Aalborg Universitet

Localizing Near and Far Field Acoustic Sources with Distributed Microhone Arrays

Hansen, Martin Weiss; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

LOCALIZING NEAR AND FAR FIELD ACOUSTIC SOURCES WITH DISTRIBUTED MICROHONE ARRAYS

Martin Weiss Hansen, Jesper Rindom Jensen and Mads Græsbøll Christensen

Audio Analysis Lab, AD:MT, Aalborg University, Denmark
{mwh, jrj, mgc}@create.aau.dk

ABSTRACT

In this paper, we consider near-field acoustic source localization using distributed microphone arrays. Range differences (RDs) are estimated using a recently proposed joint direction of arrival (DOA) and range estimator. The RDs are used to estimate the location of an acoustic source using a recently proposed cone-based localization method. The performance of the proposed localization method is compared to generalized cross-correlation with phase transform, and a pitch-based time difference of arrival (TDOA) estimator, using synthetic and real signals. Results show a decrease in the error of the estimated position when joint DOA and range estimation is used for RD estimation.

Index Terms— Acoustic source localization, multi-channel processing, noise reduction, speech enhancement.

1. INTRODUCTION

Within the field of audio processing, acoustic source localization with microphone arrays is an important aspect in noise reduction and source separation [1]. Microphone arrays are more commonly found in many modern-day devices, e.g., hearing aids, smartphones, smart TVs and laptops, making distributed processing feasible. These microphone arrays facilitate beamforming [2] and direction of arrival (DOA) estimation that can be used for, e.g., video teleconferencing [3].

Several classes of methods for source localization exist, e.g., multilateration [4], TDOA-based source localization [5], maximum likelihood source localization [6] and energy-based source localization [7]. The most popular class of methods is probably the TDOA-based localization methods [5, 8, 9, 10, 11]. Recently, methods for acoustic source localization with distributed microphone arrays, based on geometric cone constraints, have been proposed in [12], which are related to the spherical least squares (LS) errors [9]. These geometric methods are based on time difference of arrival (TDOA) measurements, estimated using GCC-PHAT [13], which are converted

to range differences (RDs).

In many TDOA-based acoustic localization methods, TDOAs are estimated using generalized cross-correlation with phase transform (GCC-PHAT), where PHAT is a heuristic weighting used to whiten the cross-spectrum [13, 14, 10]. Examples of its use can be found in [15, 16]. GCC-PHAT is a broadband heuristic method, while the methods in [17, 18] model the desired signal using a harmonic signal model, such that only the frequency components related to the signal are used in the localization. This is beneficial for statistical reasons and in multi source scenarios. Furthermore, in [18] the signal is modeled using both TDOAs and gain ratios of arrival (GROAs), which means that the acoustic source is not assumed to be placed in the far-field of the microphone arrays. This should increase performance, compared to GCC-PHAT, when the signal of interest is in the near field of one or more microphone arrays.

In this paper, we propose a method for acoustic source localization for distributed microphone arrays, based on the cone-based method of [12], which is based on geometric cone constraints. The localization method is based on RDs estimated individually for each array using the near-field localization method of [18], which is statistically efficient, and applicable in both near- and far-field scenarios, potentially with multiple sources present. When distributed microphone arrays are used for localization, the probability is high that the source is close to one or more of the arrays. Because of this, near field localization is an important aspect to consider.

The paper is organized as follows. In Section 2, we present the signal model and the near-field range estimation method. In Section 3, the proposed localization scheme is presented. Section 4 presents some experimental results, while a conclusion is presented in Section 5.

2. SIGNAL MODEL

We now introduce the signal model for each microphone array [18]. Consider a scenario where an array with K microphones captures snapshots of length N, i.e., the data in channel k at time instant n is given by

$$\mathbf{x}_k(n) = [x_k(n) \quad x_k(n+1) \quad \cdots \quad x_k(n+N-1)]^T. \quad (1)$$

