ENVIRONMENT AWARE ESTIMATION OF THE ORIENTATION OF ACOUSTIC SOURCES USING A LINE ARRAY

Piergiorgio Svaizer, Alessio Brutti, Maurizio Omologo

Fondazione Bruno Kessler-Irst
via Sommarive 18, 38123 Trento, Italy
{svaizer|brutti|omologo}@fbk.eu

ABSTRACT

This paper addresses the problem of estimating the orientation of an acoustic source with a line microphone array by exploiting the information provided by the multipath propagation in an enclosure. Although the device has a one-dimensional geometry, in presence of wavefronts reflected by surfaces within a room it is possible to obtain an accurate estimation of the emission point (even in terms of 3D coordinates) provided that the information conveyed by mutual delays between the arrivals of direct and reflected waves is exploited. This paper introduces a novel method that, properly modeling the multipath propagation, yields an accurate estimation of the source orientation. Experimental results obtained on real data acquired in a reverberant room as well as on synthetic data are presented to show the effectiveness of the proposed method and to investigate on its behavior under different environmental conditions in terms of noise and reverberation.

Index Terms— Source orientation, image method, GCC-PHAT, acoustic maps.

1. INTRODUCTION

The presence of reflecting surfaces in an enclosure has generally a negative influence on acoustic source localization based on Time Difference Of Arrival (TDOA) at multiple microphones [1]. Mirrored image sources appear and produce multiple wavefronts interfering with each other, which complicates the TDOA estimation for the direct sound [2]. If the propagation model is made explicitly aware of the multipath aspect, single reflections can be considered in order to take advantage of the additional information they convey. In the presence of a few discernible early reflections it is possible not only to improve localization performance [3, 4] but also to achieve further insight about the directional characteristics of the source.

In a previous work [5] we introduced the MultiPath Global Coherence Field (MP-GCF) method for source localization relying on exploitation of both direct and reflected wavefronts. Although in order to best exploit the benefits of a 3D sound propagation/reflection modeling, the geometry of a microphone array for source localization should span all the three dimensions of space, MP-GCF ensures accurate 3D estimations of the source position also using a line array [6], i.e. a compact 1D device with a large number of closely-spaced microphones.

In this paper we extend the MP-GCF approach and present a method to estimate the orientation of a non-omnidirectional sound source by using a simple geometric model of the acoustic propagation within an enclosure. The idea is based on the fact that given a spatial source location, the amount of energy received by the microphones through each propagation path is related to the orientation and directivity of the source. This information is partially embedded in the Generalized Cross Correlation-Phase Transform (GCC-PHAT) [7] function in the form of multiple peaks besides the one associated to the line-of-sight (direct sound). Similarly to what done in the Oriented Global Coherence Field (OGCF) approach [8], the source orientation is inferred by analyzing the patterns of GCC-PHAT functions evaluated at several distributed microphone pairs.

The use of a line array is one of the most important aspects of the proposed method. In general, the problem of estimating the orientation of an acoustic source is tackled by adopting distributed acquisition settings that surround the area of interest. For instance, even if rather different techniques are employed, in [8, 9, 10, 11, 12] microphones distributed on the walls are used and specifically in [13] a huge array consisting of 512 microphones is adopted. All the above methods rely on the direct wavefronts only, and for this reason they require a uniform coverage of the angular space. Conversely, the approach presented here exploits the information brought by multipath propagation to avoid such a constraint. A similar approach, based on eigenspace of the correlation matrix instead of GCC-PHAT, is presented in [14] where a compact circular microphone array is adopted.

This paper is organized as follows: Sections 2 and 3 recall the MP-GCF approach which exploits direct as well as reflected wavefronts. Section 4 introduces the proposed method for source orientation estimation. Section 5 reports on experimental results obtained from real data and under different simulated environmental conditions. Section 6 concludes the paper with final remarks.

