KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Dietzen T., De Sena E., van Waterschoot T. **Low-Complexity Steered Response Power Mapping based on Low-Rank and Sparse Interpolation** STADIUS Technical Report TR 23-90, KU Leuven, pp. 1–13, June 2023.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	
Journal homepage	
Author contact	thomas.dietzen@esat.kuleuven.be
Abstract	For acoustic source localization, a map of the acoustic scene as obtained by the steered response power (SRP) approach can be employed. In SRP, the frequency-weighted output power of a beamformer steered towards a set of candidate locations is obtained from generalized cross-correlations (GCCs). Due to the dense grid of candidate locations, conventional SRP exhibits a high computational complexity. While a number of low-complexity SRP-based localization approaches using non-exhaustive spatial search have been proposed, few studies aim to construct a full SRP map at reduced computational cost. In this paper, we propose two scalable approaches to this problem. Expressing the SRP map as a matrix transform of frequency-domain GCCs, we decompose the SRP matrix into a sampling matrix and an interpolation matrix. While the sampling operation can be implemented efficiently by the inverse fast Fourier transform (iFFT), we propose to use optimal low-rank or sparse

approximations of the interpolation matrix for further complexity reduction. The proposed approaches, refered to as sampling + low-rank interpolation-based SRP (SLRI-SRP) and sampling + sparse interpolation-based SRP (SSPI-SRP), are evaluated in a near-field (NF) and a far-field (FF) localization scenario and compared to a state-of-the-art low rank-based SRP approach (LR-SRP). The results indicate that SSPI-SRP outperforms both SLRI-SRP and LR-SRP over a wide complexity range in terms of approximation error and localization accuracy, achieving a complexity reduction of two to three orders of magnitude as compared to conventional SRP. A MATLAB implementation is available online.

IR

(article begins on next page)

Low-Complexity Steered Response Power Mapping based on Low-Rank and Sparse Interpolation

Thomas Dietzen, Enzo De Sena, *Senior Member,* Toon van Waterschoot, *Member*

***Abstract*—For acoustic source localization, a map of the acoustic scene as obtained by the steered response power (SRP) approach can be employed. In SRP, the frequency-weighted output power of a beamformer steered towards a set of candidate locations is obtained from generalized cross-correlations (GCCs). Due to the dense grid of candidate locations, conventional SRP exhibits a high computational complexity. While a number of low-complexity SRP-based localization approaches using non-exhaustive spatial search have been proposed, few studies aim to construct a full SRP map at reduced computational cost. In this paper, we propose two scalable approaches to this problem. Expressing the SRP map as a matrix transform of frequency-domain GCCs, we decompose the SRP matrix into a sampling matrix and an interpolation matrix. While the sampling operation can be implemented efficiently by the inverse fast Fourier transform (iFFT), we propose to use optimal low-rank or sparse approximations of the interpolation matrix for further complexity reduction. The proposed approaches, refered to as sampling + low-rank interpolation-based SRP (SLRI-SRP) and sampling + sparse interpolation-based SRP (SSPI-SRP), are evaluated in a near-field (NF) and a far-field (FF) localization scenario and compared to a state-of-the-art low-rank-based SRP approach (LR-SRP). The results indicate that SSPI-SRP outperforms both SLRI-SRP and LR-SRP over a wide complexity range in terms of approximation error and localization accuracy, achieving a complexity reduction of two to three orders of magnitude as compared to conventional SRP. A MATLAB implementation is available online.**

***Index Terms*—Acoustic source localization, steered response power, complexity reduction, sampling, low-rank interpolation, sparse interpolation**

I. Introduction

Acoustic source localization [1]–[3] in noisy and reverberant environments remains an essential task in many multi-microphone signal processing [2], [4], [5] applications. Next to classical localization techniques using explicit models exploiting time differences of arrival (TDOAs), such as approaches based on the generalized cross-correlation (GCC) [6], subspace-based approaches like MUSIC [7] or ESPRIT [8], sparse estimation approaches like SPICE [9], or steered response power (SRP) mapping [10]–[30], also learning-based approaches [26], [30], [31] have been proposed recently.

In SRP, a map of the acoustic scene is computed from the frequency-weighted output power of a delay-and-sum (DSB) beamformer steered towards a set of candidate locations. Given this map, the location of a single source may be inferred from its maximum. It is well known [10], [22] that SRP may equivalently be expressed in terms of time-domain GCCs (TD GCCs) [6] of the microphone signals at lags equal to the candidate locations' TDOAs. The major disadvantage of conventional SRP is posed by its high computational complexity due to the dense grid of candidate locations.

In the literature, several low-complexity SRP approaches [11], [13]–[15], [17]–[21], [24], [25], [28] have been proposed. The majority of these approaches [11], [14], [15], [17]–[21], [24] essentially performs a non-exhaustive search as a strategy to reduce computational cost: In stochastic region contraction [14], [17] or coarse-to-fine search strategies [11], [15], the search space is iteratively contracted based on the SRP map obtained in the previous iteration. Volumetric [18]–[21] or density-based [24] approaches accumulate the SRP function corresponding to the volume surrounding the candidate locations [18]–[21] or a surface element in the spherical harmonics domain [24]. This allows for a coarser spatial grid [18]–[20] and (similarly to [11], [14], [15], [17]) enables iterative or hierarchical [19], [20], [24] search approaches.

Recent advances in acoustic mapping and localization using distributed processing [29], [30] or deep learning [26], [30] require a full SRP map, where the non-exhaustive search approaches described above are not immediately applicable. Few studies [13], [25], [28] aim to construct a full SRP map at a reduced computational cost: In [13], instead of searching over space, the search is performed over TDOAs, at each of which the SRP map is updated for all corresponding candidate locations. In [25], a low-rank approximation of the SRP matrix is proposed. In our previous work [28], we have introduced a low-complexity SRP approach based on Nyquist-Shannon sampling and interpolation.

In this paper, we propose two extensions to our previous work [28]. As compared to other SRP approaches using interpolation [12], [16], the main contribution of this work consists in the derivation of scalable, low-complexity interpolation schemes based on global optimality criteria as well as the thorough investigation of the trade-off between complexity and performance. While the sampling operation can be implemented efficiently by the inverse fast Fourier transform (iFFT), we propose to use optimal low-rank or sparse approximations of the interpolation matrix for further complexity reduction. The proposed approaches, refered to as sampling + low-rank interpolation-based SRP (SLRI-SRP) and sampling + sparse interpolation-based SRP (SSPI-SRP), are evaluated in a near-field (NF) and a far-field (FF) localization scenario and compared to a state-of-the-art low rank-based SRP approach (LR-SRP) [25]. The results indicate that SSPI-SRP outperforms both SLRI-SRP and LR-SRP over a wide complexity range

T. Dietzen and T. van Waterschoot are with KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium. Enzo De Sena is with University of Surrey, Institute of Sound Recording, Guildford, UK.

Manuscript received 2023.

in terms of approximation error and localization accuracy, achieving a complexity reduction of two to three orders of magnitude as compared to conventional SRP. A MATLAB implementation is available online [32].

The paper is organized as follows. Sec. II introduces conventional SRP. In Sec. III, we review [25] and our previous work [28]. The proposed low-rank and sparse interpolation approaches are introduced in Sec. IV and evaluated in Sec. V. Sec. VI concludes the paper.

II. Steered Response Power Mapping

The SRP map is defined over a set of candidate locations. Towards each candidate location, one steers a DSB and obtains the SRP value as the frequency-weighted output power. We define the signal and propagation model in Sec. II-A, revisit SRP in Sec. II-B, outline its relation to the TD GCC in Sec. II-C, define a matrix-vector representation in II-D, and discuss its complexity in Sec. II-E.

Notation: column vectors and matrices are denoted by lower- and upper-case boldface letters. For matrices $\mathbf{A}$ (and vectors $\mathbf{a}$ where applicable), $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{\mathrm{H}}$, $\Re[\mathbf{A}]$, $\mathrm{tr}[\mathbf{A}]$, $\mathrm{rank}[\mathbf{A}]$, $\mathrm{vec}[\mathbf{A}]$, $\|\mathbf{A}\|_{\mathrm{F}}$, and $[\mathbf{A}]_{m,m'}$ respectively denote the transpose, the Hermitian transpose, the real part, the trace, the rank, the vectorization, the Frobenius norm, and the element of $\mathbf{A}$ at index (m, m'). The operation $\mathrm{blkdiag}[\mathbf{A}_1, \mathbf{A}_2, \dots]$ creates a block-diagonal matrix from its input arguments. For vectors $\mathbf{a}$, we respectively use $\|\mathbf{a}\|$, $\|\mathbf{a}\|_0$, $\mathrm{iFFT}[\mathbf{a}]$, and $\hat{\mathbf{a}}$ to denote the Euclidean norm, the zero norm, the iFFT, and an estimate of $\mathbf{a}$. For scalars a, the complex conjugate is denoted by a^*, and $\min[a, b]$ and $\mathrm{mod}[a, b]$ denote the minimum of a and b and a modulo b, respectively. The arcus cosine of a is denoted by $\mathrm{acos}[a]$, Euler's number by e and the imaginary unit by j.

A. Signal and Propagation Model

Let the microphone signals be critically sampled with period T, let ω denote the radial frequency and M the number of microphones, and let $y_m(\omega)$ with $m \in \{1, \dots, M\}$ denote the m^{th} microphone signal in the frequency domain. Critical sampling implies that the microphone signals are bandlimited by

$$\omega_{\mathrm{b}} = \pi/T, \tag{1}$$

that is we assume that $y_m(\omega) = 0$ for $\omega \geq \omega_{\mathrm{b}}$.

In SRP, the DSB is usually not directly applied to the microphone signals, but to a frequency-weighted version thereof in order to improve spatial resolution [6], [10], [22], [33]. To this end, one defines the deterministic frequency-domain GCC (FD GCC) matrix $\mathbf{\Psi}(\omega) \in \mathbb{C}^{M \times M}$ as

$$\begin{aligned}[\mathbf{\Psi}(\omega)]_{m,m'} &= \psi_{m,m'}(\omega) \\ &= \gamma_{m,m'}(\omega) y_m(\omega) y^*_{m'}(\omega),\end{aligned} \tag{2}$$

where $\gamma_{m,m'}(\omega)$ denotes the weighting function. A common, heuristically motivated weighting approach is the so-called phase transform (PHAT) [6], [10]–[30], [33] given by $\gamma_{m,m'}(\omega) = 1/|y_m(\omega) y^*_{m'}(\omega)|$, in which case only phase information remains on the right-hand side of (2). The remainder of this paper holds regardless of the actual weighting approach.

We consider both near- and far-field (NF and FF) sound propagation. Let the Cartesian coordinate vectors of the M microphone locations be denoted as $\mathbf{p}_m \in \mathbb{R}^3$. In case of NF propagation, let $\mathbf{q}_i \in \mathbb{R}^3$ with $i \in \{1, \dots, J\}$ be the Cartesian coordinate vectors of the J candidate locations. In case of FF propagation, let $\mathbf{q}_i$ instead denote the unit-length incident direction vectors of the J candidate locations. The TDOA $\Delta t_{m,m'}(i)$ of a sound wave originating from $\mathbf{q}_i$ as observed at the microphone pair (m, m') is then given by

$$\Delta t_{m,m'}(i) = \frac{1}{c} \begin{cases} \|\mathbf{p}_m - \mathbf{q}_i\| - \|\mathbf{p}_{m'} - \mathbf{q}_i\| & \text{for NF propagation,} \\ (\mathbf{p}_m - \mathbf{p}_{m'})^{\mathrm{T}} \mathbf{q}_i & \text{for FF propagation,} \end{cases} \tag{3}$$

where c is the speed of sound. Note that with $d_{m,m'} = \|\mathbf{p}_m - \mathbf{p}_{m'}\|$ the inter-microphone distance, we find that

$$\Delta t_{m,m'}(i) \in [-\Delta t_{m,m'|0},\, \Delta t_{m,m'|0}] \tag{4}$$

$$\text{with} \quad \Delta t_{m,m'|0} = d_{m,m'}/c, \tag{5}$$

where the limits of the interval are obtained if $\mathbf{p}_m$, $\mathbf{p}_{m'}$, and $\mathbf{q}_i$ lie on the same line in the NF case or if $\mathbf{p}_m - \mathbf{p}_{m'}$ and $\mathbf{q}_i$ are colinear in the FF case. The number of different microphone pairs (m, m') is given by

$$P = M(M-1)/2. \tag{6}$$

For convenience, we introduce the microphone pair index $p \in \{1, \dots, P\}$ and admit the alternative notation $\psi_p(\omega) = \psi_{m,m'}(\omega)$ and $\Delta t_p(i) = \Delta t_{m,m'}(i)$ for $m > m'$.

