A NOVEL SPARSE MODEL FOR MULTI-SOURCE LOCALIZATION USING DISTRIBUTED MICROPHONE ARRAY

Thi Ngoc Tho Nguyen⋆ *Cagdas Tuna*⋆ *Shengkui Zhao*⋆ *Douglas L. Jones*⋆† *

⋆ Advanced Digital Science Center (ADSC),
Illinois at Singapore, 138632, Singapore
† Department of Electrical and Computer Engineering,
University of Illinois at Urbana-Champaign, IL 61801, United States

ABSTRACT

When distances between microphone pairs are larger than the half-wavelength of signals, source localization methods using cross-correlation such as time-difference-of-arrival (TDOA), steered response power (SRP) are commonly used in practice. We present here a novel model that expresses microphone pairwise cross-correlations as a sum of autocorrelations of source signals shifted by the relative delays of the signals arriving at the microphone pairs, and weighted by the source power and the distances between the sources and the microphone pairs. The model is formulated as a linear inverse problem and is sparse with respect to the source power map. The source power map, which directly shows the locations of all the sound sources, can be reconstructed using ℓ_1-norm minimization algorithms. We demonstrate the effectiveness of our model in a wildlife monitoring application, where the goal is to locate multiple frogs in a dense chorus.

***Index Terms*—** cross-correlation, distributed microphone array, linear inverse problem, multi-source, source localization, sparse representation

1. INTRODUCTION

Sound source localization using a distributed microphone array is widely used in wildlife monitoring [1–3], environmental noise mapping [4, 5], surveillance, and other civilian and military applications [6]. As distributed arrays spanning a larger area with fewer microphones are desirable for cost reasons, improving accuracy and resolution of the localization task still remains a challenge for researchers worldwide.

When there is a single source, time-difference-of-arrival (TDOA) is an effective localization algorithm because of its accuracy and robustness [7]. Jones and Ratnam [1] exploited temporal sparseness of frog calls to locate active sources using TDOA. Valin *et al.* employed TDOA to detect a dominant speaker using a mobile microphone array [8]. When there are multiple sources, steered response power (SRP) is shown to be a better choice [7, 9]. In general, SRP methods beamform over a search grid by summing pairwise cross-correlations and locating peaks in the resulting SRP map [10]. Among all SRP methods, SRP-PHAT is the most popular, where the map is computed using phase-transform (PHAT) cross-correlations [7, 9]. Collier *et al.* [2] computed two-dimensional SRP maps to find locations of antbirds. Aarabi [11] used a spatial likelihood function which is a variant of SRP for source localization. One advantage of SRP methods is that they take into account the entire pairwise cross-correlation functions to make decisions, not only the peak values as in TDOA techniques. However, because SRP computes the output power of a delay-and-sum beamformer [9], which has some well-known limitations such as broad sidelobes [12], the produced SRP map is smeared and has low resolution. As a consequence, weaker sources are often masked by stronger sources and it is difficult to identify all the sources on the SRP map.

Here, we present a new model that explicitly expresses pairwise cross-correlations as a sum of autocorrelations of source signals, which are shifted by the relative delays of the signals arriving at the microphone pairs, weighted by the source power and the distances between the sources and the microphone pairs. The sound field to be recovered by our model represents the power of the present sound sources, which is spatially sparse. Using a linear inverse problem formulation, a solution for the sound field can be obtained using sparse reconstruction techniques such as FOCUSS [13], Basis Pursuit Denoising [14], LASSO [15], etc. After that, the source locations can be inferred directly from the reconstructed sound field. For the traditional signal model in the frequency domain, there also exist sparse signal reconstruction approaches to estimate direction-of-arrival of sound sources [16]. Here, we use LASSO to solve the inverse problem and compare the results with SRP-PHAT's results. Our simulation and experimental results show that our model outperforms SRP-PHAT.

*This study is supported by the research grant for the Human Centered Cyber-physical Systems Programme at the Advanced Digital Sciences Center from Singapore's Agency for Science, Technology and Research (A*STAR).

2. SIGNAL MODEL

We divide the region of interest into a grid of N identical points. Each point has a physical dimension of dx x dy x dz m^3. Let $\xi_\mathbf{n}$ be the 3D coordinates of a grid point n with reference to an arbitrary origin. We assume that every grid point can have at most one sound source. Let $s_n(t)$ be a wideband signal generated by a source at point n. If there is no source at n, $s_n(t) = 0\ \forall t$. Our objective is to identify all the grid points that contain a sound source.