2. GCC-PHAT IN MULTIPATH CONDITIONS

The GCC-PHAT is often exploited for TDOA estimation in acoustic source localization under the hypothesis that the spatial coherence of the direct wavefront predominates over the contributions of early reflections, reverberation and noise. In the case of directive sources and reflecting surfaces, low order reflections may have amplitude comparable to the direct sound, or even larger. In such conditions multiple peaks of similar amplitude are found in the GCC-PHAT. The amplitude of GCC-PHAT at a given time lag expresses the phase correlation between the two signals, which, in presence of reverberation, depends also on the local Direct-to-Reverberant Ratio, as discussed in [15]. The computation of the GCC-PHAT involves a whitening of the crosspower spectrum, which is a normalized-energy non-linear operation. As a consequence, the principle of superposition of effects does not hold. In practice, however, the information

This work was partially supported by the European Commission under the Project SCENIC, Future Emerging Technologies (FET), 7th FP, grant number 226007.

© EURASIP, 2012 - ISSN 2076-1465

concerning the mutual delays of the strongest wavefronts is partially preserved and still exploitable. This situation is exemplified in Figure 1, where the sound emitted by a source arrives at two microphones following the direct line and two paths reflected on the left and right walls. The associated Room Impulse Responses (RIR) are shown in the upper part of the figure, while their cross-correlation and GCC-PHAT are illustrated in the lower part. The peaks of the cross-correlation between the RIRs are labeled according to the contributing wavefronts. They provide information about all the mutual delays. As it can be noted, the GCC-PHAT only roughly corresponds to the amplitudes of the cross-correlation, due to its non-linearity. However, the position of the relevant peaks (full circles in the figure) is anyway preserved. In practice, from the signals acquired by means of a set of microphones it is not possible to measure directly the single RIR at each microphone, unless specific test signals are used (e.g., sweeping tones generated by means of loudspeakers). Consequently, the cross-correlation of impulse responses is also not observable, while the GCC-PHAT can be computed from the received signals to obtain an approximation of it.

3. MULTIPATH ACOUSTIC MAPS

Acoustic maps for source localization are functions expressing a source localization score at each point $\mathbf{p} = (x, y, z)$ of the space where the source can be active. A typical example is the Global Coherence Field (GCF) [16], also known as SRP-PHAT [17], which combines the GCC-PHAT functions computed at distributed microphone pairs to obtain a robust estimation of the source position also in presence of noise and reverberation.

In [6], a line array of 64 closely spaced (2 cm) microphones had been used to analyze multipath propagation inside an enclosure with high spatial resolution. The spatial evolution of impulse responses along the line array was exploited to build acoustic maps of a room from which a source emitting a chirp signal was localized, based on a match of the observed multipath arrivals with those generated by simulated image sources. This approach was then extended to the MP-GCF for the case of uncontrolled sources (e.g., producing speech signals), by analyzing the evolution of the GCC-PHAT along the array [5]. Using a fixed reference microphone m_0 at one side of the line array, the GCC-PHAT can be computed for all the microphone pairs $m_0 - m_n$ ($n = 1..N$ where $N = 63$). Let us introduce the column vector $\mathbf{C}_n = [C_n(\tau)]$ representing the GCC-PHAT values, with the time lag τ corresponding to mutual delay (integer sample resolution in the range $-\tau_{\max} \leq \tau \leq \tau_{\max}$). Let us then define a matrix $\mathbf{M} = [\mathbf{C}_1 \cdots \mathbf{C}_N]^T$. The elements $M(n, \tau)$ of this matrix can be effectively visualized in a diagram as a gray-level image, where darker pixels mean higher peaks in the GCC-PHAT.