B. Steered Response Power

The DSB steering vector $\mathbf{h}(\omega, i) \in \mathbb{C}^M$ [2], [4], [5] towards the i^{th} candidate location relative to the first microphone is solely defined by $\Delta t_{m,1}(i)$ as

$$[\mathbf{h}(\omega, i)]_m = e^{-j\omega \Delta t_{m,1}(i)}. \tag{7}$$

With $\mathbf{\Psi}(\omega)$ in (2) and $\mathbf{h}(\omega, i)$ in (7), we can define the frequency-dependent SRP map

$$\begin{aligned} z(\omega, i) &= \mathbf{h}^{\mathrm{H}}(\omega, i) \mathbf{\Psi}(\omega) \mathbf{h}(\omega, i) - \mathrm{tr}[\mathbf{\Psi}(\omega)] \\ &= 2 \sum_{m'=1}^{M} \sum_{m=m'+1}^{M} \Re\big[\psi_{m,m'}(\omega) e^{j\omega \Delta t_{m,m'}(i)}\big] \\ &= 2 \sum_{p=1}^{P} \Re\big[\psi_p(\omega) e^{j\omega \Delta t_p(i)}\big], \end{aligned} \tag{8}$$

where $\mathrm{tr}[\mathbf{\Psi}(\omega)]$ is commonly subtracted [10]–[30] as it constitutes an offset independent of i and hence does not contribute to the objective of spatial mapping. The second transition in (8) is obtained from the relation $\Delta t_{m,1}(i) - \Delta t_{m',1}(i) = \Delta t_{m,m'}(i)$ and the Hermitian structure of $\mathbf{\Psi}(\omega)$. Here, the sum is to be taken over all combinations (m, m') with $m > m'$, i.e. over all P microphone pairs. From (8), the broadband SRP value $z(i)$ is obtained by integrating $z(\omega, i)$ over frequency,

$$z(i) = \int_0^{\omega_{\mathrm{b}}} z(\omega, i)\, d\omega, \tag{9}$$

where the upper integration limit coincides with the assumed bandlimit.

The NF or FF location $\mathbf{q}_\mathrm{s}$ of a single source can be estimated from the candidate location index i that maximizes $z(i)$, i.e.

$$\hat{\mathbf{q}}_\mathrm{s} = \mathbf{q}_{i_\mathrm{max}} \tag{10}$$

$$\text{with} \quad i_\mathrm{max} = \arg\max_i z(i). \tag{11}$$

C. Relation to TD GCC

The integral in (9) may be reformulated into a simple relation between $z(i)$ and the TD GCCs of the P microphone pairs [10], [22]. Let $\xi_p(\tau)$ denote the TD GCC defined by the inverse Fourier transform of the FD GCC $\psi_p(\omega)$, i.e.

$$\begin{aligned}\xi_p(\tau) &= \int_{-\omega_\mathrm{b}}^{\omega_\mathrm{b}} \psi_p(\omega) e^{j\omega\tau}\, d\omega \\ &= 2\int_0^{\omega_\mathrm{b}} \Re[\psi_p(\omega) e^{j\omega\tau}]\, d\omega,\end{aligned} \tag{12}$$

where the second transition is valid because $\psi_p(\omega)$ is complex symmetric over ω. Then, using (8) in (9) and comparing to (12) one finds that $z(i)$ becomes

$$z(i) = \sum_{p=1}^{P} \xi_p(\Delta t_p(i)), \tag{13}$$

i.e. a summation of TD GCCs $\xi_p(\tau)$ at lags τ corresponding to the candidate-location TDOAs $\Delta t_p(i)$. Regarding (12) and (13), recall that $\Delta t_p(i) \in [-\Delta t_{p|0},\, \Delta t_{p|0}]$ according to (4), while $\xi_p(\tau)$ itself has unlimited support in τ.

D. Matrix-Vector Form

In the following, we represent the computation of the SRP map as a matrix-vector product using discrete frequencies. The formulation is based on the TD GCC formulation in Sec. II-C. We assume that the discrete version of the FD GCC is computed from a discrete Fourier transform (DFT) of the original signals. If the DFT is of length $2K$, we obtain discrete frequencies

$$\omega_k = k\omega_\mathrm{b}/K \tag{14}$$

with the two-sided frequency bin index $k \in \{-K+1, \ldots, K\}$. Noting that $\psi_p(\pm\omega_\mathrm{b}) = 0$ due to the bandlimit in (1), the integral in (12) can then be replaced by

$$\begin{aligned}\xi_p(\tau) &= \sum_{k=-K+1}^{K-1} \psi_p(\omega_k) e^{j\omega_k\tau} \\ &= 2\sum_{k=1}^{K-1} \Re\big[\psi_p(\omega_k) e^{j\omega_k\tau}\big],\end{aligned} \tag{15}$$

where the second transition is valid because $\psi_p(\omega)$ is complex symmetric and $\psi_p(0)$ is assumed to be negligible.

Based on (13) and the second, single-sided equality in (15), with the single-sided frequency bin index $k \in \{1, \ldots, K-1\}$, we define the SRP map vector $\mathbf{z} \in \mathbb{R}^J$, the TD GCC vector $\boldsymbol{\xi}_{\Delta t_p} \in \mathbb{R}^J$, the single-sided FD GCC vector $\boldsymbol{\psi}_p \in \mathbb{C}^{K-1}$, and the single-sided SRP matrix $\mathbf{H}_p \in \mathbb{C}^{J\times K-1}$ as

$$[\mathbf{z}]_i = z(i), \tag{16}$$

$$\big[\boldsymbol{\xi}_{\Delta t_p}\big]_i = \xi_p(\Delta t_p(i)), \tag{17}$$

$$[\boldsymbol{\psi}_p]_k = \psi_p(\omega_k), \tag{18}$$

$$[\mathbf{H}_p]_{i,k} = e^{j\omega_k \Delta t_p(i)}. \tag{19}$$

Stacking over microphone pairs, we further define $\boldsymbol{\psi} \in \mathbb{C}^{P(K-1)}$ and $\mathbf{H} \in \mathbb{C}^{J\times P(K-1)}$ as

$$\boldsymbol{\psi} = \begin{pmatrix}\boldsymbol{\psi}_1^\mathrm{T} & \cdots & \boldsymbol{\psi}_P^\mathrm{T}\end{pmatrix}^\mathrm{T}, \tag{20}$$

$$\mathbf{H} = \begin{pmatrix}\mathbf{H}_1 & \cdots & \mathbf{H}_P\end{pmatrix}. \tag{21}$$

Using the definitions in (16)–(21), $\mathbf{z}$ can be expressed as

$$\boxed{\begin{aligned}\mathbf{z} &= \sum_{p=1}^{P} \boldsymbol{\xi}_{\Delta t_p} \\ &= 2\sum_{p=1}^{P} \Re[\mathbf{H}_p \boldsymbol{\psi}_p] \\ &= 2\Re[\mathbf{H}\boldsymbol{\psi}].\end{aligned}} \tag{22}$$

In the remainder, we refer to (22) as conventional SRP.

E. Computational Complexity

The computation of the conventional SRP map $\mathbf{z}$ in (22) requires one matrix-vector product,[1] namely $\mathbf{H} \cdot \boldsymbol{\psi}$. The complexity thereof is determined by the number of (non-zero) elements in $\mathbf{H}$, given by $JP(K-1)$. Since the product is complex, but only its real part is required, we assume an additional factor 2. The operation in (22) therefore requires

$$\boxed{C = 2JP(K-1)} \tag{23}$$

multiplications. Both J and K are design parameters and commonly chosen in the order of hundreds or thousands, while P grows quadratically with M, making conventional SRP rather complex. For an example, consider the parameters in Table I chosen for the NF and FF scenario used in Sec. V.

III. Review of the State of the Art

We review two low-complexity SRP approaches. A low-rank SRP approach referred to as SVD-PHAT [25] is discussed in Sec. III-A, and the approach based on Nyquist-Shannon sampling and interpolation introduced in our prior work [28] is discussed in Sec. III-B.

A. Low-Rank SRP

We discuss the low-rank approximation in Sec. III-A1 and the computational complexity of the approach in Sec. III-A2.

1) Low-Rank Approximation: SVD-PHAT [25] can be defined based on the cost function

$$f(\mathbf{H}_\mathrm{var}) = \|\mathbf{H}_\mathrm{var} - \mathbf{H}\|_\mathrm{F}^2 \tag{24}$$

where $\mathbf{H}_\mathrm{var}$ is the optimization variable. In order to obtain a computationally less demanding approximation $\mathbf{H}_\mathrm{lr}$ of $\mathbf{H}$, (24) is minimized subject to a low-rank constraint, i.e.

$$\mathbf{H}_\mathrm{lr} = \arg\min_{\mathbf{H}_\mathrm{var}} \|\mathbf{H}_\mathrm{var} - \mathbf{H}\|_\mathrm{F}^2 \ \mathrm{s.t.}\ \mathrm{rank}[\mathbf{H}_\mathrm{var}] \le R_H, \tag{25}$$

[1]We disregard the complexity of computing the FD GCC $\boldsymbol{\psi}$ since this step is common to all presented approaches.

where the choice of the design parameter $R_H < \min[J, P(K-1)]$ determines the computational complexity, cf. Sec. III-A2. The solution to (25) is based on the singular value decomposition (SVD) of $\mathbf{H}$ [34] and can be written as an outer product of a tall matrix $\mathbf{H}_{\text{tall}} \in \mathbb{C}^{J\times R_H}$ and a fat matrix $\mathbf{H}_{\text{fat}} \in \mathbb{C}^{R_H\times P(K-1)}$, i.e.

$$\mathbf{H}_{\text{lr}} = \mathbf{H}_{\text{tall}}\mathbf{H}_{\text{fat}}, \tag{26}$$

with for instance $\mathbf{H}_{\text{tall}} = \mathbf{U}\tilde{\mathbf{\Sigma}}$ and $\mathbf{H}_{\text{fat}} = \mathbf{V}^{\text{H}}$, where the diagonal of $\tilde{\mathbf{\Sigma}} \in \mathbb{R}^{R_H\times R_H}$ contains the R_H largest singular values of $\mathbf{H}$, and the columns of $\mathbf{U} \in \mathbb{C}^{J\times R_H}$ and $\mathbf{V} \in \mathbb{C}^{P(K-1)\times R_H}$ are given by the corresponding left and right singular vectors, respectively.

The corresponding SRP map is defined similarly to (22) as

$$\boxed{\begin{aligned}\mathbf{z}_{\text{lr}} &= 2\Re[\mathbf{H}_{\text{lr}}\boldsymbol{\psi}]\\ &= 2\Re[\mathbf{H}_{\text{tall}}\mathbf{H}_{\text{fat}}\boldsymbol{\psi}].\end{aligned}} \tag{27}$$

In the remainder, we refer to (27) as low-rank SRP (LR-SRP) and denote the corresponding source-location estimate by $\hat{\mathbf{q}}_{\text{s|lr}}$.