Consider a microphone array comprised of M microphones arranged in an unrestricted fashion. Let $\xi_\mathbf{i}$ be the 3D coordinates of microphone i. The signal received at microphone i in the time domain is

$$y_i(t) = \sum_{n=1}^{N} \frac{1}{d_{i,n}} s_n(t - \frac{d_{i,n}}{c}) + v_i(t), \tag{1}$$

where $d_{i,n} = \|\xi_\mathbf{n} - \xi_\mathbf{i}\|_\mathbf{2}$, c is the speed of sound and $v_i(t)$ is the noise received at microphone i.

The cross-correlation between microphone i and j is $R_{ij}(\tau) = \mathbb{E}\{y_i^*(t)y_j(t+\tau)\}$, where $\mathbb{E}\{\cdot\}$ denotes mathematical expectation, $(\cdot)^*$ denotes complex conjugate, and τ is a time lag. We assume that source signals and the microphone noises are mutually uncorrelated. Expanding the right-hand side of the cross-correlation equation and eliminating the cross terms, we obtain

$$\begin{aligned} R_{ij}(\tau) &= \mathbb{E}\left\{\sum_{n=1}^{N} \frac{1}{d_{i,n}d_{j,n}} s_n^*(t - \frac{d_{i,n}}{c}) s_n(t - \frac{d_{j,n}}{c} + \tau)\right\} \\ &= \sum_{n=1}^{N} \frac{1}{d_{i,n}d_{j,n}} A_n(\tau - \tau_{i,j}^n), \end{aligned} \tag{2}$$

where $A_n(\tau) = \mathbb{E}\{s_n^*(t)s_n(t+\tau)\}$ is the autocorrelation of signal at grid point n, $\tau_{i,j}^n = (d_{j,n} - d_{i,n})/c$ is the TDOA of the signal from grid point n to microphone pair (i,j). Eq. (2) shows that the cross-correlation value of microphone pair (i,j) at lag τ is the sum of all signal autocorrelation functions at lag τ, shifted by TDOA $\tau_{i,j}^n$ and weighted by the distances from the sources to the pair of microphones. We denote the signal power at grid point n as x_n. From the definition of autocorrelation, $x_n = A_n(0)$. If there is no sound source at grid point n, $x_n = 0$. We denote $\bar{A}_n(\tau)$ as a normalized autocorrelation such that $\bar{A}_n(0) = 1$, and the cross-correlation becomes

$$R_{ij}(\tau) = \sum_{n=1}^{N} \frac{1}{d_{i,n}d_{j,n}} \bar{A}_n(\tau - \tau_{i,j}^n) x_n. \tag{3}$$

Since $R_{i,j}(\tau) = R_{j,i}(-\tau)$, we have $P = \binom{M}{2}$ unique pairs of microphones. For each pair of microphones, the maximum reliable relative delay [9] is

$$\tau_{i,j}^{max} = ceil\left(\frac{f_s}{c}\|\xi_\mathbf{j} - \xi_\mathbf{i}\|_\mathbf{2}\right), \tag{4}$$

where f_s is the sampling frequency. As a result, $\tau_{i,j}$ is limited to this set $D_{i,j} = \{-\tau_{i,j}^{max}, \cdots, -1, 0, 1, \cdots, \tau_{i,j}^{max}\}$.

We rewrite Eq. (3) in matrix form for all possible lags for $R_{i,j}(\tau)$ in Eq. (5). Let $l_{i,j} = 2\tau_{i,j}^{max} + 1$, then vector $\mathbf{r_{i,j}} \in R^{l_{i,j}}$, matrix $\mathbf{A_{i,j}} \in R^{l_{i,j} \times N}$, and vector $\mathbf{x} \in R^N$. The n^{th} column of matrix $\mathbf{A_{i,j}}$, denoted as $\mathbf{a_{i,j}^n}$, is the normalized autocorrelation function of a sound source at grid point n, shifted by the relative delay at microphone pair (i,j), and weighted by the distance from point n to microphone pair (i,j). The pairwise cross-correlation is a linear combination of columns in $\mathbf{A_{i,j}}$, scaled by signal power $\mathbf{x}$. Since $\mathbf{x}$ is the same for all pairs of microphones, we are able to stack all the pairwise cross-correlations into a single vector as follows:

$$\underbrace{\begin{bmatrix} \mathbf{r_{1,2}} \\ \vdots \\ \mathbf{r_{i,j}} \\ \vdots \\ \mathbf{r_{M-1,M}} \end{bmatrix}}_{\mathbf{r}} = \underbrace{\begin{bmatrix} \mathbf{A_{1,2}} \\ \vdots \\ \mathbf{A_{i,j}} \\ \vdots \\ \mathbf{A_{M-1,M}} \end{bmatrix}}_{\mathbf{W}} \mathbf{x}. \tag{6}$$

