

FRIDA: FRI-BASED DOA ESTIMATION FOR ARBITRARY ARRAY LAYOUTS

Hanjie Pan[†], Robin Scheibler[†], Eric Bezzam[†], Ivan Dokmanic[‡], and Martin Vetterli[†]

[†]École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

[‡]Institut Langevin, CNRS, ESPCI Paris, PSL Research University, France
{firstname.lastname}@epfl.ch, ivan.dokmanic@espci.fr

ABSTRACT

In this paper we present FRIDA—an algorithm for estimating directions of arrival of multiple wideband sound sources. FRIDA combines multi-band information coherently and achieves state-of-the-art resolution at extremely low signal-to-noise ratios. It works for arbitrary array layouts, but unlike the various steered response power and subspace methods, it does not require a grid search. FRIDA leverages recent advances in sampling signals with a finite rate of innovation. It is based on the insight that for any array layout, the entries of the spatial covariance matrix can be linearly transformed into a uniformly sampled sum of sinusoids.

Index Terms—Direction of arrival, finite rate of innovation, subspace method, search-free, wideband sources

1. INTRODUCTION

A wishlist for a direction of arrival (DOA) estimator may look something like this: it should be high-resolution, work at low signal-to-noise ratios (SNRs), resolve many possibly closely spaced sources, work with few arbitrarily laid out microphones, and do so efficiently, without grid searches.

It is uncommon to have all of these items checked at once. For example, the steered response power (SRP) methods [1] can be made robust, do not require a specific array geometry, and are immune to coherence in signals. Nevertheless, because they are based on beamforming, they cannot resolve close sources [2].

Close sources can be resolved by the high-resolution DOA finders. Their main representatives are subspace methods such as MUSIC [3], Prony-type methods such as root-MUSIC [4], and methods that attempt to compute the maximum likelihood (ML) estimator such as IQML [5].

Subspace methods exploit the fact that for uncorrelated signal and noise, the eigenspace of the spatial covariance matrix corresponding to largest eigenvalues is spanned by the source steering vectors [3]. These methods are fundamentally narrowband since the signal subspaces vary with frequency; they can be made wideband either by incoherently combining narrowband

estimates or, better, by combining them coherently through transforming the array manifold at each frequency to a manifold at a reference frequency (CSSM [6], WAVES [7]). These methods require a search over space unless the array is a uniform linear array (ULA) [8]. Coherent methods also require special “focusing matrices”, essentially initial guesses of the source locations. WAVES can do without focusing but at the cost of performance. In between coherent and incoherent methods is the TOPS algorithm [9], which performs well at mid-SNRs, but still requires a search and performs worse than coherent methods at low SNRs.

We propose a new finite rate of innovation (FRI) sampling-based algorithm for DOA finding—FRIDA. Among the mentioned algorithms, FRIDA is most reminiscent of IQML [5], especially for narrowband signals and ULAs. Unlike IQML, FRIDA works for wideband signal and arbitrary sensor geometries by naturally accommodating array interpolation in a way reminiscent of manifold separation [10, 11]. Moreover, it uses multi-band information coherently. Still, it requires no grid search and no sensitive preprocessing akin to focusing matrices, and it achieves very high resolution at very low SNRs, outperforming previous state-of-the-art.

FRIDA can work with fewer microphones than sources as it uses cross-correlations instead of raw microphone streams. The tradeoff is that it is not able to handle completely correlated signals. A straightforward modification of the algorithm which operates on raw signals rather than cross-correlations does not have this issue, but it requires more microphones.

The main ingredient of FRIDA is an FRI sampling algorithm [12]. FRI sampling has recently been extended to non-uniform grids along with a robust reconstruction algorithm [13]. The algorithm is an iterative algorithm similar to IQML, but with an added spectral resampling layer and a modified stopping criterion (Section 2.2.3). The key insight is that the elements of the spatial correlation matrix can be *linearly* transformed into uniformly sampled sums of sinusoids, regardless of the array geometry.