2) Computational Complexity: The decomposition of $\mathbf{H}_{\text{lr}}$ in (26) can be computed upfront and hence does not require computations during run time. The computation of the LR-SRP map $\mathbf{z}_{\text{lr}}$ in (27) requires two matrix-vector multiplications, namely first $\mathbf{H}_{\text{fat}}\cdot\boldsymbol{\psi}$ resulting in the vector $\mathbf{H}_{\text{fat}}\boldsymbol{\psi}$, and second $\mathbf{H}_{\text{tall}}\cdot\mathbf{H}_{\text{fat}}\boldsymbol{\psi}$. The computational complexity thereof is determined by the number of (non-zero) elements in $\mathbf{H}_{\text{fat}}$ and $\mathbf{H}_{\text{tall}}$, given by $R_H P(K-1)$ and JR_H, respectively. Both products are complex, but of the second product only the real part is required, such that we assume additional factors of 4 and 2, respectively. The operation in (27) therefore requires

$$\boxed{C_{\text{lr}} = 2JR_H + 4R_H P(K-1)} \tag{28}$$

multiplications. If R_H is chosen small enough, C_{lr} will be well below C in (23).

B. Sampling and Interpolation-based SRP

We discuss sampling and interpolation according to [28] in Sec. III-B1, a corresponding matrix-vector form in Sec. III-B2, and the computational complexity of the approach in Sec. III-B3.

1) Sampling and Interpolation: Recall that $\xi_p(\tau)$ is bandlimited by ω_{b}, as can be seen from (12). We can therefore invoke the Nyquist-Shannon sampling theorem [35]. Let $\xi_p(\tau)$ be critically sampled at $\tau = nT$ with $n \in \mathbb{Z}$. Using Whittaker-Shannon interpolation [35], we can then express $\xi_p(\tau)$ as a function of its samples $\xi_p(nT)$,

$$\xi_p(\tau) = \sum_{n=-\infty}^{\infty} \xi_p(nT)\,\text{sinc}(\tau/T - n), \tag{29}$$

permitting perfect reconstruction. As $\xi_p(\tau)$ is bandlimited in ω, it has unlimited support in τ [35] and thus, as stated by (29), one indeed needs infinitely many samples $\xi_p(nT)$ in order to exactly reconstruct $\xi_p(\tau)$. Nonetheless, we may approximate $\xi_p(\tau)$ based on (29) and the following arguments.

In the SRP use case, the approximation needs to be accurate only within the interval $\tau \in [-\Delta t_{p|0},\, \Delta t_{p|0}]$, cf. (3)–(4), where τ may be interpreted as the TDOA from a physical source location. Samples $\xi_p(nT)$ sufficiently far outside this interval, i.e. at $|n| \gg \Delta t_{p|0}/T$, have little effect within the interval for two reasons. First, their effect is limited by nature of the interpolating sinc-function, whose envelope decays with the inverse of its argument. Second, also the samples $\xi_p(nT)$ themselves decay as the TD GCC lag nT does not correspond to a TDOA anymore.[2] In [28], we have therefore proposed to approximate $\xi_p(\tau)$ by $\xi_{p|\text{si}}(\tau)$ by truncating (29) as

$$\xi_{p|\text{si}}(\tau) = \sum_{n=-N_p}^{N_p} \xi_p(nT)\text{sinc}(\tau/T - n), \tag{30}$$

$$\text{with}\quad N_p = \big\lfloor \Delta t_{p|0}/T \big\rceil + N_{\text{aux}}, \tag{31}$$

where N_{aux} is a design parameter defining the number of auxiliary sample points beyond $[-\Delta t_{p|0},\, \Delta t_{p|0}]$. In practice, few auxiliary samples are sufficient, cf. also Sec. V and [28]. Based on the above, the Nyquist-Shannon approximation of the SRP map is obtained by first computing the required samples $\xi_p(nT)$ based on (15), then interpolating $\xi_{p|\text{si}}(\tau)$ at $\tau = \Delta t_p(i)$ according to (30), and finally accumulating over p according to (13). This procedure will become more apparent in the following subsection III-B2.

2) Matrix-Vector Form: In the following, we represent the computation of the sampling and interpolation-based SRP map as a matrix-vector product. This respresentation has not been included in our previous work [28], but is essential to the approaches proposed in Sec. IV.[3]

With $n \in \{-N_p, \ldots, N_p\}$, let the column or row index n_{mod} be defined as

$$n_{\text{mod}} = \text{mod}[n, 2N_p + 1] + 1, \tag{32}$$

yielding $n_{\text{mod}} \in \{1, \ldots, 2N_p + 1\}$. Further, let N denote the average number of required samples per microphone pair, i.e.

$$N = 1 + \frac{2}{P}\sum_{p=1}^{P} N_p. \tag{33}$$

Based on (15), we define the TD GCC sample vector $\boldsymbol{\xi}_p \in \mathbb{R}^{2N_p+1}$ and the single-sided sampling matrix $\mathbf{S}_p \in \mathbb{C}^{2N_p+1\times K-1}$ by

$$[\boldsymbol{\xi}_p]_{n_{\text{mod}}} = \xi_p(nT), \tag{34}$$

$$[\mathbf{S}_p]_{n_{\text{mod}},k} = e^{j\pi kn/K}, \tag{35}$$

which can be compared to (17) and (19). In (35), we have used $\omega_k nT = \pi kn/K$ from (1) and (14). Note that $\mathbf{S}_p$ is a submatrix of the $2K$ inverse DFT (iDFT) matrix corresponding to a subset of samples and one side of the spectrum. Stacking over microphone pairs, we define $\boldsymbol{\xi} \in \mathbb{R}^{PN}$ and $\mathbf{S} \in \mathbb{C}^{PN\times P(K-1)}$ as

$$\boldsymbol{\xi} = \begin{pmatrix}\boldsymbol{\xi}_1^{\text{T}} & \cdots & \boldsymbol{\xi}_P^{\text{T}}\end{pmatrix}^{\text{T}} \tag{36}$$

$$\mathbf{S} = \text{blkdiag}\big[\mathbf{S}_1,\, \ldots,\, \mathbf{S}_P\big], \tag{37}$$

[2]Exceptions due to, e.g., strong reflections or source-signal periodicities are possible, but irrelevant to the source localization problem.

[3]Despite its novel aspect, we introduce the matrix-vector form in the state-of-the-art section to ensure a coherent outline and facilitate the complexity analysis in the following section.

Using (34) to (37), with $\boldsymbol{\psi}$ defined in (20), we find that $\boldsymbol{\xi}_p = 2\Re[\mathbf{S}_p\boldsymbol{\psi}_p]$ and accordingly

$$\boldsymbol{\xi} = 2\Re[\mathbf{S}\boldsymbol{\psi}]. \tag{38}$$

Based on (30), we define the interpolated TD GCC vector $\boldsymbol{\xi}_{\Delta t_p|\mathrm{si}} \in \mathbb{R}^J$ and the interpolation matrix $\boldsymbol{\Lambda}_p \in \mathbb{R}^{J\times(2N_p+1)}$ as

$$\left[\boldsymbol{\xi}_{\Delta t_p|\mathrm{si}}\right]_i = \xi_{p|\mathrm{si}}(\Delta t_p(i)), \tag{39}$$

$$[\boldsymbol{\Lambda}_p]_{i,n_{\mathrm{mod}}} = \mathrm{sinc}(\Delta t_p(i)/T - n). \tag{40}$$

Stacking over microphone pairs, we define $\boldsymbol{\Lambda} \in \mathbb{R}^{J\times PN}$ as

$$\boldsymbol{\Lambda} = \begin{pmatrix}\boldsymbol{\Lambda}_1 & \cdots & \boldsymbol{\Lambda}_P\end{pmatrix}. \tag{41}$$

Using (39) to (41), with $\boldsymbol{\xi}_p$ and $\boldsymbol{\xi}$ defined in (34) and (36), we find that $\boldsymbol{\xi}_{\Delta t_p|\mathrm{si}} = \boldsymbol{\Lambda}_p\boldsymbol{\xi}_p$ and accordingly define the corresponding SRP map $\mathbf{z}_{\mathrm{si}}$ based on (13) as

$$\boxed{\begin{aligned}\mathbf{z}_{\mathrm{si}} &= \sum_{p=1}^{P} \boldsymbol{\xi}_{\Delta t_p|\mathrm{si}} \\ &= \sum_{p=1}^{P} \boldsymbol{\Lambda}_p\boldsymbol{\xi}_p \\ &= \boldsymbol{\Lambda}\boldsymbol{\xi}.\end{aligned}} \tag{42}$$

With (38) and (42), we have approximated $\mathbf{H}$ in (22) by

$$\mathbf{H}_{\mathrm{si}} = \boldsymbol{\Lambda}\mathbf{S}. \tag{43}$$

Note that like (26), we can under most cases of practical interest interpret (43) as low-rank: with $\boldsymbol{\Lambda} \in \mathbb{R}^{J\times PN}$ and $\mathbf{S} \in \mathbb{C}^{PN\times P(K-1)}$ we typically find that $\boldsymbol{\Lambda}$ is tall and $\mathbf{S}$ is fat since in usual setups $J > NP$ and $N < K-1$,[4] respectively, cf. also the parameters in Table I.

In the remainder, we refer to (42) as sampling + interpolation-based SRP (SI-SRP) and denote the corresponding source-location estimate by $\hat{\mathbf{q}}_{\mathrm{s}|\mathrm{si}}$.

3) Computational Complexity: The components of $\mathbf{H}_{\mathrm{si}}$ in (43) can be defined upfront and hence do not require computations during run time. The computation of the SI-SRP map $\mathbf{z}_{\mathrm{si}}$ in (42) requires sampling and interpolation. The sampling operation, i.e. the computation of $\boldsymbol{\xi}$, can be performed in two different ways. We may either use the matrix vector product $\mathbf{S}\cdot\boldsymbol{\psi}$ in (38), whereof the complexity is determined by the number of non-zero elements[5] in $\mathbf{S}$, i.e. $PN(K-1)$. Since the product is complex, but only its real part is required, we assume an additional factor 2. Alternatively, noting that $\mathbf{S}_p$ is a submatrix of the $2K$-iDFT matrix, we may use the iFFT.[6] Let $\bar{\boldsymbol{\psi}}_p$ be a two-sided version of $\boldsymbol{\psi}_p$ as defined in the Appendix. Then, $\boldsymbol{\xi}_p$ can be obtained by performing the iFFT on $\bar{\boldsymbol{\psi}}_p$ and selecting the required $2N_p+1$ samples according to

$$\bar{\boldsymbol{\xi}}_p = \mathrm{iFFT}[\bar{\boldsymbol{\psi}}_p], \tag{44}$$

$$[\boldsymbol{\xi}_p]_{n_{\mathrm{mod}}} = [\bar{\boldsymbol{\xi}}_p]_{n'_{\mathrm{mod}}}, \tag{45}$$

with n_{mod} in (32) and the sample selection index n'_{mod} defined by

$$n'_{\mathrm{mod}} = \mathrm{mod}[n, 2K] + 1, \tag{46}$$

yielding $n'_{\mathrm{mod}} \in \{1,\dots,N_p+1\}\cup\{2K-N_p+1,\dots,2K\}$ for $n \in \{-N_p,\dots,N_p\}$. From $\boldsymbol{\xi}_p$ in (45), we obtain the stacked TD GCC vector $\boldsymbol{\xi}$ using (36). One iFFT of size $2K$ requires $2K\log_2[2K]$ complex multiplications, where we assume an additional factor 4. Computing $\boldsymbol{\xi}$ therefore requires C_{samp} multiplications with

$$C_{\mathrm{samp}} = \begin{cases}2PN(K-1) & \text{when using (38),}\\ 8PK\log_2[2K] & \text{when using (44), (45), (36).}\end{cases} \tag{47}$$

The FFT-based computation of $\boldsymbol{\xi}$ is less complex if $N > 4\frac{K}{K-1}\log_2[2K]$, which is often the case in usual setups, cf. the parameters in Table I.

