INFERENCE OF ACOUSTIC SOURCE DIRECTIVITY USING ENVIRONMENT AWARENESS

A. Brutti, M. Omologo and P. Svaizer

Fondazione Bruno Kessler CIT-irst
via Sommarive 18, 38123 Trento, Italy
phone: + 39 0461314529, email: brutti@fbk.eu

ABSTRACT

Acoustic maps derived from the Generalized Cross-Correlation Phase Transform (GCC-PHAT) computed on the signals acquired by a set of distributed microphones can be effectively used for the localization of active acoustic sources. When the microphone pairs surround a given area with a good angular coverage, directional characteristics of the sources can also be inferred, based on the relative amplitudes of the GCC-PHAT peaks and the geometry of propagation in the given environment. This paper presents a novel method for estimating the radiation pattern of an acoustic source which combines GCC-PHAT observations with accurate descriptors of the environment characteristics, i.e. reverberation time. Experiments on simulated data show that the source emission pattern can be estimated in an effective way under noisy and reverberant conditions.

1. INTRODUCTION

Acoustic scene analysis for audio processing takes advantage of an accurate characterization of the sound sources, for instance in terms of spatial positioning and radiation properties. As far as the latter is concerned, in general acoustic sources present emission patterns which are far from being omnidirectional in particular at higher frequencies [10]. As an example, Figure 1 sketches the typical radiation pattern of humans at two frequencies [5, 7].

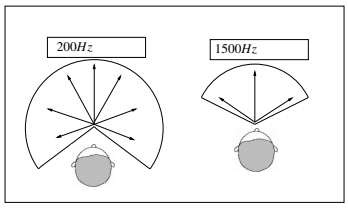


Figure 1: Rough shape of the head radiation pattern at two frequencies.

To this regard, the SCENIC project[1] focuses on techniques for achieving acoustic awareness of an environment and on methods for employing the acquired awareness in source characterization and enhancement of the recorded signals. The specific radiation pattern, and consequently the orientation of the source, are crucial features in many speech related algorithms which rely on the assumption that direct wave-fronts prevail in the multipath propagation [3, 16]. The source emission pattern plays a double role since it influences not only the direct path but also the whole Room Impulse Response (RIR), controlling the amount of energy irradiated along each propagation path. Consequently, the RIR, which fully describes the point-to-point acoustic propagation in a given enclosure, could be used to extrapolate the properties of the source emission pattern. An accurate derivation of the RIR from the captured audio signal is possible only if the originally emitted sound is known. Alternatively blind channel identification methods are adopted [8], which typically operate using arrays of microphones. In this work instead we are interested in using the information provided by the GCC-PHAT [9] computed at several surrounding microphone pairs. It has been shown that such information is related to the orientation of a non omnidirectional source and that it can be employed to infer the direction of sound emission [4]. The radiation pattern influences in a very articulated way the behaviour of the GCC-PHAT, affecting not only the direct path but the whole multipath propagation. Therefore, deriving a direct relationship between the emission pattern and the observed GCC-PHAT measurements is not trivial. In this work we suggest adopting an environment aware method, which, provided that accurate descriptors of the environment are available, approximates the acoustic propagation by considering low order microphone mirrors [2] and consequently models the corresponding expected GCC-PHAT. A similar approach was followed in [14] which presents a method, relying on first order mirror images, for source orientation estimation using eigenvalues of the cross-correlation matrix.

The problem of estimating the radiation pattern of an acoustic source has not received much attention by the research community so far. Few works are available in the literature on the topic. In [11] the acoustic source is approximated as a circular piston whose radius determines the emission pattern. Based on the energy received at a microphone array the piston radius is derived. More recently in [12] a rough representation of the radiation pattern, in combination with the source position and orientation, is derived from a weighted delay-and-sum beamforming.

This paper is organized as follows. Section 2 introduces the adopted acoustic propagation model in presence of a non omnidirectional source. Section 3 presents the maximum likelihood estimation framework whose performance, measured on simulated data, is discussed in Section 4. A final discussion concludes the paper.

2. ACOUSTIC PROPAGATION WITH DIRECTIVE SOURCES

Let us assume that N microphones are available in a given enclosure where a directional acoustic source is active with a certain orientation. Each microphone is identified by the couple (r_n, θ_n) indicating its distance r_n and azimuth θ_n with respect to the source (see Figure 2). The azimuth corresponds to the angular distance between the direction the source is aimed at and the line connecting the source and the microphone. Assuming an FIR modeling of the RIR $h_n(t)$ between the source and the n-th microphone, a common approach is to split the acoustic propagation into the direct path and the reverberated part:

$$h_n(t) = h_{n,d}\,\delta(t - r_n/c) + h_{n,r}(t) \quad (1)$$

where the attenuation factor $h_{n,d}$ includes both the propagation loss and the directivity gain (in case the source is not omnidirectional), while c is the speed of sound. The reverberant part $h_{n,r}(t)$ includes