2. PROBLEM FORMULATION

Throughout the paper, matrices and vectors are denoted by bold upper and lower case letters. The Euclidean norm of a vector \mathbf{x} is denoted by $\|\mathbf{x}\|_2 = (\mathbf{x}^H \mathbf{x})^{1/2}$, where $(\cdot)^H$ is the Hermitian transpose. We denote by \mathbb{S} the unit circle. Unit propagation vectors will be denoted by $\mathbf{p} = [\cos \varphi, \sin \varphi]^T$, where φ is the corresponding azimuth.

This work was supported by the Swiss National Science Foundation grant 20FP-1.151073 — Inverse problems regularized by sparsity. ID was funded by LABEX WIFI (Laboratory of Excellence within the French Program “Investments for the Future”) under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL* and by Agence Nationale de la Recherche under reference ANR-13-JS09-0001-01. All the code used to produce the results of this paper is available at <http://github.com/LCAV/FRIDA>.

2.1. Source Signal and Measurements

We assume a setup with Q microphones located at $\{\mathbf{r}_q \in \mathbb{R}^2\}_{q=1}^Q$, and K monochromatic and uncorrelated point sources in the far-field indexed by the letter k . Each propagates in the direction of the unit vector $\mathbf{p}_k = [\cos \varphi_k, \sin \varphi_k]^\top$, where φ_k is the azimuth of the k -th source. Within a narrow band centered at frequency ω , the baseband representation of the signal coming from direction $\mathbf{p} \in \mathbb{S}$ reads $x(\mathbf{p}, \omega, t) = \tilde{x}(\mathbf{p}, \omega) e^{j\omega t}$, where $\tilde{x}(\mathbf{p}, \omega)$ is the emitted sound signal by a source located at \mathbf{p} and frequency ω . The *intensity* of the sound field is

$$I(\mathbf{p}, \omega) \stackrel{\text{def}}{=} \mathbb{E} [|x(\mathbf{p}, \omega, t)|^2] = \sum_{k=1}^K \sigma_k^2(\omega) \delta(\mathbf{p} - \mathbf{p}_k),$$

where $\sigma(\omega)_k^2$ is the power of the k -th source and $\delta(\mathbf{p})$ is the Dirac delta function on \mathbb{S} . We assume frame-based processing, and the expectation is over the randomness of \tilde{x} from frame to frame. As \tilde{x} carries the phase, the assumption $\mathbb{E}[\tilde{x}] = 0$ holds.

The received signal at the q -th microphone located at \mathbf{r}_q is the integration of all plane waves along the unit circle:

$$y_q(\omega, t) = \int_{\mathbb{S}} x(\mathbf{p}, \omega, t) e^{-j\omega \langle \mathbf{p}, \frac{\mathbf{r}_q}{c} \rangle} d\mathbf{p} \quad \text{for } q = 1, \dots, Q,$$

where c is the speed of sound. In this paper, we will take as measurements the cross-correlations¹ between the received signals for a microphone pair (q, q') :

$$V_{q,q'}(\omega) \stackrel{\text{def}}{=} \mathbb{E} [y_q(\omega, t) y_{q'}^*(\omega, t)] \quad (1)$$

for $q, q' \in [1, Q]$ and $q \neq q'$. In practice, $V_{q,q'}$ is estimated by averaging over frames; for simplicity we use the same symbol for the empirical version. Since we assume uncorrelated sources, the cross-correlation reduces to:

$$\begin{aligned} V_{q,q'}(\omega) &= \int_{\mathbb{S}} \int_{\mathbb{S}} \mathbb{E} [x(\mathbf{p}, t) x^*(\mathbf{p}', t)] e^{-j\omega \langle \mathbf{p}, \frac{\mathbf{r}_q}{c} \rangle} e^{j\omega \langle \mathbf{p}', \frac{\mathbf{r}_{q'}}{c} \rangle} d\mathbf{p} d\mathbf{p}' \\ &= \int_{\mathbb{S}} I(\mathbf{p}, \omega) e^{-j\omega \langle \mathbf{p}, \Delta \mathbf{r}_{q,q'} \rangle} d\mathbf{p} = \sum_{k=1}^K \sigma_k^2(\omega) e^{-j\omega \langle \mathbf{p}_k, \Delta \mathbf{r}_{q,q'} \rangle}, \end{aligned} \quad (2)$$

where $\Delta \mathbf{r}_{q,q'} \stackrel{\text{def}}{=} \frac{\mathbf{r}_q - \mathbf{r}_{q'}}{c}$ is the normalized baseline. Instead of the more conventional approach to FRI sampling where the microphone signals y_q would be used as input [14], we use the correlations $V_{q,q'}$. This effectively increases the number of measurements and allows us to use a small number of microphones.