The interpolation operation requires one real matrix-vector product, namely $\boldsymbol{\Lambda}\cdot\boldsymbol{\xi}$ in (42). The computational complexity thereof is determined by the number of (non-zero) elements in $\boldsymbol{\Lambda}$, given by JPN. In total, the operation in (42) therefore requires

$$\boxed{C_{\mathrm{si}} = JPN + C_{\mathrm{samp}}} \tag{48}$$

multiplications. For usual setups, we typically find that C_{si} will be well below C in (23), cf. also the low-rank interpretation of (43) and Table I.

IV. Low-Rank and Sparse Interpolation

The SI-SRP approach presented in Sec. III-B reduces computational complexity depending on the acoustic scenario and algorithmic parameter choices, but is hardly scalable in the sense of trading-off complexity against performance. In this section, we therefore propose two scalable extensions of the SI-SRP approach. While the sampling operation of the SI-SRP approach can already be carried out in a computationally efficient manner using the inverse FFT, the complexity of the interpolation operation depends on the number of candidate locations, which typically poses the largest factor of the computational load. We therefore further reduce complexity based on low-rank and sparse approximations of the interpolation matrix. In Sec. IV-A, we derive the proposed approaches, and in Sec. IV-B discuss the complexity of the resulting algorithms.

A. Low-Rank and Sparse Interpolation-based SRP

We define the cost function

$$f(\boldsymbol{\Lambda}_{\mathrm{var}}) = \|\boldsymbol{\Lambda}_{\mathrm{var}}\mathbf{S} - \mathbf{H}\|_{\mathrm{F}}^2 \tag{49}$$

where $\boldsymbol{\Lambda}_{\mathrm{var}} \in \mathbb{R}^{J\times PN}$ is the optimization variable. Similarly to (24), the cost function in (49) penalizes the approximation error with respect to $\mathbf{H}$.

To reduce complexity, (49) could be minimized subject to low-rank or sparsity constraints on $\boldsymbol{\Lambda}_{\mathrm{var}}$.[7] Such a problem

[4] In order to ensure a good TD GCC estimate within the TDOA range and limit the effects of windowing and circular convolution, the frame length $2K$ should well exceed twice the largest possible TDOA $\Delta t_{p|0}$, which implies that $N \ll 2K$, cf. (33) and (31).

[5] Recall that $\mathbf{S}$ is block-diagonal, cf. (37).

[6] The use of the iFFT has not been considered in our previous work [28].

[7] In addition, one would either require a realness constraint on $\boldsymbol{\Lambda}_{\mathrm{var}}$ or alternatively use a two-sided version of (49), cf. also the corresponding two-sided definitions $\bar{\mathbf{H}}$ and $\bar{\mathbf{S}}$ in the Appendix.

could for instance be solved iteratively by means of the proximal gradient method [36], that is by alternatingly performing a gradient-descent step based on the convex cost function in (49) and applying a proximal operator based on the constraints. Instead however, we propose to simplify the cost function in order to obtain a simpler minimization problem that can be solved without requiring alternations.

To this end, we first approximate (49) by replacing $\mathbf{H}$ by $\mathbf{H}_{\mathrm{si}}$ yielding

$$f(\mathbf{\Lambda}_{\mathrm{var}}) \approx \|\mathbf{\Lambda}_{\mathrm{var}}\mathbf{S} - \mathbf{H}_{\mathrm{si}}\|_{\mathrm{F}}^2, \tag{50}$$

and then exploit the properties of $\mathbf{S}$ as outlined in the following. With $\mathbf{H}_{\mathrm{si}} = \mathbf{\Lambda}\mathbf{S}$ from (43), we find that the right-hand side of (50) becomes

$$\begin{aligned}\|\mathbf{\Lambda}_{\mathrm{var}}\mathbf{S} - \mathbf{H}_{\mathrm{si}}\|_{\mathrm{F}}^2 &= \|(\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda})\mathbf{S}\|_{\mathrm{F}}^2 \\ &\approx \frac{1}{2}\|(\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda})\bar{\mathbf{S}}\|_{\mathrm{F}}^2 \\ &= K\|\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda}\|_{\mathrm{F}}^2, \end{aligned} \tag{51}$$

where $\bar{\mathbf{S}}$ is a two-sided version of $\mathbf{S}$ as defined in the Appendix and the second transition holds for $\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda}$ being real. The error in this transition is solely due to the columns of $\bar{\mathbf{S}}$ relating to $\omega_k = 0$ and $\omega_k = \pm\omega_{\mathrm{b}}$, which are disregarded in the definition of $\mathbf{S}$, cf. (35) and (37), and is considered negligible. The third transition holds because $\|\mathbf{A}\|_{\mathrm{F}}^2 = \mathrm{tr}[\mathbf{A}\mathbf{A}^{\mathrm{H}}]$ for any matrix $\mathbf{A}$ and the rows of $\bar{\mathbf{S}}$ are orthogonal with $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\mathrm{H}} = 2K\mathbf{I}$.

Hence, the last equality in (51) is nearly equivalent to (49) and can be used as a cost function instead. In order to obtain computationally less demanding approximations $\mathbf{\Lambda}_{\mathrm{lr}}$ and $\mathbf{\Lambda}_{\mathrm{sp}}$ of $\mathbf{\Lambda}$, we propose to minimize this cost function subject to low-rank or sparsity constraints, respectively, i.e.

$$\mathbf{\Lambda}_{\mathrm{lr}} = \arg\min_{\mathbf{\Lambda}_{\mathrm{var}}} \|\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda}\|_{\mathrm{F}}^2 \text{ s.t. } \mathrm{rank}[\mathbf{\Lambda}_{\mathrm{var}}] \leq R_\Lambda, \tag{52}$$

$$\mathbf{\Lambda}_{\mathrm{sp}} = \arg\min_{\mathbf{\Lambda}_{\mathrm{var}}} \|\mathbf{\Lambda}_{\mathrm{var}} - \mathbf{\Lambda}\|_{\mathrm{F}}^2 \text{ s.t. } \|\mathrm{vec}[\mathbf{\Lambda}_{\mathrm{var}}]\|_0 \leq Q_\Lambda, \tag{53}$$

where the choices of the design parameters $R_\Lambda < \min[J, PN]$ and $Q_\Lambda < JPN$ determine the computational complexity, cf. Sec. IV-B.

The solution to (52) is based on the SVD of $\mathbf{\Lambda}$ [34] and can be written as an outer product of a tall matrix $\mathbf{\Lambda}_{\mathrm{tall}} \in \mathbb{R}^{J\times R_\Lambda}$ and a fat matrix $\mathbf{\Lambda}_{\mathrm{fat}} \in \mathbb{R}^{R_\Lambda \times PN}$, i.e.

$$\mathbf{\Lambda}_{\mathrm{lr}} = \mathbf{\Lambda}_{\mathrm{tall}}\mathbf{\Lambda}_{\mathrm{fat}}, \tag{54}$$

where $\mathbf{\Lambda}_{\mathrm{tall}}$ and $\mathbf{\Lambda}_{\mathrm{fat}}$ can be defined similarly to $\mathbf{H}_{\mathrm{tall}}$ and $\mathbf{H}_{\mathrm{fat}}$ in Sec. III-A1. The solution to (53) is obtained from $\mathbf{\Lambda}$ by keeping the Q_Λ largest elements and discarding all others.

Using (52) to (54), with $\boldsymbol{\xi}$ defined in (36), we define the corresponding SRP maps $\mathbf{z}_{\mathrm{slri}}$ and $\mathbf{z}_{\mathrm{sspi}}$ similarly to (42) as

$$\boxed{\begin{aligned}\mathbf{z}_{\mathrm{slri}} &= \mathbf{\Lambda}_{\mathrm{lr}}\boldsymbol{\xi} \\ &= \mathbf{\Lambda}_{\mathrm{tall}}\mathbf{\Lambda}_{\mathrm{fat}}\boldsymbol{\xi}, \end{aligned}} \tag{55}$$

$$\boxed{\mathbf{z}_{\mathrm{sspi}} = \mathbf{\Lambda}_{\mathrm{sp}}\boldsymbol{\xi}.} \tag{56}$$

With (38) and (55) or (56), we have approximated $\mathbf{H}$ in (22) by

$$\mathbf{H}_{\mathrm{slri}} = \mathbf{\Lambda}_{\mathrm{lr}}\mathbf{S}, \tag{57}$$

$$\mathbf{H}_{\mathrm{sspi}} = \mathbf{\Lambda}_{\mathrm{sp}}\mathbf{S}, \tag{58}$$

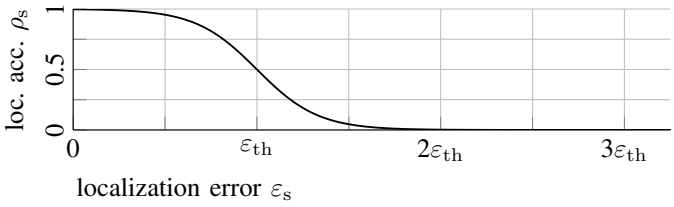


Fig. 1: The localization accuracy ρ_{s} as a function of the localization error ε_{s} as defined in (65) for $\sigma = 6$.

respectively. In the remainder, we refer to (55) and (56) as sampling + low-rank interpolation-based SRP (SLRI-SRP) and sampling + sparse interpolation-based SRP (SSPI-SRP) and denote the corresponding source-location estimate by $\hat{\mathbf{q}}_{\mathrm{s|slri}}$ and $\hat{\mathbf{q}}_{\mathrm{s|sspi}}$, respectively.

B. Computational Complexity

The decompositions of $\mathbf{\Lambda}_{\mathrm{lr}}$, $\mathbf{H}_{\mathrm{slri}}$, and $\mathbf{H}_{\mathrm{sspi}}$ in (54), (57), and (58) can be computed upfront and hence do not require computations during run time. The computation of the SLRI-SRP map $\mathbf{z}_{\mathrm{slri}}$ and the SSPI-SRP $\mathbf{z}_{\mathrm{sspi}}$ in (55) and (56) requires sampling and interpolation, where the complexity of the sampling operation C_{samp} remains unaltered and is given by (47).

The interpolation operation in (55) requires two real matrix-vector products, namely first $\mathbf{\Lambda}_{\mathrm{fat}} \cdot \boldsymbol{\xi}$ resulting in the vector $\mathbf{\Lambda}_{\mathrm{fat}}\boldsymbol{\xi}$, and second $\mathbf{\Lambda}_{\mathrm{tall}} \cdot \mathbf{\Lambda}_{\mathrm{fat}}\boldsymbol{\xi}$. The computational complexity thereof is determined by the number of (non-zero) elements in $\mathbf{\Lambda}_{\mathrm{fat}}$ and $\mathbf{\Lambda}_{\mathrm{tall}}$, given by $R_\Lambda PN$ and JR_Λ, respectively. The interpolation operation in (56) requires one real matrix-vector product, namely $\mathbf{\Lambda}_{\mathrm{sspi}} \cdot \boldsymbol{\xi}$, whereof the computational complexity is determined by the number of (non-zero) elements in $\mathbf{\Lambda}_{\mathrm{sp}}$, i.e. Q_Λ.

In total, the operations in (55) or (56) therefore require

$$\boxed{\begin{aligned} C_{\mathrm{slri}} &= JR_\Lambda + R_\Lambda PN + C_{\mathrm{samp}} \\ C_{\mathrm{sspi}} &= Q_\Lambda + C_{\mathrm{samp}} \end{aligned}} \tag{59, 60}$$

multiplications. If R_Λ or Q_Λ are chosen small enough, C_{slri} and C_{sspi} will be well below C in (23), respectively.

V. Simulations

We compare SLRI-SRP and SSPI-SRP to conventional SRP and LR-SRP. In Sec. V-A, we define performance measures. The simulation setup and the simulation results are described in Sec. V-B an Sec. V-C. A MATLAB implementation is available online [32].