We assume that the snapshots in (1) contain a quasi-periodic source $s(n)$ corrupted by noise $v_k(n)$, and that the environment is anechoic. Because of different distances from each microphone to the source, and due to the inverse square law for sound propogation, the signal at microphone k can be modeled as

$$x_k(n) = \beta_k s(n - f_s \tau_k) + v_k(n), \qquad (2)$$

where β_k is an attenuation factor reciprocal to the distance from the source to microphone k, f_s is the sampling frequency, τ_k is the time it takes the sound waves to travel from the source to the kth microphone. If the first microphone is chosen as reference, we have $s_1 = \beta_1 s(n - f_s \tau_1)$, and (2) can be re-written as

$$x_k(n) = \frac{r_1}{r_k} s_1(n - f_s \tau_{1k}) = \frac{r_1}{r_k} s_1(n - f_s \frac{r_k - r_1}{c}), \quad (3)$$

where r_k is the distance from the source to the kth microphone, τ_{1k} is the time it takes for the sound waves to travel from the reference microphone to the kth microphone, and c is the speed of sound. If we assume that the microphones form a uniform linear array (ULA), r_k is given by

$$r_k = \sqrt{g_k^2 d^2 + r_c^2 - 2g_k d r_c \sin\theta}, \qquad (4)$$

where $g_k = \frac{K-1}{2} - k + 1$, d is the microphone spacing, r_c is the distance from the source to the center of the array and θ is the DOA of the source. The ULA structure is used for illustration purposes, but by finding other models for r_k, other array structures can be taken into account. Since $s_1(n)$ is a quasi-periodic signal, it can be modeled as

$$s_1\left(n - f_s \frac{r_k - r_1}{c}\right) = \sum_{l=1}^{L} \gamma_l e^{jl\omega_0 n} e^{-j2\pi l f_0 \frac{r_k - r_1}{c}}, \quad (5)$$

where L is the model order, $\gamma_l = \beta_1 \alpha_l$, $\alpha_l = A_l e^{\phi l}$ is the complex amplitude of harmonic l, A_l is its real amplitude, ϕ_l is its phase, ω_0 is the fundamental frequency, and $f_0 = f_s \frac{\omega_0}{2\pi}$. Using (5), the signal model in (3) can be rewritten as

$$x_k(n) = \frac{r_1}{r_k} s_1\left(n - f_s \frac{r_k - r_1}{c}\right) + v_k(n), \qquad (6)$$

which can be used to form a vector model as

$$\mathbf{x}_k = \mathbf{Z}(\omega_0)\mathbf{D}_k(r_c, \theta)\boldsymbol{\gamma} + \mathbf{v}_k, \qquad (7)$$

where $\mathbf{x}_k = [x_k(0) \quad x_k(1) \quad \cdots \quad x_k(N-1)]^T$, with a similar definition for $\mathbf{v}_k$. Also,

$$\mathbf{Z}(\omega_0) = [\mathbf{z}(\omega_0) \quad \cdots \quad \mathbf{z}(L\omega_0)], \qquad (8)$$

$$\mathbf{z}(\omega) = [1 \quad e^{j\omega} \quad \cdots \quad e^{j(N-1)\omega}]^T, \qquad (9)$$

$$[\mathbf{D}_k(r_c, \theta)]_{ll} = \sqrt{\frac{g_1^2 d^2 + r_c^2 - 2g_1 d r_c \sin\theta}{g_k^2 d^2 + r_c^2 - 2g_k d r_c \sin\theta}} \times \qquad (10)$$

$$e^{-j2\pi l f_0 \frac{\sqrt{g_k^2 d^2 + r_c^2 - 2g_k d r_c \sin\theta} - \sqrt{g_1^2 d^2 + r_c^2 - 2g_1 d r_c \sin\theta}}{c}},$$

$$\boldsymbol{\gamma} = [\gamma_1 \quad \cdots \quad \gamma_L]^T, \qquad (11)$$

and $[\mathbf{D}_k(r_c, \theta)]_{pq} = 0$ for $p \neq q$. In the remainder of the paper, we omit writing the dependency on ω_0 to simplify the notation. If the noise in each channel is assumed to be white Gaussian and uncorrelated across channels, the log-likelihood function is given by [19, 20]

$$\ln p(\{\mathbf{x}_k(n)\}; \boldsymbol{\psi}) = -NK\ln\pi - N\sum_{k=1}^{K}\ln\sigma_k^2 - \qquad (12)$$

$$\sum_{k=1}^{K}\frac{\|\mathbf{v}_k\|^2}{\sigma_k^2},$$

where $\boldsymbol{\psi}$ is the unkown parameter vector, σ_k^2 is the variance of the noise at microphone k and $\|\cdot\|$ denotes the l_2-norm.