2.2. Point Source Reconstruction

Following the generalized FRI sampling framework [13], we will first identify the set of unknown sinusoidal samples and its relation with the given measurements (1). Then, the DOA estimation is cast as a constrained optimization, cf. (6).

¹ Alternatively, we can estimate the DOA directly from the received microphone signals $y_q(\omega, t)$. We leave the detailed discussions for a future publication.

2.2.1. Relation between Measurements and the Uniform Samples of Sinusoids

Since \mathbf{p} is supported on the circle, the intensity of (2.1) can be equally written as a Fourier series expansion:

$$I(\mathbf{p}, \omega) = \sum_{m \in \mathbb{Z}} \hat{I}_m(\omega) Y_m(\mathbf{p}),$$

where $Y_m(\mathbf{p})$ is the Fourier series basis $Y_m(\mathbf{p}) = Y_m(\varphi) = e^{jm\varphi}$, and $\hat{I}_m(\omega)$ is the associated expansion coefficient for a sub-band centered at frequency ω :

$$\hat{I}_m(\omega) = \frac{1}{2\pi} \int_{\mathbb{S}} I(\mathbf{p}, \omega) Y_m^*(\mathbf{p}) d\mathbf{p} = \frac{1}{2\pi} \sum_{k=1}^K \sigma_k^2(\omega) e^{-jm\varphi_k}. \quad (3)$$

Notice that the Fourier series coefficients $\hat{I}_m(\omega)$ for $m \in \mathbb{Z}$ are uniform samples of sinusoids, which are related with the cross-correlation (2) as:

$$\begin{aligned} V_{q,q'}(\omega) &= \int_{\mathbb{S}} \sum_{m \in \mathbb{Z}} \hat{I}_m(\omega) Y_m(\mathbf{p}) e^{-j\omega \langle \mathbf{p}, \Delta \mathbf{r}_{q,q'} \rangle} d\mathbf{p} \\ &\stackrel{(a)}{=} 2\pi \sum_{m \in \mathbb{Z}} (-j)^m J_m(\|\omega \Delta \mathbf{r}_{q,q'}\|_2) Y_m\left(\frac{\Delta \mathbf{r}_{q,q'}}{\|\Delta \mathbf{r}_{q,q'}\|_2}\right) \hat{I}_m(\omega) \end{aligned} \quad (4)$$

where (a) is from Jacobi-Anger expansion [15] of the complex exponential and $J_m(\cdot)$ is Bessel function of the first kind.

Therefore, we establish a linear mapping from the uniformly sampled sinusoids \hat{I}_m to the given measurements $V_{q,q'}$. Concretely, denote a lexicographically ordered vectorization of the cross-correlations $V_{q,q'}(\omega)$, $q \neq q'$ by $\mathbf{a}(\omega) \in \mathbb{C}^{Q(Q-1)}$, and let the vector $\mathbf{b}(\omega)$ be the Fourier series coefficients $\hat{I}_m(\omega)$ for $m \in \mathcal{M}$, where \mathcal{M} is a set of considered Fourier coefficients². Define also a $Q(Q-1) \times |\mathcal{M}|$ matrix $\mathbf{G}(\omega)$ as

$$g_{(q,q'),m}(\omega) \stackrel{\text{def}}{=} (-j)^m J_m(\|\omega \Delta \mathbf{r}_{q,q'}\|_2) Y_m\left(\frac{\Delta \mathbf{r}_{q,q'}}{\|\Delta \mathbf{r}_{q,q'}\|_2}\right), \quad (5)$$

where rows of \mathbf{G} are indexed by microphone pairs (q, q') , and columns of \mathbf{G} are indexed by Fourier bins m . We can then concisely write (4) as $\mathbf{a}(\omega) = \mathbf{G}(\omega) \mathbf{b}(\omega)$.