A. Performance Measures

We define the following performance measures. For $C_{\mathrm{type}} \in \{C_{\mathrm{lr}}, C_{\mathrm{si}}, C_{\mathrm{slri}}, C_{\mathrm{sspi}}\}$, we define the relative complexity with respect to C as

$$C_{\mathrm{rel}} = C_{\mathrm{type}}/C. \tag{61}$$

The SRP matrix error $\varepsilon_H(\mathbf{H}_{\mathrm{type}})$ with respect to $\mathbf{H}$ for $\mathbf{H}_{\mathrm{type}} \in \{\mathbf{H}_{\mathrm{lr}}, \mathbf{H}_{\mathrm{si}}, \mathbf{H}_{\mathrm{slri}}, \mathbf{H}_{\mathrm{sspi}}\}$ and the SRP map error

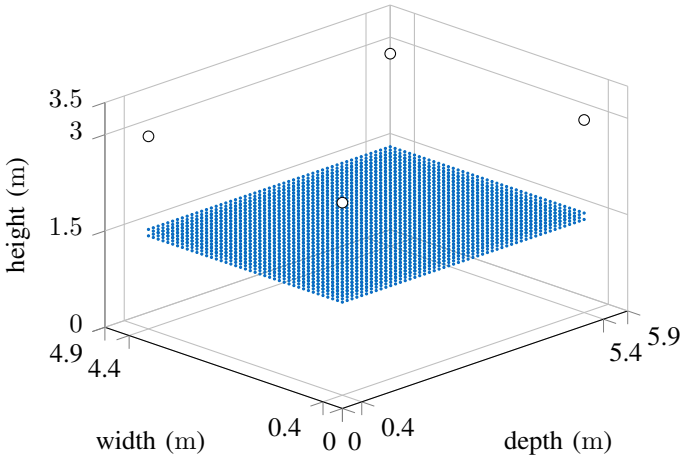


Fig. 2: Microphone [○] and candidate locations [·] in the NF scenario. The three-dimensional search grid samples a volume of $(4 \times 5 \times 0.1)\mathrm{m}$. For better visibility, the search grid is shown with a reduced resolution of 10 cm instead of the simulated resolution of 3.33 cm per spatial dimension. The axes limits correspond to the room dimensions.

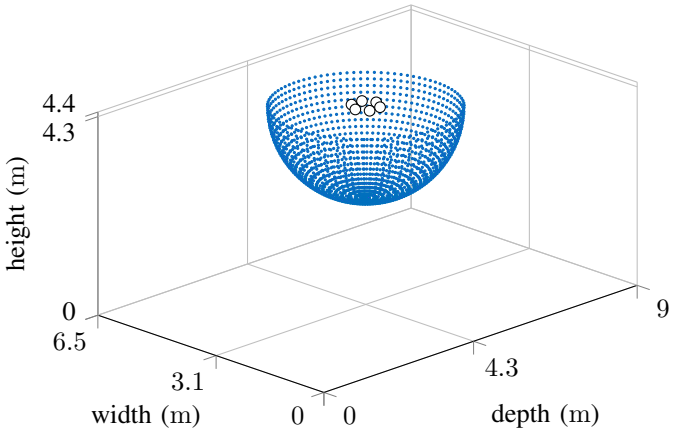


Fig. 3: Microphone [○] and candidate locations [·] in the FF scenario. The two-dimensional search grid samples the lower half-sphere of incident directions, indicated at 2 m distance from the center of the microphone array beyond which the FF assumption is assumed to be valid. For better visibility, the search grid is shown with a reduced resolution of 4° instead of the simulated resolution of 2° per angular dimension. The axes limits correspond to the room dimensions.

$\varepsilon_z(\mathbf{z}_{\mathrm{type}})$ with respect to $\mathbf{z}$ for $\mathbf{z}_{\mathrm{type}} \in \{\mathbf{z}_{\mathrm{lr}}, \mathbf{z}_{\mathrm{si}}, \mathbf{z}_{\mathrm{slri}}, \mathbf{z}_{\mathrm{sspi}}\}$ are defined as

$$\varepsilon_H = 10 \log_{10} \frac{\|\mathbf{H}_{\mathrm{type}} - \mathbf{H}\|_{\mathrm{F}}^2}{\|\mathbf{H}\|_{\mathrm{F}}^2} \,\mathrm{dB}, \tag{62}$$

$$\varepsilon_z = 10 \log_{10} \frac{\|\mathbf{z}_{\mathrm{type}} - \mathbf{z}\|^2}{\|\mathbf{z}\|^2} \,\mathrm{dB}. \tag{63}$$

Further, for the NF or FF source-location estimates $\hat{\mathbf{q}}_{\mathrm{s}|\,\mathrm{type}} \in \{\hat{\mathbf{q}}_{\mathrm{s}}, \hat{\mathbf{q}}_{\mathrm{s}|\mathrm{lr}}, \hat{\mathbf{q}}_{\mathrm{s}|\mathrm{si}}, \hat{\mathbf{q}}_{\mathrm{s}|\mathrm{slri}}, \hat{\mathbf{q}}_{\mathrm{s}|\mathrm{sspi}}\}$, we define the localization error ε_{s} as the distance between the estimated and the true source location and the angle between the estimated and the true incident direction, respectively, i.e.

$$\varepsilon_{\mathrm{s}} = \begin{cases} \|\hat{\mathbf{q}}_{\mathrm{s}|\,\mathrm{type}} - \mathbf{q}_s\| & \text{for NF localization,} \\ \mathrm{acos}[\hat{\mathbf{q}}_{\mathrm{s}|\,\mathrm{type}}^{\mathrm{T}} \mathbf{q}_{\mathrm{s}}] & \text{for FF localization.} \end{cases} \tag{64}$$

TABLE I: Parameters relevant to computational complexity in the NF and FF scenario. In LR-SRP, SLRI-SRP, and SSPI-SRP, C_{rel} is scalable and varied by means of the low-rank or sparsity constraint.

parameter\scenario		NF	FF
grid resolution		3.33 cm	2°
# candidate locations	J	73084	8101
# microphones	M	4	6
# microphone pairs	P	6	15
largest microphone distance	d_p	6.4 m	0.6 m
twice the largest TDOA limit	$2\Delta t_{p\|0}$	37.7 ms	3.5 ms
frame duration	$2KT$	128 ms	64 ms
sampling rate	$1/T$	4 kHz	16 kHz
number of frequency bins	$K-1$	255	511
mean # required samples	N	125	46.6
complexity of conv. SRP	C	$111.8 \cdot 10^6$	$62.1 \cdot 10^6$
rel. complexity of SI-SRP	C_{rel}	0.25	0.05

TABLE II: DRR and critical distance for the simulated NF and FF scenario at given reverberation times.

reverb. time T_{60} (s)	mean DRR (dB)		critical dist. (m)	
NF and FF	NF	FF	NF	FF
0	∞	∞	∞	∞
0.2	-6.3	-2.6	1.69	2.69
0.4	-10.2	-6.7	1.19	1.90
0.8	-13.6	-10.3	0.84	1.35

As the distributions of ε_{s} were observed to have a nearly uniform tail, suggesting that erroneous location estimates are uniformly distributed over the search grid, cf. also Sec. V-C1, we further define the localization accuracy ρ_{s} as a function of ε_{s} by

$$\rho_{\mathrm{s}} = \frac{1}{1 + e^{\sigma(\varepsilon_{\mathrm{s}} - \varepsilon_{\mathrm{th}})/\varepsilon_{\mathrm{th}}}}, \tag{65}$$

where we use the NF and FF thresholds $\varepsilon_{\mathrm{th}} = 0.2\,\mathrm{m}$ and $\varepsilon_{\mathrm{th}} = 2.5°$, respectively, and $\sigma = 6$ in both cases. The localization accuracy function is shown in Fig. 1. We observe that ρ_{s} approaches one for $\varepsilon_{\mathrm{s}} \ll \varepsilon_{\mathrm{th}}$ and zero for $\varepsilon_{\mathrm{s}} \gg \varepsilon_{\mathrm{th}}$, i.e. its distribution is nearly binary.

B. Simulation Setup

We simulate an NF and an FF scenario within shoe-box shaped rooms. The geometries of the simulated scenarios are illustrated in Fig. 2 and Fig. 3, and the parameters relevant to computational complexity are summarized in Table I.

In the NF scenario, $M = 4$ microphones are placed near the corners of a room with dimensions $(4.9 \times 5.9 \times 3.5)\mathrm{m}$ at a height of 3 m and in a distance of 0.4 m to 0.5 m to the walls. The three-dimensional search grid samples a volume of $(4 \times 5 \times 0.1)\mathrm{m}$ and lies below the microphones at around a height of 1.5 m. Its spatial resolution is 3.33 cm, yielding $J = 73084$ candidate locations. In the FF scenario, a circular microphone array with 30 cm radius and $M = 6$ microphones is placed in the horizontal plane 10 cm below the ceiling and near the center of a room with dimensions $(6.5 \times 9 \times 4.5)\mathrm{m}$. The two-dimensional search grid samples the polar and azimuth angle of the candidate locations and covers the lower half-sphere with zenith and azimuth direction along

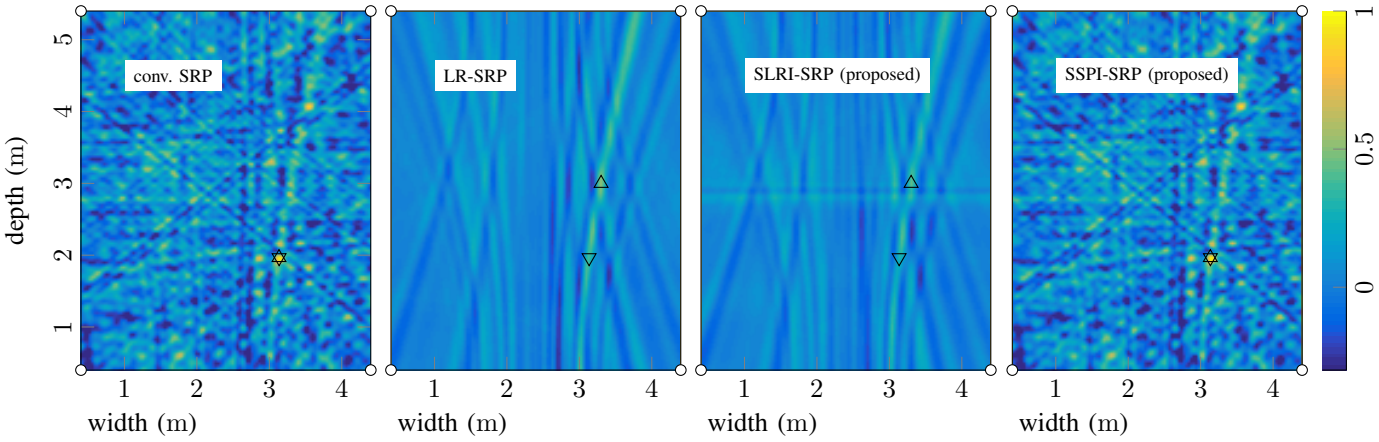


Fig. 4: Example map of conv. SRP and corresponding maps of LR-SRP, SLRI-SRP (proposed), and SSPI-SRP (proposed) with relative complexity $C_{\mathrm{rel}} \approx 2 \cdot 10^{-2}$ (i.e. nearly two orders of magnitude less than conv. SRP) in the NF scenario at $SNR = 0\,\mathrm{dB}$ and reverberation time $T_{60} = 0.2\,\mathrm{s}$. The maps represent the horizontal slice of the (three-dimensional) search grid closest to the true source location. The (projected) microphone locations, the true source location, and the estimated source location are marked by [○], [▽], and [△], respectively.

.