3. SOURCE LOCALIZATION

The proposed method for acoustic source localization is based on the cone-based method in [12], with RDs estimated using the joint DOA and range method presented in [18]. In order to find the DOA and range, the log-likelihood function (12) is maximized by differentiating with respect to $\boldsymbol{\gamma}$, and setting equal to zero, i.e.,

$$\widehat{\boldsymbol{\gamma}} = \left(\sum_{k=1}^{K}\frac{\mathbf{D}_k^H \mathbf{Z}^H \mathbf{Z} \mathbf{D}_k}{\sigma_k^2}\right)^{-1}\sum_{k=1}^{K}\frac{\mathbf{D}_k^H \mathbf{Z}^H \mathbf{x}_k}{\sigma_k^2}. \qquad (13)$$

We also solve for the unknown noise variance, which yields

$$\widehat{\sigma}_k^2 = \frac{\|\mathbf{x}_k - \mathbf{Z}\mathbf{D}_k(r_c, \theta)\boldsymbol{\gamma}\|^2}{N}. \qquad (14)$$

The amplitude estimates (13) and the noise variance (14) are interdependent. A way to deal with this is to estimate the parameters iteratively and by initializing, e.g., $\sigma_k^2 = 1$, for all k. After convergence, the noise variance estimate (14) can be inserted into (12), after which the DOA and range can be estimated by minimizing the expression

$$\{\widehat{\theta}, \widehat{r_c}\} = \operatorname*{arg\,min}_{\{\theta, r_c\} \in \Theta \times \mathcal{R}_c} \sum_{k=1}^{K}\ln\|\mathbf{x}_k - \mathbf{Z}\mathbf{D}_k(r_c, \theta)\boldsymbol{\gamma}\|^2, \quad (15)$$

where Θ and $\mathcal{R}_c$ are sets of candidate DOAs and ranges, respectively. Note that computationally simpler algorithms are presented in [18]. The estimated distance r_c from the source to the center of the array can be used together with (4), to find the distance from the source to each of the sensors in the microphone array, assuming a ULA structure.

Equipped with these ranges, we now proceed by using the source localization method [12]. In this localization method, an extended coordinate system is formed by adding a RD coordinate to the coordinates of the source. Each point $\mathbf{p} = [x, y, s]^T$ is mapped onto the 4D space-range $[\mathbf{p}^T, w]^T$, where

w is the range coordinate. Consider an array consisting of K mics. The RD of mic. k is

$$w_k = \|\mathbf{p}_s - \mathbf{m}_k\| - \|\mathbf{p}_s - \mathbf{m}_1\| = r_k - r_1, \qquad (16)$$

where $\mathbf{p}_s = [x_s, y_s, z_s]^T$ is the source position, $\mathbf{m}_k = [x_k, y_k, z_k]^T$ is the position of microphone k and $\mathbf{m}_1 = [0, 0, 0]^T$ is the position of the reference sensor. In practice, the range of microphone k in each array, is given by the estimate of r_c in (15) and (4). The microphones $\mathbf{m}_k$ can be represented by cones with apex $[\mathbf{m}_k, w_k]^T$. With noiseless measurements, $\mathbf{p}_s$ should fall in the intersection of all such cones (see [12] for further details). Because of noisy measurements, the source location is found as the point $\widehat{\mathbf{p}}_s$ with minimum distance to the surface of the cones, i.e.,

$$\widehat{\mathbf{p}}_s = \arg\min_{\mathbf{p}_s \in \mathbf{P}_s} \|\mathbf{e}(\mathbf{p}_s)\|^2, \qquad (17)$$

where $\mathbf{P}_s$ is a set of candidate source positions, $\mathbf{e}(\mathbf{p}_s) = [e_1(\mathbf{p}_s) \cdots e_K(\mathbf{p}_s)]^T$, and $e_k(\mathbf{p}_s) = (x_s - x_k)^2 + (y_s - y_k)^2 + (z_s - z_k)^2 - (w_s - w_k)^2$. To address synchronization in the case of U distributed arrays, a difference between the distance from the source to the reference and to the local reference of the uth array is defined as [21]

$$\Delta z^{(u)} = \sqrt{\Delta_{x_u}^2 + \Delta_{y_u}^2 + \Delta_{z_u}^2} - \sqrt{\Delta_{x_1}^2 + \Delta_{y_1}^2 + \Delta_{z_1}^2}, \quad (18)$$

where $\Delta_{x_u} = x_{1,u} - x_s$, $x_{1,u}$ is the position of the reference sensor in the uth array, and $u = 1, \ldots, U$. By adding the displacements (18) to the cone errors in (17), the RD estimates refer to a single global reference microphone.