2.2.2. Annihilation on the Circle

Since \hat{I}_m in (3) is a weighted sum of uniformly sampled sinusoids, we know that \hat{I}_m should satisfy a set of annihilation equations [12]: $\hat{I}_m * h_m = 0$. Here h_m is the unknown annihilating filters to be recovered. A polynomial, whose coefficients are specified by the filter h_m , has roots located at $e^{-j\varphi_k}$ [12]. The source azimuths φ_k are subsequently reconstructed with polynomial root-finding.

In a multi-band setting, the uniform sinusoidal samples $\hat{I}_m(\omega)$ are different for each sub-band. This is because the signal power σ_k^2 varies with the mid-band frequency ω in general. However, since we have *the same* source locations φ_k for each sub-band, we only need to find one filter h_m (depending solely on the source locations φ_k) that annihilates $\hat{I}_m(\omega)$ for all ω -s:

$$\hat{I}_m(\omega) * h_m = 0 \quad \forall \omega.$$

²Note that these correspond to the spatial Fourier transform of I over the circle, not to sources' temporal spectra.

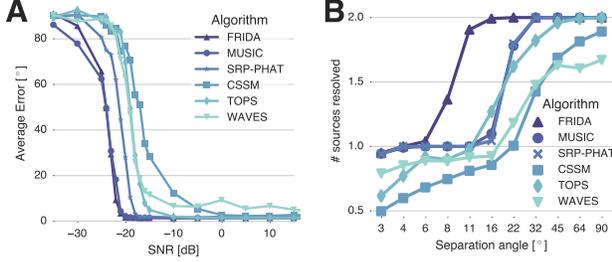


Fig. 1. **A** Average DOA reconstruction error as a function of SNR. Lower is better. **B** Average number of sources reconstructed for the case of two sources separated by a fixed angle.

2.2.3. Reconstruction Algorithm

Following the discussion in the previous section, we reconstruct the source locations jointly across all sub-bands. More specifically, suppose we consider J sub-bands centered around frequencies $\{\omega_j\}_{j=1}^J$. Then, we formulate the FRIDA estimate as a solution of the following constrained optimization:

$$\min_{\mathbf{h} \in \mathcal{H}} \sum_{i=1}^J \|\mathbf{a}_i - \mathbf{G}_i \mathbf{b}_i\|_2^2 \quad (6)$$

subject to $\mathbf{b}_i * \mathbf{h} = \mathbf{0}$ for $i = 1, \dots, J$,

Here \mathbf{a}_i , \mathbf{b}_i and \mathbf{G}_i are the cross-correlation, uniform sinusoidal samples, and the linear mapping between them for the i -th sub-band (5); \mathcal{H} is a feasible set that the annihilating filter coefficients belong to, e.g., $\mathcal{H} = \{\mathbf{h} \in \mathbb{C}^{K+1} : \|\mathbf{h}\|_2 = 1\}$.

Note that (6) is a simple quadratic minimization with respect to \mathbf{b}_i -s for a given annihilating filter \mathbf{h} . By substituting the solution of \mathbf{b}_i (in function of \mathbf{h}), we end up with an optimization for \mathbf{h} alone:

$$\min_{\mathbf{h} \in \mathcal{H}} \mathbf{h}^H \mathbf{\Lambda}(\mathbf{h}) \mathbf{h}, \quad (7)$$

where

$$\mathbf{\Lambda}(\mathbf{h}) = \sum_{i=1}^J \mathbf{T}^H(\beta_i) \left[\mathbf{R}(\mathbf{h}) (\mathbf{G}_i^H \mathbf{G}_i)^{-1} \mathbf{R}^H(\mathbf{h}) \right]^{-1} \mathbf{T}(\beta_i).$$

Here $\beta_i = (\mathbf{G}_i^H \mathbf{G}_i)^{-1} \mathbf{G}_i^H \mathbf{a}_i$; $\mathbf{T}(\cdot)$ builds a Toeplitz matrix from the input vector; and $\mathbf{R}(\cdot)$ is the *right-dual* matrix associated with $\mathbf{T}(\cdot)$ such that $\mathbf{T}(\mathbf{h})\mathbf{b} = \mathbf{R}(\mathbf{h})\mathbf{b}$, $\forall \mathbf{b}, \mathbf{h}$. This follows from the commutativity of convolution: $\mathbf{b} * \mathbf{h} = \mathbf{h} * \mathbf{b}$.