Let $L = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} l_{i,j}$, vector $\mathbf{r} \in R^L$, matrix $\mathbf{W} \in R^{L \times N}$. Each column of $\mathbf{W}$, $\mathbf{w^n} = \left[\mathbf{a_{1,2}^n} \cdots \mathbf{a_{M-1,M}^n}\right]^\mathbf{T}$, is a shifted and weighted normalized autocorrelation function of the source signal at grid point n. Eq. (6) is a novel interpretation of the pairwise cross-correlations in terms of the autocorrelations of source signals. From the viewpoint of deconvolution, column $\mathbf{w^n}$ can be thought of as a point spread function (PSF). From the signal representation perspective, $\mathbf{w^n}$ is a basis vector and $\mathbf{W}$ is a dictionary. If we back-project each component $\mathbf{a_{i,j}^n}$ of column $\mathbf{w^n}$ onto the grid and sum all of the projected maps, we obtain the beam pattern generated by a source at grid point n, which is similar to SRP map. The columns in $\mathbf{W}$ are not orthogonal. Each column $\mathbf{w^n}$ embeds information on both the location and the characteristic of the source signal. Back to Eq. (5), the maximum value of $\mathbf{a_{i,j}^n}$, which is $\frac{1}{d_{i,n}d_{j,n}}\bar{A}_n(0)$, is shifted from the center by $\tau_n^{i,j}$, which is the TDOA of a source at grid point n to microphone pair (i,j). By locating these maximum values in P segments $\mathbf{a_{i,j}^n}$ of $\mathbf{w^n}$, we have P values of TDOA from the source to all microphone pairs. This set of P TDOA values uniquely identify the location of the source provided if we have enough microphones. Hahn [17] shows that it requires a minimum of 3 to 4 sensors to locate a source in two dimensions, and 4 or 5 sensors to locate a source in three dimensions. The spectral characteristic of the signal is encoded in the normalized autocorrelation $\bar{A}_n$.

One question that arises is how to construct $\mathbf{W}$ if the autocorrelation functions of all the signals present in the sound field and their locations are unknown. In certain scenarios where all the sources produce similar sounds (for example same-species frog chorus [1] or bird vocalization [2,3]), it is acceptable to approximate $\bar{A}_n$ as the same for all grid points

$$\underbrace{\begin{bmatrix} R_{i,j}(-\tau_{i,j}^{max}) \\ \vdots \\ R_{i,j}(0) \\ \vdots \\ R_{i,j}(\tau_{i,j}^{max}) \end{bmatrix}}_{\mathbf{r_{i,j}}} = \underbrace{\begin{bmatrix} \frac{1}{d_{i,1}d_{j,1}}\bar{A}_1(-\tau_{i,j}^{max}-\tau_{i,j}^1) & \cdots & \frac{1}{d_{i,n}d_{j,n}}\bar{A}_n(-\tau_{i,j}^{max}-\tau_{i,j}^n) & \cdots & \frac{1}{d_{i,N}d_{j,N}}\bar{A}_N(-\tau_{i,j}^{max}-\tau_{i,j}^N) \\ \vdots & & \vdots & & \vdots \\ \frac{1}{d_{i,1}d_{j,1}}\bar{A}_1(0-\tau_{i,j}^1) & \cdots & \frac{1}{d_{i,n}d_{j,n}}\bar{A}_n(0-\tau_{i,j}^n) & \cdots & \frac{1}{d_{i,N}d_{j,N}}\bar{A}_N(0-\tau_{i,j}^N) \\ \vdots & & \vdots & & \vdots \\ \frac{1}{d_{i,1}d_{j,1}}\bar{A}_1(\tau_{i,j}^{max}-\tau_{i,j}^1) & \cdots & \frac{1}{d_{i,n}d_{j,n}}\bar{A}_n(\tau_{i,j}^{max}-\tau_{i,j}^n) & \cdots & \frac{1}{d_{i,N}d_{j,N}}\bar{A}_N(\tau_{i,j}^{max}-\tau_{i,j}^N) \end{bmatrix}}_{\mathbf{A_{i,j}}} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \\ \vdots \\ x_N \end{bmatrix}}_{\mathbf{x}}. \quad (5)$$

n. We can use some known example of the sound source of interest or use the current segment of the recorded signal to estimate $\bar{A}_n$. This is a trade-off between accuracy of our estimation and prior knowledge we have about the signals encoded in $\mathbf{W}$. From our simulation and experimental results, this approximation produces a reasonable estimation for $\mathbf{x}$.