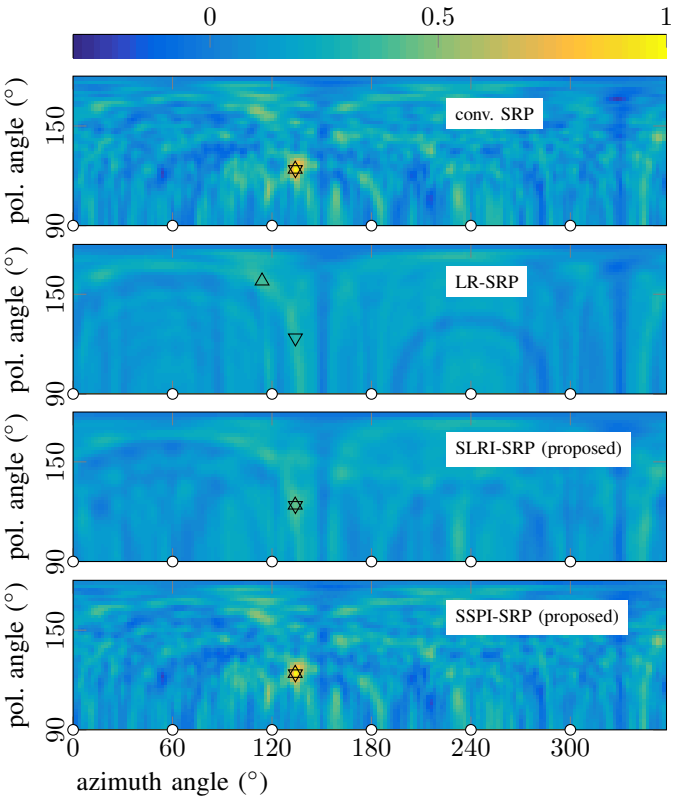


Fig. 5: Example map of conv. SRP and corresponding maps of LR-SRP, SLRI-SRP (proposed), and SSPI-SRP (proposed) with relative complexity $C_{\mathrm{rel}} \approx 1 \cdot 10^{-2}$ (i.e. two orders of magnitude less than conv. SRP) in the FF scenario at $SNR = 0\,\mathrm{dB}$ and reverberation time $T_{60} = 0.4\,\mathrm{s}$. The microphone locations, the true source location, and the estimated source location are marked by [○], [▽], and [△], respectively.

.

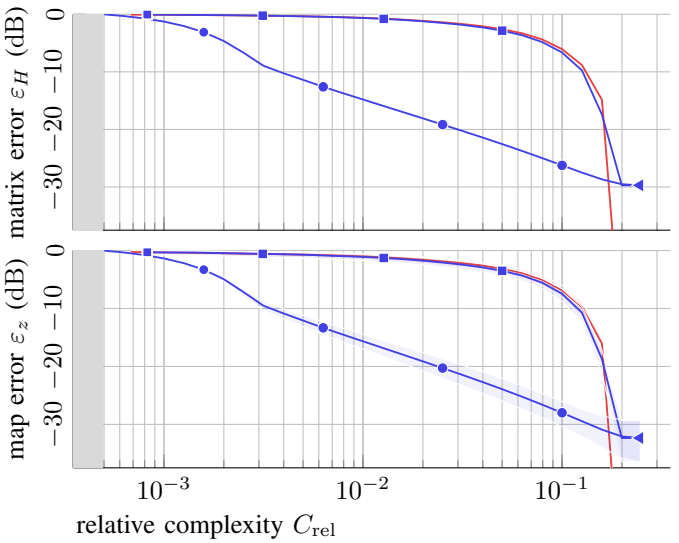


Fig. 6: Matrix error ε_H (top) and median map error ε_z (bottom) vs. relative complexity C_{rel} for LR-SRP [——], SI-SRP (single point) [◂], SLRI-SRP (proposed) [—▪—], and SSPI-SRP (proposed) [—●—] in the NF scenario. The shaded areas in the map error indicate the range from the 10th to the 90th percentile. The gray area indicates the range where $C_{\mathrm{rel}} < C_{\mathrm{samp}}$, which cannot be reached by SLRI-SRP and SSPI-SRP.

.

the third (i.e., the vertical) and the second room dimension. The angular resolution is 2°, yielding $J = 8101$ candidate locations.

In both scenarios, a single source is placed in 512 randomly selected locations, which are not restricted to be on the search grid. In the NF scenario, the random locations are restricted to be inside the volume spanned by the search grid. In the FF scenario, the distance of the source to the center of the microphone array is restricted to be at least 2 m.

Both female and male speech [37] are used as source signals. Diffuse babble noise [38], [39] is added at varying SNRs of $SNR = \{6, 0, -6, -12\}\mathrm{dB}$. The shoe-box rooms are simulated using the randomized image method [40] at varying reverberation times $T_{60} = \{0, 0.2, 0.4, 0.8\}\mathrm{s}$. Note that due to different geometries, the same reverberation time results in different direct-to-reverberant ratios (DRRs) and

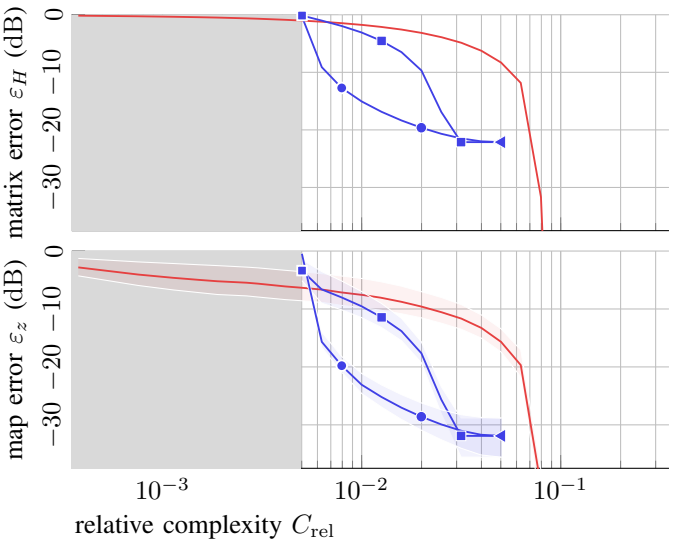


Fig. 7: Matrix error ε_H (top) and median map error ε_z (bottom) vs. relative complexity C_{rel} for LR-SRP [——], SI-SRP (single point) [◂], SLRI-SRP (proposed) [—▪—], and SSPI-SRP (proposed) [—•—] in the FF scenario. The shaded areas in the map error indicate the range from the 10th to the 90th percentile. The gray area indicates the range where $C_{\mathrm{rel}} < C_{\mathrm{samp}}$, which cannot be reached by SLRI-SRP and SSPI-SRP.

critical distances [41] across the two scenarios, with the NF scenario being more adversely affected as shown in Table II.

The speed of sound is set to $c = 340\,\mathrm{m/s}$, and the sampling rates of the NF and the FF scenario are respectively set to $4\,\mathrm{kHz}$ and $16\,\mathrm{kHz}$, implying the bandlimits $\omega_{\mathrm{b}}/2\pi = 2\,\mathrm{kHz}$[8] and $\omega_{\mathrm{b}}/2\pi = 8\,\mathrm{kHz}$.

The microphone signals are processed by a short-time Fourier transform (STFT) using square-root Hann windows of $2K = 512$ samples in the NF scenario and $2K = 1024$ in the FF scenario, corresponding to $128\,\mathrm{ms}$ and $64\,\mathrm{ms}$ and well exceeding the possible TDOA range, cf. Table I. For each random source location, SRP maps are computed for 32 different frames of microphone data.

As shown in Table I, the computational complexity of conventional SRP in the NF and FF scenario is $C = 111.8 \cdot 10^6$ and $C = 62.1 \cdot 10^6$, respectively. In SI-SRP, SLRI-SRP and SSPI-SRP, we set N_{aux} in (31) to $N_{\mathrm{aux}} = 2$, resulting in $N = 125$ and $N = 46.6$ samples per microphone pair in the NF and FF scenario. For SI-SRP, with $\boldsymbol{\xi}$ computed using the iFFT, we obtain the relative complexities $C_{\mathrm{rel}} = 0.25$ and $C_{\mathrm{rel}} = 0.05$. In LR-SRP, SLRI-SRP, and SSPI-SRP, C_{rel} is scalable and varied by means of the low-rank or sparsity constraint.

C. Simulation Results

We first discuss the appearance of SRP maps for the studied approaches in Sec. V-C1, and compare the matrix and map error in Sec. V-C2 and the localization accuracy in Sec. V-C3.

[8]In agreement with the spatial resolution of the search grid, the NF bandlimit is chosen such that the main lobes of the resulting TD GCCs $\xi_p(\tau)$ are wide enough to be sampled by at least one candidate location TDOA $\Delta t_p(i)$, cf. (13) and [23], [27]. While the bandlimit of speech signals is larger than $2\,\mathrm{kHz}$, we low-pass filter the signals accordingly in the NF scenario.

1) Appearance of SRP Maps: In an ideal acoustic map, the power in each candidate location would be influenced only by a source in its close vicinity. The appearance of SRP maps by contrast is dominated by hyperboloids of constant TDOAs [23], with one hyperboloid per microphone pair and (image) source. The P hyperboloids due to the source intersect at the source location, yielding a large SRP value. This justifies the commonly used peak picking-based location estimation in (10)–(11). The intersections of the hyperboloids due to image sources cause a multitude of more or less uniformly distributed local maxima, which can result in localization errors in particular if additional noise is present. For non-stationary source signals, the strength of individual hyperboloids varies over frames, causing fluctuations in the SRP map.

Fig. 4 shows an NF example of a conventional SRP map as well as the corresponding LR-SRP, SLRI-SRP, and SSPI-SRP maps with relative complexity $C_{\mathrm{rel}} \approx 2 \cdot 10^{-2}$ at $SNR = 0\,\mathrm{dB}$ and reverberation time $T_{60} = 0.2\,\mathrm{s}$. The hyperboloid-based appearance described above can be verified in the conventional SRP map in Fig. 4. We observe that conventional SRP correctly localizes the source (i.e. the estimated [△] coincides with the true [▽] source location). The LR-SRP and the SLRI-SRP maps are less spatially variant and appear to be defined by only a subset of the hyperboloids apparent in the conventional SRP map. Most of the $P = 6$ hyperboloids due to the source cannot be visually identified anymore and seem to be lost through the low-rank approximation. As a consequence, the maximum of the SRP maps does not correspond to the source location, resulting in erroneous localization. The SSPI-SRP map in contrast agrees very well with the conventional SRP map, resulting in equally accurate localization.

Fig. 5 shows an FF example of a conventional SRP map as well as the corresponding LR-SRP, SLRI-SRP, and SSPI-SRP maps with relative complexity $C_{\mathrm{rel}} \approx 1 \cdot 10^{-2}$ at $SNR = 0\,\mathrm{dB}$ and reverberation time $T_{60} = 0.4\,\mathrm{s}$. The FF maps are determined by the asymptotic behaviour of the hyperboloids of constant TDOAs, which are transformed to angular coordinates. Again, we observe that conventional SRP and SSPI-SRP correctly localize the source, and that the LR-SRP and the SLRI-SRP maps appear to be defined by only a subset of the hyperboloids apparent in the conventional SRP map. While LR-SRP fails to localize the source, the SLRI-SRP map preserves a few more features of the conventional SRP map and achieves accurate localization, although the maximum of the map is much less pronounced.