In general, it is challenging to solve (7) directly. We use an iterative strategy, building $\mathbf{\Lambda}(\mathbf{h})$ with the reconstructed \mathbf{h} from the previous iteration. However, unlike similar approaches (e.g. [5]), we do not aim at obtaining a convergent solution of (7) but rather a valid solution such that the reconstructed sinusoidal samples \mathbf{b}_i -s explain the given measurements up to a certain approximation level (ϵ^2): $\sum_{i=1}^J \|\mathbf{a}_i - \mathbf{G}_i \mathbf{b}_i\|_2^2 \leq \epsilon^2$. Readers are referred to [13] for detailed discussions on the algorithmic details, e.g., choice of ϵ , implementation details, etc. We summarize the iterative algorithm in Algorithm 1.

3. EXPERIMENTS

In this section, we demonstrate the effectiveness of the proposed algorithm through numerical simulations and practical experiments. We compare the performance of FRIDA to that of other

Algorithm 1: FRIDA: FRI-based DOA estimation

Input : cross-correlation of the microphone signals \mathbf{a}_i , transformation matrix \mathbf{G}_i , noise level ϵ^2

Output: uniform sinusoidal samples \mathbf{b}_i , annihilating filter coefficients \mathbf{h}

```

for loop  $\leftarrow$  1 to max. initializations do
1   Initialize  $\mathbf{h}$  with a random vector  $\mathbf{h}^{(0)}$ ;
   for  $n \leftarrow$  1 to max. iterations do
2     Build  $\mathbf{\Lambda}(\mathbf{h})$  with  $\mathbf{h} = \mathbf{h}^{(n-1)}$  and update  $\mathbf{h}^{(n)}$ 
       by solving (7);
3     Re-synthesize  $\mathbf{b}_i^{(n)}$  with the updated annihilating
       filter  $\mathbf{h} = \mathbf{h}^{(n)}$  as:
       
$$\mathbf{b}_i(\mathbf{h}) = \beta_i - (\mathbf{G}_i^H \mathbf{G}_i)^{-1} \mathbf{R}^H(\mathbf{h})$$

       
$$\cdot (\mathbf{R}(\mathbf{h}) (\mathbf{G}_i^H \mathbf{G}_i)^{-1} \mathbf{R}^H(\mathbf{h}))^{-1} \mathbf{R}(\mathbf{h}) \beta_i$$

       if  $\sum_{i=1}^J \|\mathbf{a}_i - \mathbf{G}_i \mathbf{b}_i^{(n)}\|_2^2 \leq \epsilon^2$  then
4       | Terminate both loops;
       end
   end
   end
5  $\mathbf{b}_i \leftarrow \mathbf{b}_i^{(n)}$ ,  $\mathbf{h} \leftarrow \mathbf{h}^{(n)}$ .

```

wideband algorithms: incoherent MUSIC [3], SRP-PHAT [2], CSSM [6], WAVES [7], and TOPS [9].

The sampling frequency is fixed at 16 kHz. The narrow-band sub-carriers are extracted by a 256-point short-time Fourier transform (STFT) with a Hanning window and no overlap. We use a triangular array of 24 microphones. Each edge is 30 cm long and carries 8 microphones. The spacing of microphones ranges from 8 mm to 25 cm. This geometry is that of the *Pyramic* compact array designed at EPFL [16] and used to collect the recordings for the practical experiments, see Fig. 2A.

The number of frequency bands used (out of the 128 narrow-bands) is a key parameter for performance and was tuned for each algorithm. FRIDA, MUSIC and SRP-PHAT use 20 bands, CSSM and WAVES 10 bands, and TOPS 60 bands. In the synthetic experiments, the source signals are all white noise to simplify the choice of the sub-bands. For speech recordings, the STFT bins with the largest power are chosen. All implementation details are in the supplementary material.

The reconstruction errors are quantified according to the distance on the unit circle defined as

$$d_s(\varphi, \hat{\varphi}) = \min_{s \in \{\pm 1\}} s(\varphi - \hat{\varphi}) \bmod 2\pi. \quad (8)$$

For multiple DOA, the originals and their reconstructions are matched to minimize the sum of errors.