In reality, Eq. (6) is not exact because of modelling errors such as the sound sources are not monopole, sound propagates differently across the sound field, and there are reflections from ground or other objects. We combine all of these errors into an additive Gaussian noise terms $\mathbf{e}$, and Eq. (6) becomes

$$\mathbf{r} = \mathbf{W}\mathbf{x} + \mathbf{e}. \quad (7)$$

Because the number of sources is much smaller than the number of grid points N, $\mathbf{x}$ is sparse. In addition, x is non-negative because it represents signal power. As a result, we can solve the linear inverse problem in Eq. (7) using sparse methods such as FOCUSS [13], Basic Pursuit Denoising [14], LASSO [15], etc. with non-negative constraint. We name our model as the auto-cross correlation (ACC) model. In this paper, we use LASSO to reconstruct the sound field $\mathbf{x}$ and compare the results of our ACC model with the map obtained using SRP-PHAT [9].

3. SOLVING A LINEAR INVERSE PROBLEM WITH SPARSITY CONSTRAINT

The sound field $\mathbf{x}$ can be obtained by solving the following optimization

$$\min_{\mathbf{x}} \|\mathbf{r} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \quad \mathbf{s.t.} \quad \mathbf{x} \geq \mathbf{0}, \quad (8)$$

where parameter λ controls the trade-off between the residual norm and the sparsity of $\mathbf{x}$. Eq. (8) is a convex optimization problem and can be solved using off-the-shelf packages. In this paper, we use *l1_ls* package [18] implemented in Matlab to solve Eq. (8). We manually choose a value for λ to produce a reasonably sparse estimation for $\mathbf{x}$.

4. PRACTICAL IMPLEMENTATION

Computation of cross-correlation: We use phase transform generalized cross-correlation (GCC-PHAT) [9] to estimate microphone pairwise cross-correlations.

Max filter to compensate measurement errors: One of the challenges for source localization using distributed arrays is to measure microphone locations accurately. As distributed arrays span a large area, error in array location measurement is inevitable. This error causes the estimated pairwise cross-correlations to be shifted by a few lag points. In addition, inaccurate measurement of the speed of sound also contributes to the time lag shift. To compensate for these errors, we introduce a max filter for the cross-correlation as follows $R_{i,j}(\tau) = max(R_{i,j}(\tau - \tau_{comp}), \ldots, R_{i,j}(\tau + \tau_{comp}))$, where τ_{comp} is a small lag value that is proportional to the measurement errors. The max filter introduces a standard deviation of $\pm\tau_{comp}c/f_s$ (meter) in the estimated source locations. In our experiment, we use $\tau_{comp} = 5$, thus there is an additional error of ± 8.7 cm in our location estimations.

Decimation of $R_{i,j}(\tau)$: The max filter results in redundant values of $R_{i,j}(\tau)$. Therefore we decimate $R_{i,j}(\tau)$ by a factor of $2\tau_{comp}+1$ and remove the corresponding rows in $\mathbf{W}$. This reduces the size of $\mathbf{r}$ and $\mathbf{W}$, and speed up the computation. The max filter and the decimation steps are not necessary if the measurement error is insignificant.

Using a subset of microphone pairs: When distances between microphones are large, signal coherence between two microphones reduces, thus their pairwise cross-correlation contains more noise. We select pairs of microphones with $\tau_{i,j}^{max} < \tau^{max}$ to reduce errors in our estimation. In our experiments, we choose $\tau^{max} = 400$.

