disjoint orthogonality of speech," in *Acoustics, Speech and Signal Processing, 2002. ICASSP 2002. IEEE International Conference on*, vol. 1, April 2002.

[14] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *Signal Processing, IEEE Transactions on*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[15] Y. Chan, R. Hattin, and J. Plant, "The least squares estimation of time delay and its use in signal detection," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.*, vol. 3, April 1978, pp. 665 – 669.

[16] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.

[17] V. Pan, "Algebraic complexity of computing polynomial zeros," *Computers & Mathematics with Applications*, vol. 14, no. 4, pp. 285–304, 1987.