4. EXPERIMENTS

We now present some experiments demonstrating the performance of the proposed method for acoustic source localization, using joint DOA and range estimation [18] for RD estimation, compared to GCC-PHAT [13, 14] and a non-linear least squares (NLS) method for joint DOA and pitch estimation [17], which is modified to estimate TDOAs instead of DOAs. In the experiments, the pitch is assumed known. In practice it can be done using one of the methods in [22].

Three experiments have been conducted, two using a synthetic harmonic signal, and one using a real speech signal. In all three experiments, the acoustic environment is set up using the RIR Generator for MATLAB [23], which is based on the image method [24]. The room dimensions are 4 by 4 by 3 m. Four microphone arrays each consisting of four microphones are used. The arrays are placed with their reference microphones at the middle of each side of the room, along the walls. The microphone spacing is $d = 5$ cm. Fig. 1 shows the placement of the microphones. The speed of sound is assumed to be $c = 343$ m/s. The RIRs are used to generate a multi-channel signal according to the total number of microphones. After generating a multichannel signal

with the above-mentioned setup, a number of channels of diffuse Gaussian white noise, corresponding to the total number of microphones, is added to the signals at the microphones, resulting in an SNR $\text{SNR}_\text{m} = 30$ dB. White Gaussian noise is added to the source signal, resulting in different SNRs SNR_s. It is worth noting that methods based on a harmonic signal model are more robust when there is no noise present in the signal, i.e, when $\text{SNR}_\text{s} \to \infty$. The noisy signals are processed for each array, since we are considering a distributed network of microphone arrays. For the joint DOA and range estimator, the DOA search grid spacing is 1 degree, and the range grid spacing is 4 cm. For the pitch-based TDOA estimator, the maximum possible TDOA corresponds to the distance between the sensors in the array. Because of this, the TDOAs are estimated using a grid ranging from -1.5 to 1.5 samples. The search grid spacing is 0.01 samples. For GCC-PHAT the same TDOA grid was used, and an FFT of length 512 samples was used, and we integrated over frequencies in the range [300, 4000] Hz [25]. The localization search grid spacing is 1 cm in all directions. The metric used to evaluate and compare the methods is the magnitude of the distance from the true source position to the estimated source position.

The synthetic signal used in the two first experiments is composed of $L = 10$ harmonic sinusoids, with a sampling rate of $f_s = 8$ kHz. The experimental data is generated by performing 200 Monte-Carlo simulations for each data point. For each simulation, the fundamental frequency is sampled from the interval $f_0 \in [300, 400]$ Hz, and the phases of the harmonics are randomized. In the first experiment, the magnitude of distance error is calculated for six different source-to-array distances, varied in the direction of the y-axis, $y_s \in [0.1, 2]$ m, while the x-values are sampled from the interval $x_s \in [1.9, 2.1]$ m, and the z-values are sampled from the interval $z_s \in [1.4, 1.6]$ m. The frame size is $N = 100$ samples. Fig. 2 shows the results. We note that the magnitude distance error decreases by a factor of two when the joint DOA and range method is used to estimate the RDs used for localization, compared to the GCC-PHAT and pitch-based TDOA methods.

In the second experiment, the error is calculated for different SNRs at the microphones, $\text{SNR}_\text{m} = \{0, 5, 10, 20, 30, 40\}$ dB. The source position coordinates are sampled from the intervals $x_s \in [1.9, 2.1]$, $y_s \in [0.0, 0.2]$ and $z_s \in [1.4, 1.6]$. Fig. 3 shows the results. Again, the results show a decrease in the error when the joint DOA and range method is used. It seems like the performance does not increase for SNR_m, which it should in theory. This is most likely because the resolution of the RD and localization grids, and because of the noise which is added to the source, which means that the signal model does not fit the signal perfectly.