5. RESULTS AND DISCUSSION

We experimented with our model on green tree frog data collected in a breeding pond (Creekfield Lake) at Brazos Bend State Park, TX (USA). Readers can refer to [19] for more information on the experimental set up. Our array consisted of 15 omnidirectional microphones, divided into 3 identical modules as shown in Fig.1. In each module, 4 microphones were mounted on the ends of a 1.4 m cross-arm positioned $2.65 - 2.9$ m above the ground, and a fifth microphone was mounted 1 m below the cross at the end of a single cross-arm of 0.7 m. Microphone signals were sampled simultaneously at 20 ksps, 24 bit. The estimated speed of sound was 346.8 m/s. We bandpass-filtered the signal in the frog frequency range ($700 - 5500$ Hz) before computing cross-correlations. The length of one data segment was 0.15 seconds. Our search

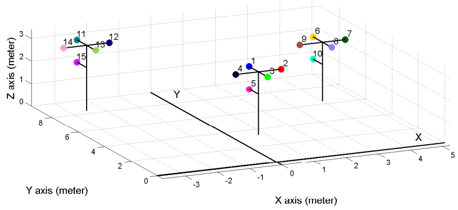
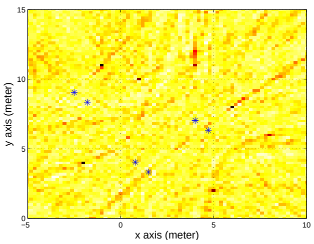
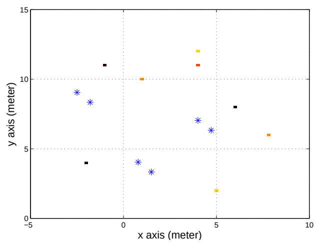


Fig. 1: Microphone array location.



(a) SRP map at z = 0 m



(b) ACC map at z = 0 m

Fig. 2: Simulation results for 8 sound sources.

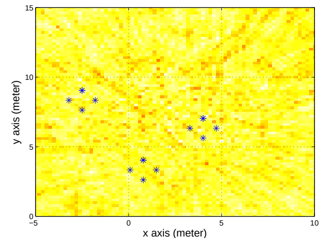
grid was between $[-5, 10)$ m in x axis, $[0, 15)$ m in y axis, and $[0, 0.5]$ m in z axis. We used $dx = 0.2$ m, $dy = 0.2$ m, $dz = 0.5$ m. The grid had 11552 points and covered 225 m^2.

5.1. Simulation results

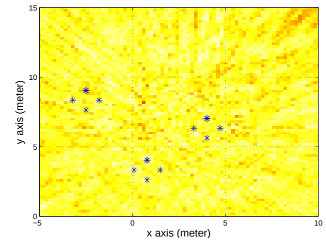
For the simulation, we used only 6 microphones (1, 2, 6, 7, 11, 12) to simulate a very sparse array. We randomly generated 8 white noise sources on the plane $z = 0$ m. We also added white noise of 5dB SNR into the signal. Since we knew the array location exactly, we did not use the max filter and decimation of $R_{i,j}(\tau)$. We used the normalized autocorrelation of white noise to compute $\mathbf{W}$ using Eq. (5). Fig. 2 displays results obtained by SRP-PHAT and our auto-cross correlation (ACC) model. The blue stars indicate the microphone locations. The sound field obtained by our ACC model is clean and sparse. The ACC model recovered the exact locations of all the 8 sources. In contrast, the SRP map produces multiple false peaks. In addition, for two nearby sources in the top right corner, the peak corresponding to the second source is not welldefined. From the simulation results, our model has better resolution and accuracy than the SRP-PHAT method.

5.2. Experimental results

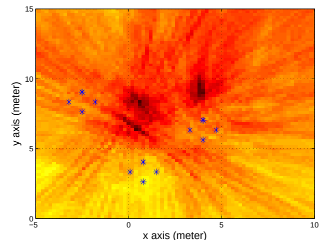
We extracted one frog call from our measurement data to compute the normalized autocorrelation. Then, we estimated $\mathbf{W}$ as in Eq. (5) assuming the same autocorrelation for all n. We selected 76 out of 105 pairs of microphones. After the max filter and decimation, the size of $\mathbf{W}$ was 2993 x 11552. We set $\lambda = 0.01$. Fig. 3 (a-d) shows SRP maps without and with max filter. Without the max filter, the pairwise cross-correlations when back-projected on the grid search are misaligned, and produce noisy maps. With the max filter, the peaks of the cross-correlations are aligned, and we could easily determine the source locations. Besides the ACC model, the max filter is also a novel contribution of this paper. Fig. 3 (e-f) shows our model estimation results. For experimental data, we observe that the hills on the SRP map are wide. Thus nearby sources can be potentially merged into one. In contrast, our model returns a clean power map. It is much easier to determine the source locations from the ACC map than SRP map. In [19], Jones *et al.* located 6 frogs A, B, C, D, E and F. Fig. 3 (e-f) shows locations of frogs A, B and D. We processed the data to determine the mean locations of all the 6 frogs. Using [19] as a ground truth, our root mean square error is 0.54 m. We conclude here that our new model is promising for source localization using distributed arrays.