3.1. Influence of Noise

We study the influence of noise on the algorithms through numerical simulation. One source playing white noise is placed at random on the unit circle. The propagation of sound is simulated by applying fractional delay filters to generate the microphone signals based on the array geometry. Finally, the algorithms are run with additive white Gaussian noise of variance corresponding to a wide range of SNR. The algorithms are fed with 256

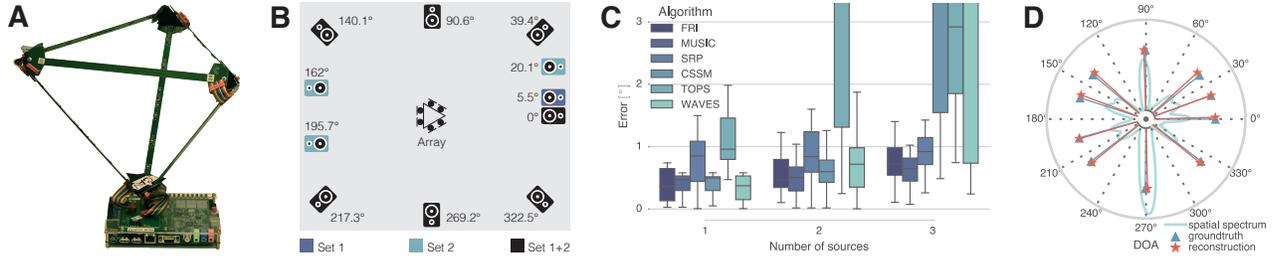


Fig. 2. **A** Pyramic array, a compact microphone array with 48 MEMS microphones distributed on the edges of a tetrahedron. For the experiments, only the top triangle is used. **B** Locations of the loudspeakers and microphone array in experiments. **C** Reconstruction error for the different algorithms applied to the recorded speech signals. **D** Reconstruction of 10 sources from only 9 microphones. The average reconstruction error is within 2° .

DOA	FRIDA	MUSIC	SRP-PHAT
0°	$-0.5 \pm 0.4^\circ$	$1.6 \pm 0.3^\circ$	$1.4 \pm 0.2^\circ$
5.5°	$4.6 \pm 0.2^\circ$	$-93.9 \pm 41.2^\circ$	$-38.1 \pm 8.6^\circ$

Table 1. The accuracy of the reconstruction for recordings with sources closely located at 0° and 5.5° . The mean is computed as the logarithm of the average of complex exponentials with argument given by the reconstruction angle. The second number is the average distance (8) from the sample to the mean.

snapshots of 256 samples each. It should be noted that 256 snapshots correspond to a processing gain of about 24 dB. We run 500 rounds of Monte-Carlo simulation for each SNR value.

The simulation results in Fig. 1A show that FRIDA and MUSIC are the most robust with a breaking points slightly below -20 dB. Next are SRP-PHAT and TOPS, breaking around 2 dB and 4 dB higher, respectively. While WAVES initially seems to perform as well as TOPS, it never reaches zero error. Least resistant to noise is CSSM, breaking down as early as -5 dB. The poor performance of WAVES and CSSM might be attributed to poor initial estimates of the focusing frequencies.

3.2. Resolving Close Sources

Next, we study the minimum angle of separation necessary to resolve distinct sources. We simulate two sources of white noise at angles φ and $\varphi + \delta$ where δ is varied from 90° to 2.8° . The average error is then computed over ten realizations of the noise for 120 values of φ . We mark a DOA as successfully recovered if the reconstruction error is less than $\delta/2$. This criterion is crude for large δ , but for small δ , where performance is critical, it is stringent. We use 256 snapshots and the SNR is set to 0 dB.

As seen in Fig. 1B, we find that FRIDA largely outperforms the other algorithms. It always separates sources located as close as 11.2° , while the closest contenders, MUSIC and SRP-PHAT, have difficulties for sources closer than 22.5° . The coherent methods perform worse than the incoherent ones; they even suffer from a lack of precision in estimating a single source.