Once an observation $\mathbf{M}$ is available, a corresponding matrix of the same size $\mathbf{S}(\mathbf{p})$ with elements $S(n, \tau|\mathbf{p})$, function of the unknown spatial position $\mathbf{p}$, can be produced by simulating with the image method the cross-correlation between the RIRs at the various microphone pairs when a source is emitting in $\mathbf{p}$. We do not consider, for the moment, the effects of source directivity and orientation (i.e. we assume here an omnidirectional source). Figure 2 shows an example of the $\mathbf{M}$ matrix measured with the line array and of the corresponding matrix $\mathbf{S}$, simulated at the correct source position, when only the direct wavefront and the first order reflections on the left wall, the right wall and the floor are taken into account (see an outline of the geometric layout in Figure 3). With the 4 considered wavefront arrivals there are 4^2 mutual delays, yielding the curves which build up the pattern of Figure 2b, where the labels indicate what wavefronts are being correlated (D=direct, L=left wall, R=right wall, F=floor). For example DR denotes the curve of mutual delays of the wavefront reflected on the right wall with respect to the direct arrival. The curve with label FF is not represented in the figure, as it is very close to the DD curve, since the D and the F wavefronts have very similar azimuth of arrival at the array. For the sake of simplicity in the simulated pattern of $\mathbf{S}(\mathbf{p})$ a unitary amplitude may be assigned to all the peaks corresponding to mutual delays. Given the observed data $\mathbf{M}$, a localization score MP-GCF($\mathbf{p}$) can then be derived at any hypothesized source position $\mathbf{p}$:

$$\text{MP-GCF}(\mathbf{p}) = \sum_n \sum_\tau M(n, \tau) \cdot S(n, \tau|\mathbf{p}) \qquad (1)$$

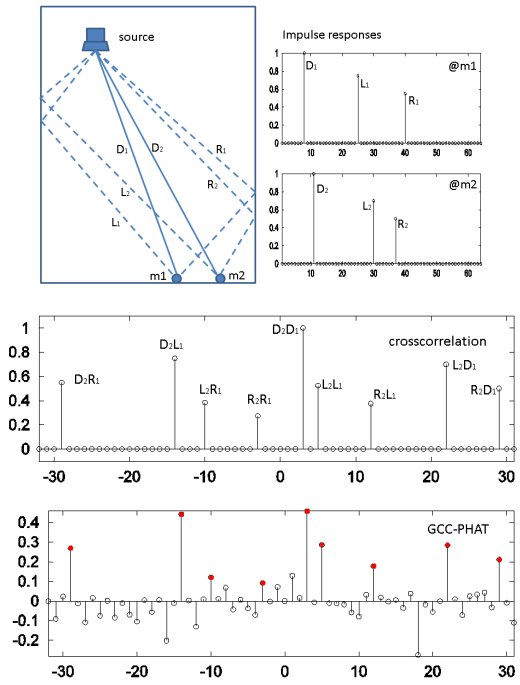


Fig. 1. Example of multiple wavefront arrivals at a microphone pair generated in the case of two reflecting surfaces. The GCC-PHAT is only an approximation of the cross-correlation between the two impulse responses.

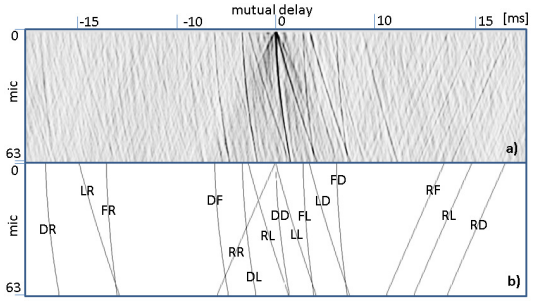


Fig. 2. a) Gray-level representation of the matrix $\mathbf{M}$ obtained from GCC-PHAT computed along the line array with microphone m_0 as fixed reference. b) Simulated cross-correlation pattern of the RIRs along the array (matrix $\mathbf{S}$). Labels indicate which wavefronts pairs are producing each curve (D denotes the direct wavefront. L, R and F denote the reflections on left wall, right wall and floor respectively).

A maximization of MP-GCF($\mathbf{p}$) provides the coordinates that are in best accordance with the pattern of mutual delays measured at the microphone array. An example of the MP-GCF acoustic map obtained from the GCC-PHAT pattern of Figure 2a and calculated on a xy plane is shown in Figure 3. Note that various contributions due to different reflection paths sum up constructively in the map to produce a peak in correspondence of the source position (highlighted by the circle).