The third experiment is conducted using a speech signal ("Why were you away a year, Roy?"). The signal was downsampled to $f_s = 8$ kHz, and the pitch and model order for each frame were estimated using the joint ANLS method in
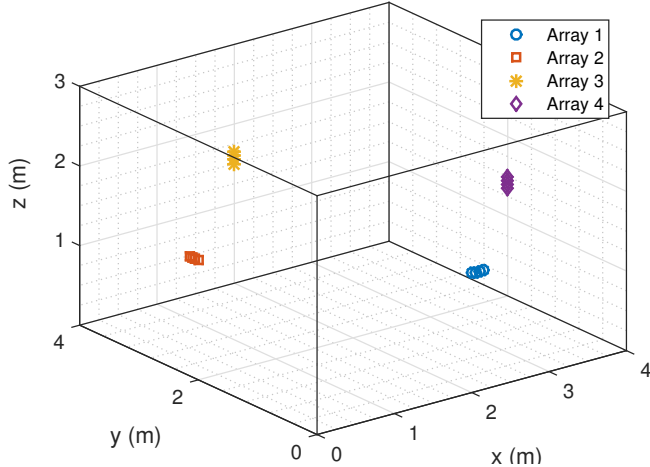
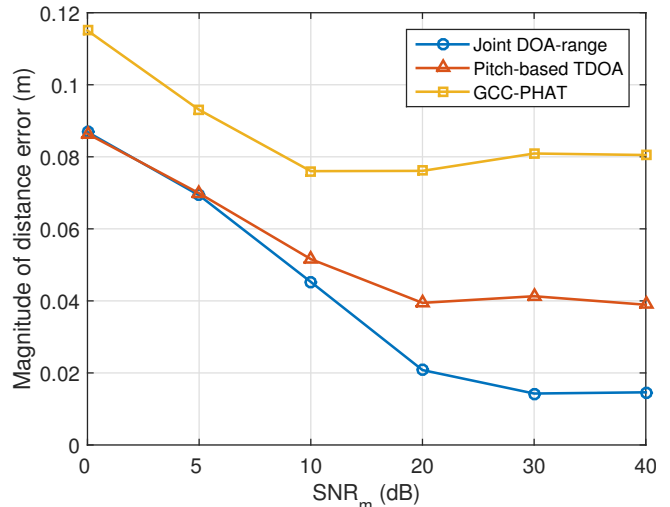**Fig. 1**. Room dimensions and array configuration.
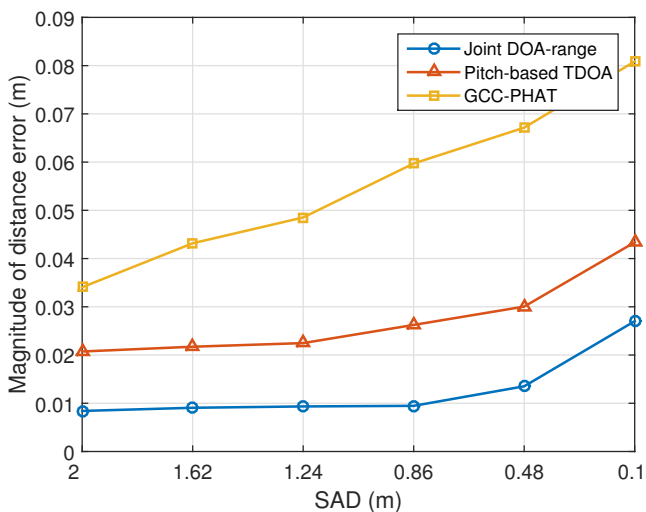
Fig. 2. Magnitude of error (m) as a function of source-to-array distance (SAD) (m).

Fig. 3. Magnitude of error (m) as a function of SNR_m (dB).

Fig. 4. Top: Spectrogram of trumpet signal. Bottom: Magnitude of error (m) as a function of SNR_m (dB).
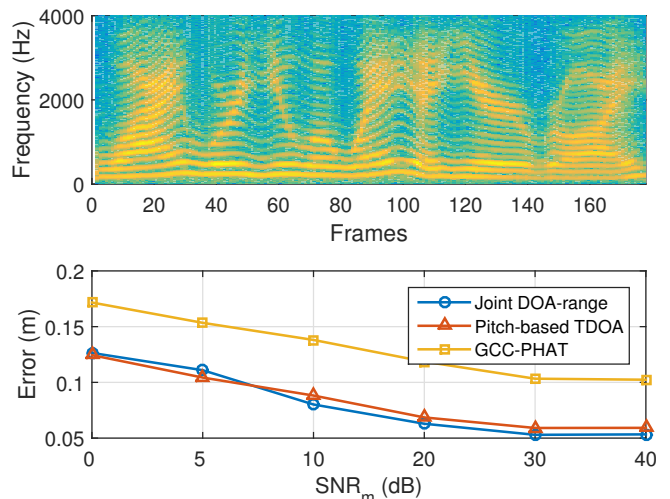
[22]. The signal is processed in blocks of length $N = 100$ samples, and the data is generated by performing 100 simulations on consecutive and non-overlapping frames. The source position coordinates are sampled from the intervals $x_s \in [1.9, 2.1]$, $y_s \in [0.0, 0.2]$ and $z_s \in [1.4, 1.6]$. Fig. 4 shows the results. The results show a decrease in the error when the joint DOA and range method is used, which is consistent with the results of the other experiments.