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

[1] Further details are available at: http://www.thescenicproject.eu

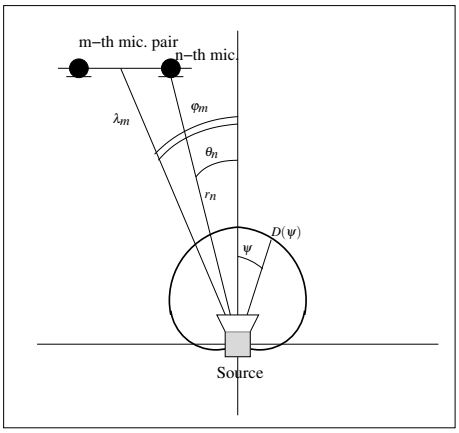


Figure 2: Reference diagram for the notation adopted in the paper.

all the other arrivals due to the multipath propagation.
Now, let us consider M microphone pairs obtained out of the N microphones and denote with λ_m and φ_m the distance and azimuth of the center of the m-th pair. Given the geometrically computed time difference of arrival $\tau_m(\mathbf{p})$ at microphone pair m for a source in spatial position $\mathbf{p}$, we indicate with $\gamma_m(t)$ the value of the GCC-PHAT C_m at microphone pair m computed for time lag τ_m at time instant t:

$$\gamma_m(t) = C_m(t, \tau_m(\mathbf{p})) \tag{2}$$

It can be shown that under diffuse reverberation and ideal propagation, $\gamma_m(t)$ is proportionally related to the Direct-to-Reverberant ratio DRR at pair m. If the emitting source has directional properties, the DRR is influenced not only by the multipath propagation but also by the specific azimuthal distance between source and microphones (i.e. the angle θ_n in Figure 2). Even though the radiation properties of an acoustic source are frequency dependent, for the sake of simplicity we consider here an average directivity gain $D(\psi)$ on a horizontal plane, defined in a way that its maximum is normalized to 1 at $\psi = 0$ and is proportional to the overall power radiated in the various directions for azimuth $-\pi \leq \psi < \pi$. The attenuation factor associated to the direct propagation path can hence be defined as [10]:

$$h_{n,d} = \frac{D(\theta_n)}{4\pi r_n} \tag{3}$$

In case of directional microphones the receiver gain should be included in the above equation. According to eq. 3, if a microphone pair is not frontal to a directional source, the DRR, and hence $\gamma_m(t)$, is reduced leading to the derivation of cues related to the source orientation or emission characteristics [13, 18].

Unfortunately a closed form relationship between $\gamma_m(t)$ and φ_m for a given source directivity gain is not available. In particular, the source directivity affects also the reverberant part of the RIR, since some paths are receiving lower energy, which, in combination with the non-linearity of the GCC-PHAT, complicates the problem considerably. Nonetheless, it is possible to take advantage of the connection between $\gamma_m(t)$ and φ_m to classify the source directivity pattern using a sufficient number of observations with a reasonable angular resolution. Only a relative comparison between the emission patterns of two different sources irradiating sound in the same conditions is actually achievable in practice. As a matter of fact, GCC-PHAT is influenced by so many factors that just slightly changing the position of the source would result in a quite different distribution of $\gamma_m(t)$ for any possible angle φ_m, preventing a practical derivation of an absolute descriptor of the radiation pattern.

3. DIRECTIVITY ESTIMATION

A more effective solution can be devised if knowledge of the environment is embedded in terms of reverberation time (or wall absorption coefficients), source position and orientation, and microphone locations. Although a plethora of robust algorithms related to source localization [19] and estimation of source orientation exists [4, 13, 18], only recently methods enabling the inference of acoustic properties of an enclosure have appeared in the literature. In [15] a method for blind estimation of the reverberation time is presented, while in [6, 17] methods for reflector characterization and room geometry inference are described. If that information was available, the RIR at the n-th microphone could be approximated using the image method [1] by mirroring the microphone position with respect to the 6 surfaces of a parallelepipedic enclosure up to the images of order I. Denoting with (n,i) the i-th mirror image of the n-th microphone, an approximated RIR can be obtained as:

$$\hat{h}_n^I(t) = \sum_{i=0}^{6^I} h_{n,i}\,\delta(t - r_{n,i}/c) \tag{4}$$

Assuming a constant absorption coefficient β for all the surfaces, the attenuation $h_{n,i}$ of the propagation path associated to the i-th image of the n-th microphone is:

$$h_{n,i} = \beta^{w_{n,i}} \frac{D(\theta_{n,i})}{4\pi r_{n,i}} \tag{5}$$

where $w_{n,i}$ is the number of reflections associated to the specific propagation path, $\theta_{n,i}$ is the azimuth of the image microphone and $r_{n,i}$ is the length of the path. Note that $i = 0$ refers to the direct path, i.e. when the microphone is not mirrored. The above derivation can be easily generalized for the case in which β is not constant. From the approximated RIRs $\hat{h}_n^I(t)$, an expected value of the GCC-PHAT at a generic microphone pair m can be derived as [9]:

$$\hat{\gamma}_m = \int_{-\infty}^{\infty} \frac{\hat{H}_{m1}(f)\hat{H}_{m2}^*(f)}{|\hat{H}_{m1}(f)||\hat{H}_{m2}(f)|} e^{j2\pi f \tau_m} \mathrm{d}f \tag{6}$$

where $\hat{H}_{m1}$ and $\hat{H}_{m2}$ are the Fourier Transforms of the two approximated RIRs of the microphones forming the m-th pair. Given $\hat{\gamma}_m$, it is reasonable to model the observed $\gamma_m(t)$ as Gaussian random variables:

$$\gamma_m(t) \in \mathcal{N}(\hat{\gamma}_m, \sigma_\gamma)$$

where, for the sake of simplicity, the standard deviation σ_γ is assumed to be equivalent at all microphone pairs. Consequently we can set up a maximum likelihood framework for estimating the directivity pattern of the emitting source. To simplify the estimation task we parametrize the average emission pattern as follows:

$$D(\psi, \rho) = \left(\frac{1 + \cos(\psi)}{2}\right)^\rho \tag{7}$$

which is basically a cardioid-like emission pattern where the exponent ρ determines the directivity of the source (when $\rho = 0$ the emission pattern is omnidirectional). The above model allows us to define the expected value $\hat{\gamma}_m$ as function of ρ: $\hat{\gamma}_m(\rho)$. Given the set $\boldsymbol{\gamma}(t) = [\gamma_0(t), \gamma_1(t), \ldots, \gamma_{M-1}(t)]$ of GCC-PHAT observations at time instant t, we are interested in maximizing the probability $P(\rho|\boldsymbol{\gamma}(t))$ of ρ given $\boldsymbol{\gamma}(t)$ for any time instant t. For the sake of simplicity we drop hereafter the dependency on time. Applying the Bayes rule and under the assumption that observations are independent and identically distributed, the solution is equivalent to maximizing $P(\boldsymbol{\gamma}|\rho)$ which can be written as:

$$P(\boldsymbol{\gamma}|\rho) = \prod_m P(\gamma_m|\rho) \tag{8}$$

$$= \prod_m \frac{1}{\sqrt{2\pi}\sigma_\gamma} e^{-\frac{1}{2}\left(\frac{\gamma_m - \hat{\gamma}_m(\rho)}{\sigma_\gamma}\right)^2} \tag{9}$$

Applying logarithm and removing constants and terms not depending on ρ, we can reformulate the above equation as:

$$\mathscr{L}(\boldsymbol{\gamma}|\rho) = \sum_m \left(2\gamma_m \hat{\gamma}_m(\rho) - \gamma_m^2 - \hat{\gamma}_m^2(\rho)\right) \quad (10)$$

from which we can define the maximum likelihood estimator for the directivity parameter ρ:

$$\overline{\rho}(t) = \arg\max_{\rho} \mathscr{L}(\boldsymbol{\gamma}|\rho) \quad (11)$$

Note that the estimation obtained in eq. 11 is for a specific time instant and uses a single set of observations $\boldsymbol{\gamma}(t)$.

Finally, it is worth mentioning that the proposed method can work only in presence of reverberation. As a matter of fact, under anechoic conditions all the emission patterns are equivalent to an omnidirectional one from the GCC-PHAT perspective. The model $\hat{h}_n^I(t)$ must hence capture the reverberation properties of the environment considering at least the low order reflections (i.e. $I > 0$). This way the adopted method is also robust against the presence of strong reflections that may artificially increase some of the $\hat{\gamma}_m$ resulting in erroneous estimates of the emission pattern.

4. EXPERIMENTAL ANALYSIS

To validate the proposed estimator, we simulated a dense microphone distribution where omnidirectional sensors are positioned all around the walls of a 5x6 meters room. Given an inter-microphone distance of 20cm, 110 microphones are considered, all of them at the same height. Each microphone is coupled for GCC-PHAT computation with the adjacent one resulting in $M = 110$ pairs. Figure 3 shows a simplified outline of the experimental set up and the source position. For each microphone a RIR at 16kHz is generated

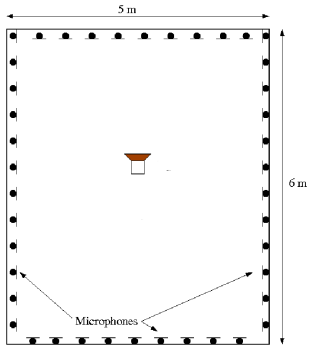


Figure 3: Simplified outline of the experimental set up (with only 42 microphones).