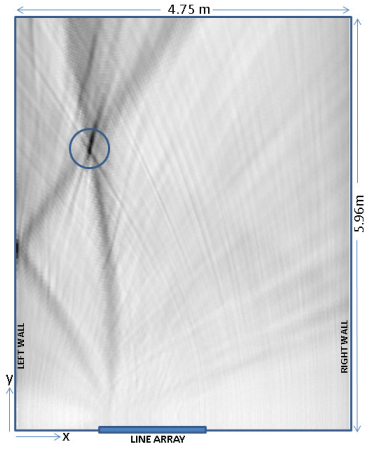


Fig. 3. Example of an acoustic map computed taking into account first order reflections on lateral walls and on the floor. The circle shows the localization of the source in correspondence of the maximum peak of the map.

4. MULTIPATH ORIENTED GLOBAL COHERENCE FIELD

From Figure 3, it is clear that some information related to the source orientation is available in the MP-GCF map. The neat reflection occurring on the left wall indicates that a considerable amount of energy is irradiated by the source toward that wall. This observation led us to investigate on the possibility to derive clues about the source orientation by analyzing the presence and the strength of the effects of reflections in the MP-GCF and hence in the $\mathbf{M}$ matrix.

Following the definition of OGCF, the MP-OGCF is introduced as a function of both source position $\mathbf{p}$ and orientation θ, through a generalization of MP-GCF. For a given $\mathbf{p}$, MP-OGCF is obtained by weighting the multipath contributions in eq. 1 as follows:

$$\text{MP-OGCF}(\mathbf{p},\theta) = \sum_{n}\sum_{\tau} M(n,\tau)\left[S(n,\tau|\mathbf{p})\cdot w(n,\tau|\mathbf{p},\theta)\right] \tag{2}$$

where the weights $w(n,\tau|\mathbf{p},\theta)$ depend on source orientation and source directivity. Basically the omnidirectional reference matrix, which only accounts for the attenuation due to path lengths and reflection coefficients associated to a given number of wavefront arrivals, is modified to account also for the source directivity and orientation by means of the weights w. Denoting as $\phi(n,\tau|\mathbf{p})$ the departure angle of the propagation path associated to the mutual delay τ at microphone pair n, a possible definition of the weights models the source radiation pattern as a cardioid elevated to a power ρ:

$$w(n,\tau|\mathbf{p},\theta) = \left(\frac{1+\cos\left(\phi\left(n,\tau|\mathbf{p}\right)\right)}{2}\right)^{\rho} \tag{3}$$

where the parameter ρ determines the degree of directivity. Note that if only the direct path is considered, eq. 2 reduces to the OGCF formulation [8].

Unfortunately eq. 2 relies on the weak information associated to the relative amplitudes of GCC-PHAT peaks associated to different propagation paths. Conversely, the original formulation of OGCF operates on the relative amplitudes of peaks associated to direct paths only, observed by different pairs. If this simplified propagation model is effective in detecting the source position through eq. 1, where mainly the presence of peaks due to reflected wavefronts matters rather than their dynamics, it is not accurate enough in the orientation estimation context. In particular, due to the GCC-PHAT non linearities, it is not sufficient to reshape the omnidirectional GCC-PHAT by applying a weighting derived from the source radiation pattern.

In order to improve the match between the data in matrix $\mathbf{M}$ and the simulated template, a new reference matrix $\mathbf{S}_{\rho}(\mathbf{p},\theta)$ is introduced, which is a simulated version of the GCC-PHAT, given a directivity parameter ρ (with the same meaning as in eq. 3), and is function of both position and orientation of the source (if $\rho = 0$ the omnidirectional version of the reference matrix is obtained). The simplified RIRs, simulated in the omnidirectional case, are hence modified so that the source directivity is accounted for. $\mathbf{S}_{\rho}(\mathbf{p},\theta)$ is obtained by computing the GCC-PHAT from the modified RIRs. In practice, weights w are directly embedded in the RIR modeling: since GCC-PHAT is non linear it is not possible to separate the weights contributions from the GCC-PHAT computation. Given the source position obtained from the simplified computation of eq. 1, the orientation is estimated by maximizing the match between the accurate orientation-dependent reference and observed matrices. Consequently, the new enhanced eMP-OGCF acoustic map is defined as follows:

$$\text{eMP-OGCF}(\mathbf{p},\theta) = \sum_{n}\sum_{\tau} M(n,\tau)\cdot S_{\rho}(n,\tau|\mathbf{p},\theta) \tag{4}$$