The appearance of the LR-SRP, SLRI-SRP, and SSPI-SRP maps can be interpreted as follows. Due to the nature of low-rank approximations, which approximate matrices by a truncated sum of outer products, LR-SRP and SLRI-SRP introduce additional dependencies between the SRP values across different candidate locations. More precisely, we note that the SRP vectors $\mathbf{z}_{\mathrm{lr}}$ and $\mathbf{z}_{\mathrm{slri}}$ are forced to lie in the (real part of the) subspace spanned by the columns of $\mathbf{H}_{\mathrm{tall}} \in \mathbb{C}^{J \times R_H}$ and $\boldsymbol{\Lambda}_{\mathrm{tall}} \in \mathbb{R}^{J \times R_\Lambda}$, respectively, cf. (27) and (55). However, restricting the subspace wherein the SRP vector resides conflicts with the ideal of a spatial map wherein the power in each candidate location would be influenced by a source in its close vicinity, that is where each source location is represented by a

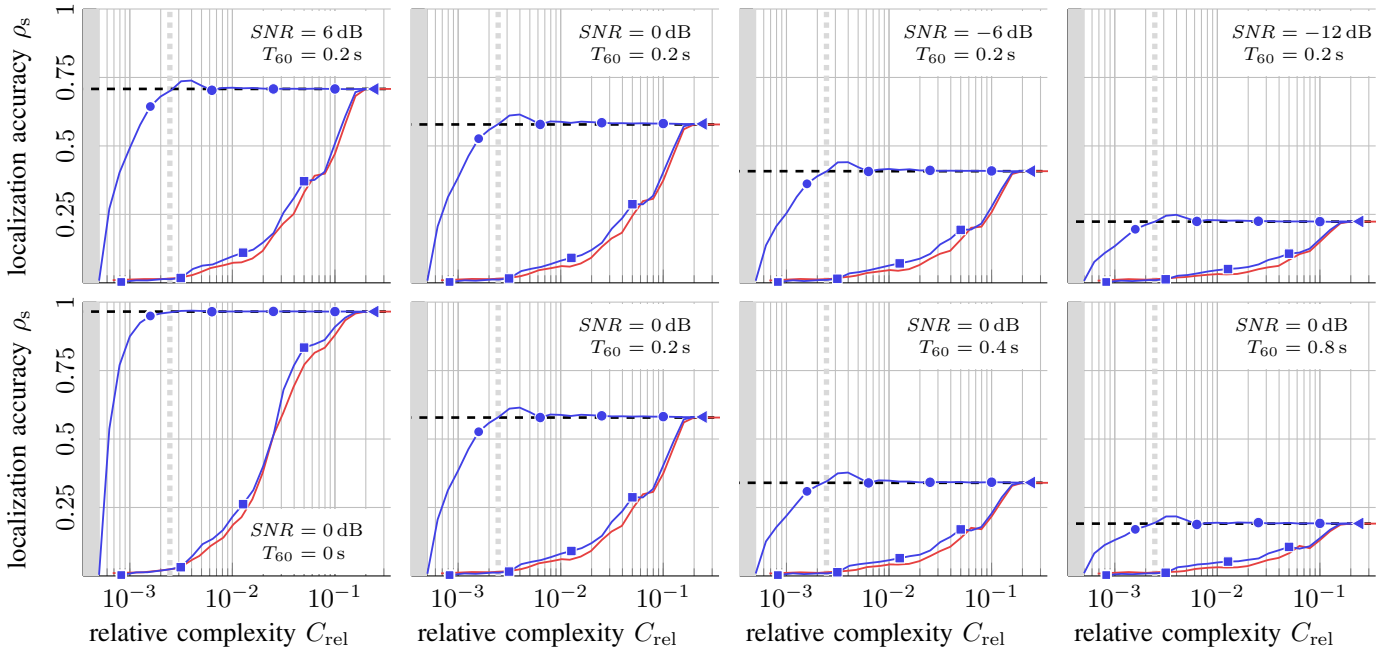


Fig. 8: Mean localization accuracy ρ_s vs. relative complexity C_{rel} of LR-SRP [——], SI-SRP (single point) [◂], SLRI-SRP (proposed) [—▪—], and SSPI-SRP (proposed) [—•—] as compared to ρ_s of conv. SRP [- - -] for different SNR values and reverberation times T_{60} in the NF scenario. From left to right, the top and the bottom row respectively illustrate the effect of decreasing SNR and increasing T_{60} (with the second column showing the same combination twice). The gray area indicates the range where $C_{\mathrm{rel}} < C_{\mathrm{samp}}$, which cannot be reached by SLRI-SRP and SSPI-SRP. The vertical dotted grid line indicates C_{rel} of SSPI-SRP for $Q_\Lambda = JP$, i.e. using on average one sample per candidate location and microphone pair for interpolation.

.

standard basis vector in a J-dimensional vector space. We can therefore argue that the low-rank approximations in LR-SRP and SLRI-SRP might be useful to the extent that the TDOA-based nature of power computation itself forces dependencies across candidate locations for a given geometry, but are likely to have a deteriorating effect beyond. SSPI-SRP on the other hand does not restrict the subspace wherein the SRP vector resides. Moreover, due to the decaying nature of sinc-functions, we can expect many elements of the interpolation matrix $\mathbf{\Lambda}$ in (40)–(41) to be negligible, which lends it to sparse approximation.

2) Matrix and Map Error: Fig. 6 and 7 show the matrix error ε_H and the map error ε_z vs. relative complexity C_{rel} for LR-SRP [——], SI-SRP [◂], SLRI-SRP [—▪—], and SSPI-SRP [—•—] in the NF and FF scenario, respectively. While ε_H can be computed before run time, ε_z is averaged over all 512 source locations, 32 frames, and 16 combinations of SNR and T_{60}. In both the NF and the FF scenario, the characteristics of ε_H and ε_z widely agree across the different algorithms, such that in practice it suffices to compute ε_H in order to judge ε_z.

In both scenarios, LR-SRP can save a bit more or less than one order of magnitude of computational complexity without introducing significant errors. However, the errors increase rapidly at lower relative complexities. Unlike LR-SRP, the complexity of SLRI-SRP and SSPI-SRP is restricted to the range between C_{samp} and C_{rel} of SI-SRP, which spans almost three decades in the NF scenario and one decade in the FF scenario. As compared to LR-SRP, SLRI-SRP performs similarly in the NF scenario and somewhat better in the FF scenario. The better performance of SLRI-SRP in the FF scenario may be due to the more effective SI-SRP approximation as compared to the NF scenario, which leaves room for a less aggressive low-rank approximation of the interpolation matrix. However, in both scenarios, LR-SRP and SLRI-SRP are significantly outperformed by SSPI-SRP over a large range of C_{rel}. These results confirm the discussion provided in Sec. V-C1.

3) Localization Accuracy: For the NF and the FF scenario, Fig. 8 and Fig. 9 respectively illustrate the localization accuracy ρ_s vs. relative complexity C_{rel} of LR-SRP [——], SI-SRP (single point) [◂], SLRI-SRP [—▪—], and SSPI-SRP [—•—] as compared to ρ_s of conventional SRP [- - -] for different SNR values and reverberation times T_{60}. In each subplot, the mean of ρ_s is taken over all 512 source locations and 32 frames. While the absolute performance of the different SRP algorithms largely depends on the SNR and T_{60}, their relative performance exhibits the same trends across different conditions. As can be seen, the NF scenario is more sensitive to T_{60}, cf. also Table II.

As compared to conventional SRP, we observe that the perfomance of LR-SRP quickly declines for $C_{\mathrm{rel}} < 2 \cdot 10^{-1}$ and $C_{\mathrm{rel}} < 8 \cdot 10^{-2}$ in the NF and the FF scenario, respectively. As compared to LR-SRP, SLRI-SRP performs similarly in the NF scenario and somewhat better in the FF scenario, where its performance drops for $C_{\mathrm{rel}} < 3 \cdot 10^{-2}$. Again, the better performance of SLRI-SRP in the FF scenario may be due to the more effective SI approximation as compared to the NF scenario. Both LR-SRP and SLRI-SRP are significantly outperformed by SSPI-SRP, again confirming the

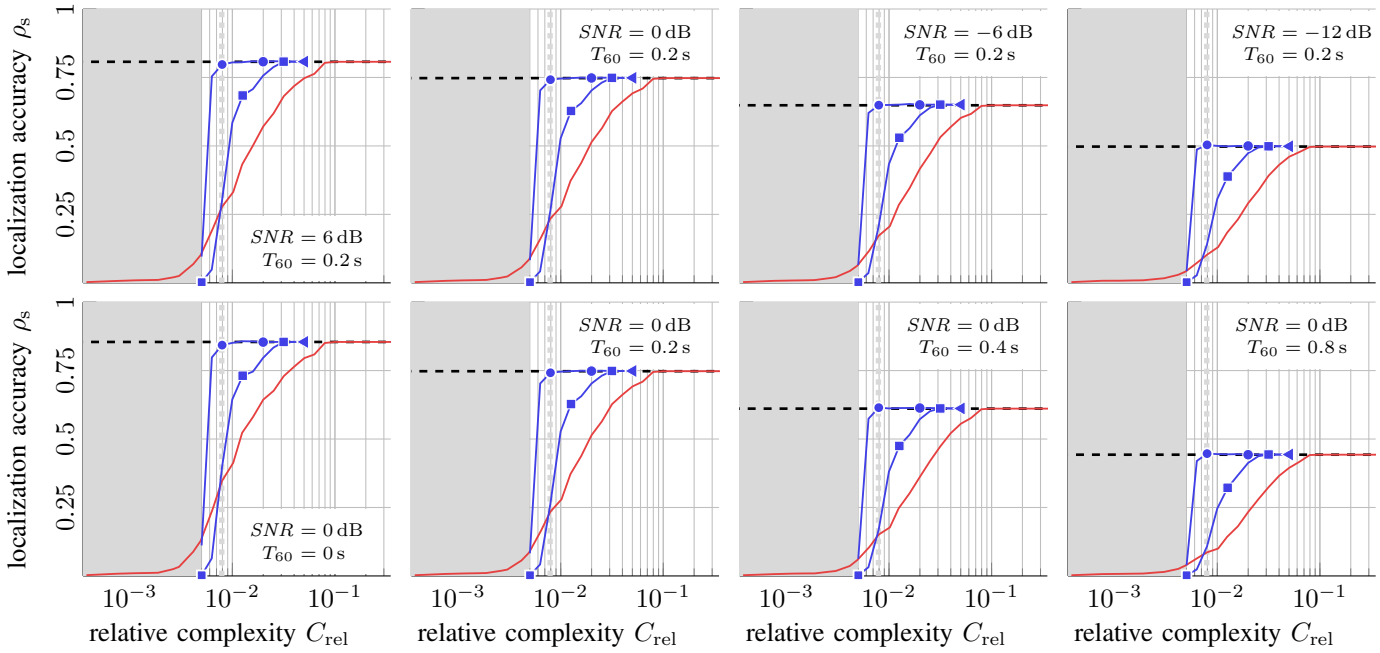


Fig. 9: Mean localization accuracy ρ_s vs. relative complexity C_{rel} of LR-SRP [——], SI-SRP (single point) [◂], SLRI-SRP (proposed) [—▪—], and SSPI-SRP (proposed) [—●—] as compared to ρ_s of conv. SRP [- - -] for different SNR values and reverberation times T_{60} in the FF scenario. From left to right, the top and the bottom row respectively illustrate the effect of decreasing SNR and increasing T_{60} (with the second column showing the same combination twice). The gray area indicates the range where $C_{\text{rel}} < C_{\text{samp}}$, which cannot be reached by SLRI-SRP and SSPI-SRP. The vertical dotted grid line indicates C_{rel} of SSPI-SRP for $Q_\Lambda = 3JP$, i.e. using on average three samples per candidate location and microphone pair for interpolation.

discussion in Sec. V-C1. SSPI-SRP drops off only below $C_{\text{rel}} = 2.3 \cdot 10^{-3}$ and $C_{\text{rel}} = 8 \cdot 10^{-3}$ in the NF and the FF scenario, respectively, which corresponds to $Q_\Lambda = JP$ and $Q_\Lambda = 3JP$, i.e. using on average one and three samples per candidate location and microphone pair for interpolation. The difference in required interpolation accuracy between the NF and the FF scenario may be due to the scenario-dependent relations between candidate location, TDOA, and error metric, cf. (3) and (64). In the NF scenario, we further observe that SSPI-SRP slightly outperforms conventional SRP at around $C_{\text{rel}} = 4 \cdot 10^{-3}$ (corresponding to $Q_\Lambda = 2JP$), indicating that sparse interpolation can even be beneficial to localization performance.