5. DISCUSSION

In this paper, we presented a near-field acoustic source localization methodology for scenarios where distributed microphone arrays are used. The methodology involves the joint DOA and range estimator of [18] and the cone-based localization method of [12]. Three experiments have been conducted by means of Monte-Carlo simulations on synthetic and real signals. The evaluation metric used in the experiments is the magnitude of the distance error. The results show an increase in performance when the joint DOA and range method is used for RD estimation. The increased localization performance could, for example, lead to better (distributed) beamforming performance. Future work could include the development of an algorithm where the variance of the error for each array could be used to weight the estimates of the individual arrays to account for different SNRs at the different microphones or arrays. This would most likely increase the localization performance even further. It would also be interesting to extend the signal model to take reverberation into account.

6. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing. Springer, 2001.

[2] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[3] B. Kapralos, M. R. M. Jenkin, and E. Milios, "Audio-visual localization of multiple speakers in a video tele-conferencing setting," *Int. J. of Imaging Systems and Technology*, vol. 13, no. 1, pp. 95–105, 2003.

[4] T. E. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*, Academic Press, 2009.

[5] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 12, pp. 1661–1669, 1987.

[6] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1843–1854, August 2002.

[7] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 17–29, March 2002.

[8] Y. Huang, J. Benesty, G. W.W. Elko, and R. M. Mersereati, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 943–956, 2001.

[9] Y. Huang, J. Benesty, and G. W. Elko, "Source localization," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pp. 229–253. Springer US, 2004.

[10] Y. Huang, J. Benesty, and J. Chen, "Time delay estimation and source localization," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds., pp. 1043–1063. Springer Berlin Heidelberg, 2008.

[11] M. D. Gillette and H. F. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Process. Lett.*, vol. 15, pp. 1–4, 2008.

[12] P. Bestagini, M. Compagnoni, F. Antonacci, A. Sarti, and S. Tubaro, "TDOA-based acoustic source localization in the space-range reference frame," *Multidim. Syst. Sign. Process.*, pp. 1–23, March 2013.

[13] C. H. Knapp and G. C. 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.

[14] D. Hertz and M. Azaria, "Time delay estimation between two phase shifted signals via generalized cross-correlation methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 8, no. 2, pp. 235 – 257, 1985.

[15] D. Salvati, S. Canazza, and A. Rodà, "A sound localization based interface for real-time control of audio processing," *Proc. Int. Conf. Digital Audio Effects*, pp. 177–184, 2011.

[16] P. Annibale and R. Rabenstein, "Closed-form estimation of the speed of propagating waves from time measurements," *Multidim. Syst. Sign. Process.*, pp. 1–18, 2013.

[17] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Non-linear least squares methods for joint DOA and pitch estimation," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 21, no. 5, pp. 923–933, 2013.

[18] J. R. Jensen and M. G. Christensen, "Near-field localization of audio: A maximum likelihood approach," in *Proc. European Signal Processing Conf.*, Sept 2014, pp. 895–899.

[19] M. G. Christensen, "Multi-channel maximum likelihood pitch estimation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 409–412, 2012.

[20] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Statistically efficient methods for pitch and DOA estimation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 3900–3904, 2013.

[21] M. Compagnoni et al., "Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 20, no. 7, pp. 1964–1975, 2012.

[22] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Synthesis lectures on speech and audio processing. Morgan & Claypool Publishers, 2009.

[23] E. A. P. Habets, "Signal generator for MATLAB," Tech. Rep., Technische Universiteit Eindhoven, 2011.

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

[25] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays*, pp. 157–180. Springer, 2001.