It is worth noting that this new method is considerably more computationally demanding than that in eq. 2 for which one reference map is computed for each spatial point. In fact, eq. 4 requires that a specific reference matrix is available for each possible orientation and position. However, as shown in the next section, an efficient solution is to apply eq. 4 only locally at the coordinates $\mathbf{p}$ found by means of eq. 1.

5. EXPERIMENTAL RESULTS

A set of experiments on real and simulated data was conducted to validate the proposed approach. Performance is measured in terms of average absolute estimation error on the orientation, expressed in degrees. Results are reported for the methods described by eq. 2 and by eq. 4.

5.1. Real data

The 64-channel line array (a modified NIST MarkIII [18]) was installed in a room of $width \times length \times height = 4.75m \times 5.96m \times 4.4m$, parallel to the x axis and with the reference microphone m_0 at coordinates $x = 1.25$, $y = 0$, $z = 1.65$. The room has a reverberation time T_{60} of about 0.7 seconds and the shape of a parallelepiped but, while the floor and the walls are quite regular surfaces (concrete walls with no furniture), the ceiling is rather articulated with hanging lattices and supports which prevent a clean reflection of sound.

	Average absolute error	
Pairs	eq.2	eq.4
$N = 64$	18.3°	14.1°
$N = 8$	18.6°	14.5°

Table 1. Average absolute errors on the real data set when considering all microphone pairs or a subset of them.

Therefore, in this study we did not consider or model any reflection from the ceiling. A loudspeaker was recorded while emitting a speech sentence (about 6 seconds, male speaker, SNR $\simeq$ 25 dB) from 20 different positions on a grid with 1 meter spacing in x and y at an height of 1.55 m from the floor. At each position 5 different orientations were adopted with azimuth angles of 0°,$\pm 30^\circ$ and $\pm 60^\circ$, for a total of 100 source emissions. Figure 4 sketches the experimental set up and the source positions and orientations.

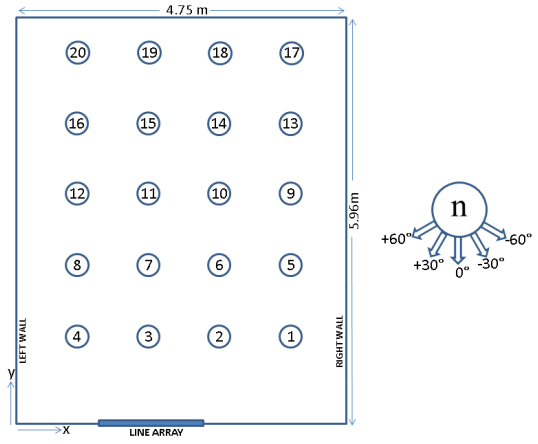
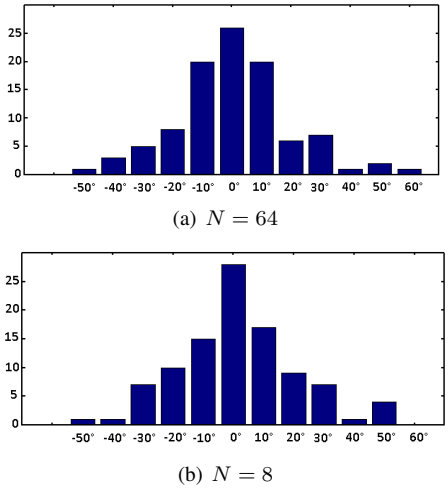


Fig. 4. Outline of the experimental room with the 20 positions x 5 orientations of the loudspeaker.