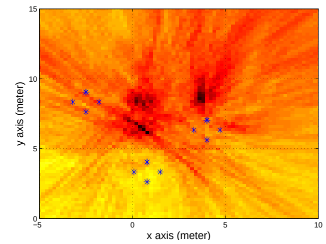
(a) SRP map without error compensation at z = 0 m



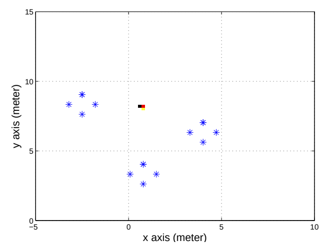
(b) SRP map without error compensation at z = 0.5 m



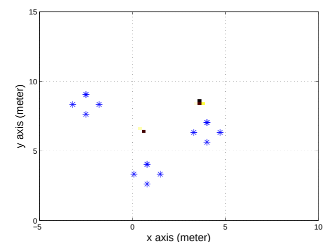
(c) SRP map at z = 0 m



(d) SRP map at z = 0.5 m



(e) ACC map at z = 0 m



(f) ACC map at z = 0.5 m

Fig. 3: Experimental results for 3 sound sources.

6. ACKNOWLEDGEMENT

We would like to thank Russell Jones and Rama Ratnam from the University of Illinois at Urbana-Champaign for providing us the green tree frog data that were used in our study.

7. REFERENCES

[1] Douglas L Jones and Rama Ratnam, "Blind location and separation of callers in a natural chorus using a microphone array.," *The Journal of the Acoustical Society of America*, vol. 126, no. 2, pp. 895–910, 2009.

[2] Travis C Collier, Alexander NG Kirschel, and Charles E Taylor, "Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network.," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 182–189, 2010.

[3] Hanbiao Wang, Chiao-En Chen, Andreas Ali, Shadnaz Asgari, Ralph E Hudson, Kung Yao, Deborah Estrin, and Charles Taylor, "Acoustic sensor networks for woodpecker localizations," *Proceedings of SPIE*, vol. 5910, 2005.

[4] Shengkui Zhao, Cagdas Tuna, Thi Ngoc Tho Nguyen, and Douglas L Jones, "Large region acoustic source mapping: A generalized sparse constrained deconvolution approach," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3186–3190.

[5] Cagdas Tuna, Shengkui Zhao, Thi Ngoc Tho Nguyen, and Douglas L Jones, "Drive-by large-region acoustic noise-source mapping via sparse beamforming tomography," *The Journal of the Acoustical Society of America*, vol. 140, no. 4, pp. 2530–2541, 2016.

[6] Feng Zhao, Jie Liu, Juan Liu, Leonidas Guibas, and James Reich, "Collaborative signal and information processing: an information-directed approach," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1199–1209, 2003.

[7] Joseph H DiBiase, Harvey F Silverman, and Michael S Brandstein, "Robust localization in reverberant rooms," *Microphone Arrays*, pp. 157–180, 2001.

[8] Valin J-M, François Michaud, Jean Rouat, and Dominic Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*. IEEE, 2003, vol. 2, pp. 1228–1233.

[9] Jacek P Dmochowski, Jacob Benesty, and Sofiène Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2510–2526, 2007.

[10] Joseph H DiBiase, *A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Array*, Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2000.

[11] Parham Aarabi, "The fusion of distributed microphone arrays for sound localization," *Eurasip Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.

[12] Petre Stoica and Randolph L Moses, *Spectral analysis of signals*, vol. 452, Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[13] Irina F Gorodnitsky and Bhaskar D Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, 1997.

[14] Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[15] Robert Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[16] Dmitry Malioutov, Müjdat Çetin, and Alan S Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[17] William R Hahn and Steven A Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.

[18] Kwangmoo Koh, Seung-Jean Kim, and Stephen Boyd, "Simple matlab solver for l1-regularized least squares problems," `https://stanford.edu/~boyd/l1_ls/`, 2008, [Online; accessed 05-July-2016].

[19] Douglas L Jones, Russell L Jones, and Rama Ratnam, "Calling dynamics and call synchronization in a local group of unison bout callers," *Journal of Comparative Physiology A*, vol. 200, no. 1, pp. 93–107, 2014.