VI. Conclusion

In this paper, we proposed two scalable approaches to construct a full SRP map at reduced computational cost. The two approaches, referred to as sampling + low-rank interpolation-based SRP (SLRI-SRP) and sampling + sparse interpolation-based SRP (SSPI-SRP), are based on a decomposition of the SRP matrix into a sampling and an interpolation matrix and optimal low-rank or sparse approximations of the latter. The proposed approaches are compared to a state-of-the-art low-rank-based SRP (LR-SRP) approach. While SLRI-SRP and LR-SRP constrain the subspace wherein the SRP vector resides, SSPI-SRP exploits the sparse nature of the sinc-based interpolation matrix. Simulation results in a near-field (NF) and a far-field (FF) localization scenario indicate that SSPI-SRP outperforms both SLRI-SRP and LR-SRP over a wide complexity range in terms of approximation error and localization accuracy.

Appendix
Two-Sided Matrix-Vector Formulations

In the Appendix, we introduce two-sided matrix-vector formulations. For the two-sided frequency bin index $k \in \{-K+1, \dots, K\}$, let the index k_{mod} be defined by

$$k_{\text{mod}} = \text{mod}[k, 2K] + 1, \tag{66}$$

yielding $k_{\text{mod}} \in \{1, \dots, 2K\}$. As two-sided counterparts of $\boldsymbol{\psi}_p$, $\mathbf{H}_p$, and $\mathbf{S}_p$ in (18), (19), and (35), respectively, we define $\bar{\boldsymbol{\psi}}_p \in \mathbb{C}^{2K}$, $\bar{\mathbf{H}}_p \in \mathbb{C}^{J \times 2K}$, and $\bar{\mathbf{S}}_p \in \mathbb{C}^{N_p \times 2K}$ as

$$[\bar{\boldsymbol{\psi}}_p]_{k_{\text{mod}}} = \psi_p(\omega_k), \tag{67}$$

$$[\bar{\mathbf{H}}_p]_{i,k_{\text{mod}}} = e^{j\omega_k \Delta t_{p(i)}}, \tag{68}$$

$$[\bar{\mathbf{S}}_p]_{n_{\text{mod}},k_{\text{mod}}} = e^{j\pi kn/K}, \tag{69}$$

Stacking over microphone pairs, as the two-sided counterparts of $\boldsymbol{\psi}$, $\mathbf{H}$, and $\mathbf{S}$ in (20), (21), and (37), we further define $\bar{\boldsymbol{\psi}} \in \mathbb{C}^{2PK}$, $\bar{\mathbf{H}} \in \mathbb{C}^{J \times 2PK}$, and $\bar{\mathbf{S}} \in \mathbb{C}^{PN \times 2PK}$ as

$$\bar{\boldsymbol{\psi}} = \begin{pmatrix} \bar{\boldsymbol{\psi}}_1^{\text{T}} & \cdots & \bar{\boldsymbol{\psi}}_P^{\text{T}} \end{pmatrix}^{\text{T}}, \tag{70}$$

$$\bar{\mathbf{H}} = \begin{pmatrix} \bar{\mathbf{H}}_1 & \cdots & \bar{\mathbf{H}}_P \end{pmatrix}, \tag{71}$$

$$\bar{\mathbf{S}} = \text{blkdiag}\big[\bar{\mathbf{S}}_1, \dots, \bar{\mathbf{S}}_P\big], \tag{72}$$

With these two-sided definitions, $\mathbf{z}$ in (22) can alternatively be expressed by

$$\mathbf{z} = \sum_{p=1}^{P} \boldsymbol{\xi}_{\Delta t_p}$$

$$= \sum_{p=1}^{P} \bar{\mathbf{H}}_p \bar{\boldsymbol{\psi}}_p$$
$$= \bar{\mathbf{H}} \bar{\boldsymbol{\psi}}. \quad (73)$$

Similarly, we can alternatively write $\boldsymbol{\xi}_p = \bar{\mathbf{S}}_p \bar{\boldsymbol{\psi}}_p$ and accordingly express $\boldsymbol{\xi}$ in (38) as

$$\boldsymbol{\xi} = \bar{\mathbf{S}} \bar{\boldsymbol{\psi}}. \quad (74)$$

The correspondence between the two-sided formulations (73) and (74) and the singe-sided formulations (22) and (38) is exact if $\psi_p(0) = 0$ and $\psi_p(\pm\omega_\mathrm{b}) = 0$, cf. Sec. II-D. Note that it can easily be verified that the rows of $\bar{\mathbf{S}}_p$ correspond to the rows of the $2K$-iDFT matrix at indeces n'_mod, with n'_mod as defined in (46). This allows the use of the iFFT in (44)–(45). From this relation to the $2K$-iDFT matrix, we further observe that the rows of $\bar{\mathbf{S}}_p$ are orthogonal such that

$$\bar{\mathbf{S}} \bar{\mathbf{S}}^\mathrm{H} = 2K\mathbf{I}, \quad (75)$$

which is exploited in (51).

ACKNOWLEDGMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven internal funds C24/16/019 "Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking" and VES/19/004, and FWO Large-scale research infrastructure "The Library of Voices - Unlocking the Alamire Foundations Music Heritage Resources Collection through Visual and Sound Technology" (I013218N). The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

REFERENCES

[1] Z. Chen, G. Gokeda, and Y. Yu, *Introduction to Direction-of-arrival Estimation*. Artech House, 2010.
[2] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Springer, 2013.
[3] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, "The LOCATA challenge: Acoustic source localization and tracking," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 1620–1643, 2020.
[4] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.
[5] E. Vincent, T. Virtanen, and S. Gannot, *Audio Source Separation and Speech Enhancement*. Hoboken, New Jersey, USA: Wiley, 2018.
[6] 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.
[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
[8] R. Roy and T. Kailath, "ESPRIT – estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 7, pp. 984–995, 1989.
[9] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, 2011.
[10] J. H. DiBiase, "A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays," Ph.D. dissertation, Brown University, Providence, RI, USA, 2000.
[11] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 499–508, Sep. 2004.
[12] H. F. Silverman, Y. Yu, J. M. Sachar, and W. R. Patterson, "Performance of real-time source-location estimators for a large-aperture microphone array," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 593–606, 2005.
[13] J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, Nov. 2007.
[14] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array," in *Proc. 2007 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2007)*, Honolulu, HI, USA, April 2007, pp. 121–124.
[15] H. Do and H. F. Silverman, "A fast microphone array SRP-PHAT source location implementation using coarse-to-fine region contraction (CFRC)," in *Proc. 2007 IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA 2007)*, New Paltz, NY, USA, Oct. 2007, pp. 295–298.
[16] S. Tervo and T. Lokki, "Interpolation methods for the SRP-PHAT algorithm," in *Proc. 2008 Int. Workshop Acoustic Echo Noise Control (IWAENC 2008)*, Seattle, WA, USA, Sep. 2008, pp. 14–17.
[17] H. Do and H. F. Silverman, "Stochastic particle filtering: A fast SRP-PHAT single source localization algorithm," in *Proc. 2009 IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA 2009)*, New Paltz, NY, USA, Oct. 2009, pp. 213–216.
[18] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Letters*, vol. 18, no. 1, pp. 71–74, Jan. 2011.
[19] A. Marti, M. Cobos, J. J. Lopez, and J. Escolano, "A steered response power iterative method for high-accuracy acoustic source localization," *J. Acoust. Soc. Amer.*, vol. 134, no. 4, pp. 2627–2630, Oct. 2013.
[20] L. O. Nunes, W. A. Martins, M. V. S. Lima, L. W. P. Biscainho, F. M. Gonçalves, A. Said, and B. Lee, "A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5171–5183, Oct. 2014.
[21] M. V. S. Lima, W. A. Martins, L. O. Nunes, L. W. P. Biscainho, T. N. Ferreira, M. V. M. Costa, and B. Lee, "A volumetric SRP with refinement step for sound source localization," *IEEE Signal Process. Letters*, vol. 22, no. 8, pp. 1098–1102, Aug. 2015.
[22] J. Velasco, C. J. Martín-Arguedas, J. Macias-Guarasa, D. Pizarro, and M. Mazo, "Proposal and validation of an analytical generative model of SRP-PHAT power maps in reverberant scenarios," *Signal Processing*, vol. 119, pp. 209–228, Feb. 2016.
[23] D. Salvati, C. Drioli, and G. L. Foresti, "Exploiting a geometrically sampled grid in the steered response power algorithm for localization improvement," *Acoust. Soc. Amer.*, vol. 141, no. 1, pp. 586–601, 2017.
[24] M. B. Çöteli, O. Olgun, and H. Hacıhabiboğlu, "Multiple sound source localization with steered response power density and hierarchical grid refinement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 11, pp. 2215–2229, 2018.
[25] F. Grondin and J. Glass, "SVD-PHAT: A fast sound source localization method," in *Proc. 2019 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2019)*, Brighton, UK, May 2019, pp. 4140–4144.
[26] D. Diaz-Guerra, A. Miguel, and J. R. Beltran, "Robust sound source tracking using SRP-PHAT and 3D convolutional neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 300–311, 2021.
[27] G. García-Barrios, J. M. Gutiérrez-Arriola, N. Sáenz-Lechón, V. J. Osma-Ruiz, and R. Fraile, "Analytical model for the relation between signal bandwidth and spatial resolution in steered-response power phase transform (SRP-PHAT) maps," *IEEE Access*, vol. 9, pp. 121 549–121 560, 2021.
[28] T. Dietzen, E. De Sena, and T. van Waterschoot, "Low-complexity steered response power mapping based on Nyquist-Shannon sampling," in *Proc. 2021 IEEE Workshop Appl. Signal Process. Audio, Acoust. (WASPAA 2021)*, New Paltz, NY, USA, Oct. 2021, pp. 206–210.
[29] B. Çakmak, T. Dietzen, R. Ali, P. A. Naylor, and T. van Waterschoot, "A distributed steered response power approach to source localization in wireless acoustic sensor networks," in *Proc. 2022 Int. Workshop Acoustic Signal Enhancement (IWAENC 2022)*, Sep. 2022, pp. 1–5.
[30] E. Grinstein, M. Brookes, and P. A. Naylor, "Graph neural networks for sound source localization on distributed microphone networks," in *Proc.*

2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2023), Rhodes Island, Greece, June 2023, pp. 1–5.

[31] P.-A. Grumiaux, S. Kitić, L. Girin, and A. Guérin, "A survey of sound source localization with deep learning methods," *Acoust. Soc. Amer.*, vol. 152, no. 1, pp. 107–151, 2022.

[32] T. Dietzen, "GitHub repository: Low-complexity steered response power mapping based on low-rank and sparse interpolation," https://github.com/tdietzen/LR-SP-INT-SRP, June 2023.

[33] C. Zhang, D. Florêncio, and Z. Zhang, "Why does PHAT work well in low noise, reverberative environments?" in *Proc. 2008 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2008)*, Las Vegas, NV, USA, Mar. 2008, pp. 2565–2568.

[34] C. D. Meyer, Ed., *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[35] R. J. I. Marks, *Introduction to Shannon sampling and interpolation theory*. Springer, 2012.

[36] N. Antonello, L. Stella, P. Patrinos, and T. van Waterschoot, "Proximal gradient algorithms: Applications in signal processing," *arXiv eprint 1803.01621*, pp. 1–30, Jan. 2020, https://arxiv.org/abs/1803.01621.

[37] C. Veaux, J. Yamagishi, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," http://homepages.inf.ed.ac.uk/jyamagis/page3/page58/page58.html, 2016.

[38] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Amer.*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

[39] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYRiAD: A multi-array room acoustic database," *EURASIP J. Audio Speech Music Process.*, vol. 2023, no. 17, pp. 1–14, Apr. 2023.

[40] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 4, pp. 774–786, Apr. 2015.

[41] H. Kuttruff, *Room Acoustics*, 5th ed. SPON Press, Apr. 2014.