Given the source position estimated by means of the MP-GCF presented in [5] (which provides an average accuracy of 7 cm over the 100 localizations), a GCC-PHAT model is computed for each pair on a range of possible emission angles extending from -90 to 90 degrees with step 10 degrees, resulting in 19 orientation-dependent maps, using a directivity parameter $\rho = 4$ (which approximates the actual radiation pattern of the loudspeaker). The models account for 3 wavefront arrivals: direct wave, reflection on the left wall and reflection on the right wall. Signals were recorded at 44100 Hz and GCC-PHAT was evaluated using blocks of 2^{17} samples. Note that this long analysis window (i.e. 3 seconds), necessary to obtain reliable peaks in the GCC-PHAT, limits the applicability of the method to stationary sources.

Besides using the full array, a set of experiments was conducted considering a subset of 8 microphones (i.e. a decimated version of the array). Table 1 reports the average estimation error (absolute value in degrees) for $N = 64$ and $N = 8$ when applying eq. 2 and eq. 4. A clear improvement is achieved when accounting for the source orientation directly in the reference matrix as done in eq. 4. Note also that only a minor performance loss is observed when limiting to 8 the number of available microphone pairs. Figure 5 reports the distribution of errors for the approach of eq. 4: the majority of them are within 10° and a slight increase in the distribution variance is observed when $N = 8$.



(a) $N = 64$

(b) $N = 8$

Fig. 5. Distribution of estimation errors when using 64 (a) and 8 (b) pairs with the estimation approach of eq. 4.

5.2. Simulated data

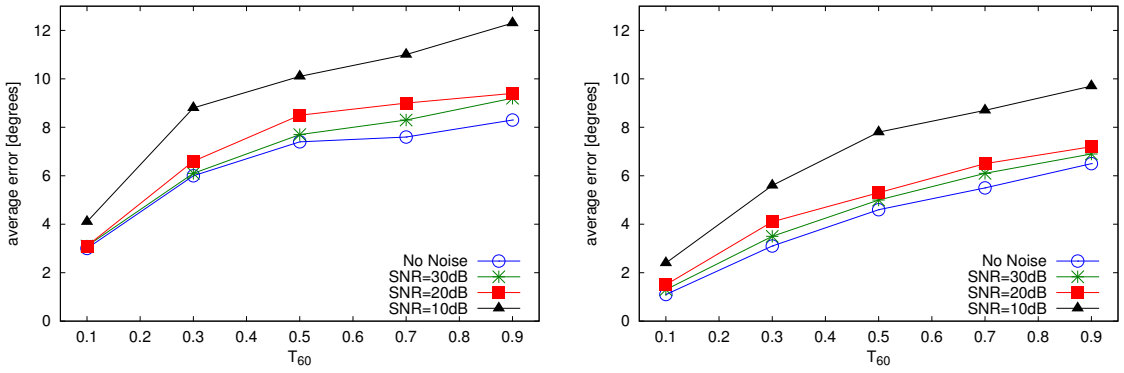
In order to analyze the effects of different environmental conditions on the enhanced method of eq. 4, a configuration as close as possible to the real one was simulated for different amounts of reverberation and additive white gaussian noise. Signals at the various microphones were obtained by convolution of a clean speech sequence with impulse responses generated by means of the image method accounting for the directivity of the source. Reverberation time values ranged between 0.1 and 0.9 seconds with four different values of SNR at the microphones: ∞, 30 dB, 20 dB and 10 dB. Two directivity patterns were considered, corresponding to ρ=2 and ρ=4. Note that, with respect to the real case discussed above, in this experimental analysis the propagation model perfectly matches the simulated RIR (it differs only for the limited amount of paths considered, i.e. 3 wavefront arrivals).

Experimental results are shown in Figure 6. First of all, note that the estimation accuracy is higher for the most directive source for which even under very challenging conditions the average absolute error is well below 10°. Instead, when a less directional source is employed the estimation error increases (2° approximately). This is explained by the fact that as the directivity of the source increases the reverberated energy captured by microphone is reduced (i.e. increasing the Direct-to-Reverberant ratio). A further aspect to take into consideration is that as the source approaches an omnidirectional one the difference between the energy irradiated along different propagation paths decreases, reducing the discriminative power of the GCC-PHAT features adopted here.