3.3. Experiments on Recorded Signals

Finally, we perform two experiments with recorded data to validate the algorithm in non-ideal, real-world conditions. In the first experiment, the Pyramic array is placed at the center of eight loudspeakers (Fig. 2B, Set 1). All the loudspeakers are between 1.45 m and 2.45 m away from the array. Recordings are made with all possible combinations of one, two, and three speakers

playing simultaneously (distinct) speech segments of 3 to 4 seconds duration. Two of the speakers are located at 5.5° of each other to test the resolving power of the algorithms.

The statistics of the reconstruction errors for the different algorithms are shown in Fig. 2C. We find the coherent methods WAVES and CSSM to perform well for one and two sources, but break down for three sources. The TOPS method maintains an acceptable but somewhat imprecise performance for more than one source. FRIDA, MUSIC and SRP-PHAT perform best with a median error within one degree from the ground truth. Where FRIDA distinguishes itself from the conventional methods is for closely spaced sources. This is highlighted in Table 1 where the average reconstructed DOA for the sources located at 0° and 5.5° is shown. While all three methods correctly identify the first source, only FRIDA is able to resolve the second.

The second experiment tests the ability of FRIDA to resolve more sources than microphones are used. We place ten loudspeakers (Fig. 2B, Set 2) around the Pyramic array and record them simultaneously playing white noise. Then, we discard the signals of all but nine microphones and run FRIDA. The algorithm successfully reconstructs all DOA within 2° of the ground truth, as shown in Fig. 2D. None of the subspace methods can achieve this result. While SRP-PHAT is not limited in this way, its resolution is lower (its error is $\sim 4^\circ$ on this recording).

4. CONCLUSION

We introduced FRIDA, a new algorithm for DOA estimation of sound sources. FRIDA relies on finite rate of innovation sampling to do so efficiently on arbitrary array geometries, avoiding any costly grid search. Its ability to use wideband signal information makes it robust to many types of noise and interference. We demonstrate that FRIDA compares favorably to the state-of-the-art, and clearly outperforms all other algorithms when it comes to resolving close sources. Moreover, FRIDA is notable for resolving more sources than microphones, as demonstrated experimentally on recorded signals. Besides the logical extension to the full spherical case, we want to extend the algorithm to work on plane waves directly, rather than the cross-correlation coefficients. This will allow the algorithm to handle correlated as well as uncorrelated sources. Finally, it is important for practical purposes to improve the computational complexity of the algorithm.

Acknowledgement We are indebted to Juan Azcarreta Ortiz and René Beuchat for their help with the Pyramic array.

5. REFERENCES

- [1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," in *Proc. IEEE*, 1969, pp. 1408–1418.
- [2] 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, 2000.
- [3] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
- [4] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, no. 1, pp. 15–29, 1993.
- [5] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, Oct. 1986.
- [6] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.
- [7] E. D. di Claudio and R. Parisi, "WAVES: weighted average of signal subspaces for robust wideband direction finding," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2179–2191, Oct. 2001.
- [8] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE ICASSP*, vol. 8. IEEE, 1983, pp. 336–339.
- [9] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "TOPS: new DOA estimator for wideband signals," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 1977–1989, May 2006.
- [10] F. Belloni, A. Richter, and V. Koivunen, "DoA estimation via manifold separation for arbitrary array structures," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4800–4810, Sep. 2007.
- [11] M. Rubsamen and A. B. Gershman, "Direction-of-Arrival estimation for nonuniform sensor arrays: From manifold separation to Fourier domain MUSIC methods," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 588–599, Jan. 2009.
- [12] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, 2002.
- [13] H. Pan, T. Blu, and M. Vetterli, "Towards generalized FRI sampling with an application to source resolution in radioastronomy," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 821–835, Feb 2017.
- [14] P. J. Hayuningtyas and P. Marziliano, "Finite rate of innovation method for DOA estimation of multiple sinusoidal signals with unknown frequency components," *Radar Conference (EuRAD)*, pp. 115–118, 2012.
- [15] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*. Springer Science & Business Media, 2012, vol. 93.
- [16] J. Azcarreta Ortiz, "Pyramic array: An FPGA based platform for many-channel audio acquisition," Master's thesis, EPFL, Lausanne, Switzerland, Aug. 2016.