Interestingly, the proposed method is robust against environmental noise and only a minor performance loss is observed when the SNR is about 10 dB.

6. CONCLUSIONS

In this paper we presented a novel method to estimate the orientation of a non-omnidirectional acoustic source using a line array and



(a) $\rho = 2$ (b) $\rho = 4$

Fig. 6. Average absolute error on the synthetic data set for two radiation patterns: $\rho = 2$ in (a) and $\rho = 4$ in (b).

taking advantage of the multipath propagation. By extending the previously presented MP-GCF, the pattern of reflected wavefronts is exploited through a simple but effective modeling of the acoustic propagation, which accounts for the radiation properties of non-omnidirectional sources. Two variations are presented which differ for the computational cost and for performance.

Experimental results on simulated and real data show that the proposed method can effectively estimate the orientation of stationary sources, with an accuracy of few degrees, even under challenging environmental conditions in terms of noise and reverberation. Experiments on real data prove that the proposed method is applicable in practical situations, in which the required environment awareness (i.e. room geometry, wall absorption coefficient, source position and emission properties) is available only with the limited accuracy typical of manual measurements or state of the art estimation methods.

7. REFERENCES

[1] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Applied Signal Processing*, vol. 2006, no. 4, pp. 1–19, 2006.

[2] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1479–1489, November 2008.

[3] F. Ribeiro, C. Zhang, D.A. Florencio, and D.E. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1781–1792, September 2010.

[4] T. Korhonen, "Acoustic localization using reverberation with virtual microphones," in *Proceedings of IWAENC*, 2008.

[5] P. Svaizer, A. Brutti, and M. Omologo, "Use of reflected wavefronts for acoustic source localization with a line array," in *Proceedings of HSCMA*, 2011.

[6] P. Svaizer, A. Brutti, and M. Omologo, "Analysis of reflected wavefronts by means of a line microphone array," in *Proceedings of IWAENC*, 2010.

[7] C. H. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoustic, Speech and Signal Processing*, vol. 24, no. 4, pp. 320–327, August 1976.

[8] A. Brutti, M. Omologo, and P. Svaizer, "Speaker localization based on Oriented Global Coherence Field," in *Proceedings of Interspeech*, 2006, pp. 2606–2609.

[9] K. Nakadai et al., "Sound source tracking with directivity pattern estimation using a 64 ch microphone array," in *Intelligent Robots and Systems (IROS)*, 2005.

[10] M. Togami and Y. Kawaguchi, "Head orientation estimation of a speaker by utilizing kurtosis of a doa histogram with restoration of distance effect," in *Proceedings of ICASSP*, 2010.

[11] B. Mungamuru and P. Aarabi, "Enhanced sound localization," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 3, pp. 1526 –1540, June 2004.

[12] A. Nakano, S. Nakagawa, and K. Yamamoto, "Automatic estimation of position and orientation of an acoustic source by a microphone array network," *Journal of Acoustic Society of America*, vol. 126, no. 6, pp. 3084–3094, 2009.

[13] A. Levi and H. Silverman, "A robust method to extract talker azimuth orientation using a large-aperture microphone array," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 277 –285, February 2010.

[14] K Niwa et al., "Estimation of sound source orientation using eigenspace of spatial correlation matrix," in *Proceedings of ICASSP*, 2010.

[15] A. Brutti, M. Omologo, and P. Svaizer, "Inference of acoustic source directivity using environment awareness," in *Proceedings of EUSIPCO*, 2011.

[16] R. DeMori, Ed., *Spoken Dialogue with Computers*, Academic Press, 1998.

[17] M. Brandstein and D. Ward, *Microphone Arrays*, Springer Verlag, 2001.

[18] L.G. Brayda, C. Bertotti, L. Cristoforetti, M. Omologo, and P. Svaizer, "Modifications on NIST MarkIII array to improve coherence properties among input signals," *118th Audio Engineering Society Convention*, 2005.