through the image method considering various reverberation times and 5 different values of the directivity parameter $\rho : 0, 2, 4, 6, 8$. Figure 4 shows the 5 resulting radiation patterns in dB: note that the two most directive patterns are quite similar. The resulting RIRs are

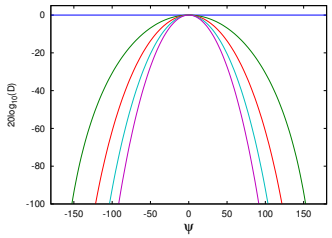


Figure 4: The 5 emission patterns taken into consideration.

used to filter a speech utterance of 4 seconds, in order to generate the signals as acquired by the microphones. White noise is then added to account for the effects of environmental noise considering 3 possible SNRs: 30dB, 20dB and 10dB. Given an analysis window of 2^{14} samples with overlap 50%, a sequence of GCC-PHAT observations is extracted from the acquired signals and is used to obtain independent estimates of the parameter ρ through eq. 11. Expected GCC-PHAT values are computed using 1st and 2nd order images. Since a closed-form solution of eq. 10 is not achievable, a grid of possible ρ is considered in the range $[0, 9.5]$ with step 0.5. The position of the source is assumed to be known, for instance as output of a source localization algorithm. We further assume that the reverberation time is known, from which β can be derived through the Sabine's formula. Performance is measured in terms of the estimation error $\varepsilon_\rho =$ averaged over all time frames: In addition, the average estimation error $\overline{\varepsilon}_\rho$, obtained as average over all directivities for a given reverberation time and SNR is evaluated:

$$\overline{\varepsilon}_\rho = \frac{1}{5}\sum_\rho |\varepsilon_\rho| \quad (12)$$

4.1 Results

Figure 5 reports on the estimation error for various reverberation times (T_{60}) and SNRs when only the 1st order mirrors are accounted for. Note that most of the errors are concentrated in the highly directive cases, with $\rho = 8$ being the most critical directivity to estimate. In particular, we observe that as the reverberation increases the directivity tends to be underestimated since the adopted propagation model does not fit the actual propagation anymore. Conversely, low directivities are overestimated ($\varepsilon_\rho > 0$). Figure 6 summarizes the

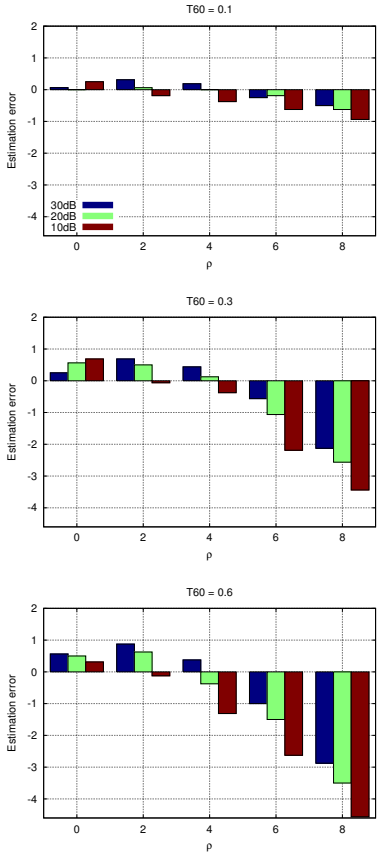


Figure 5: Estimation error ε_ρ when using only 1st order mirrors to derive the GCC-PHAT model. Results for different values of ρ, T_{60} and SNR are reported.

performance in terms of average error over the 5 directivity patterns. As expected, when reverberation increases the performance progressively degrades with a clear gap between 10dB SNR and the other 2 cases. This is a consequence of the corruption of phase information induced by an increased level of additive noise which is not captured by the adopted GCC-PHAT model.

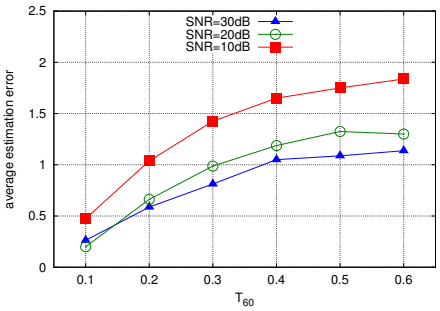


Figure 6: Average estimation error when using 1st order reflections.

In a similar way, Figures 7 and 8 present the results when the second order reflections are considered in the GCC-PHAT models. Note how the related performance is more robust against reverberation and noise, in particular when the SNR is above 10dB. Conversely, when the SNR is equal to 10dB the performance is similar to the case with $I = 1$ and no gain is provided by using 2nd order mirror images.

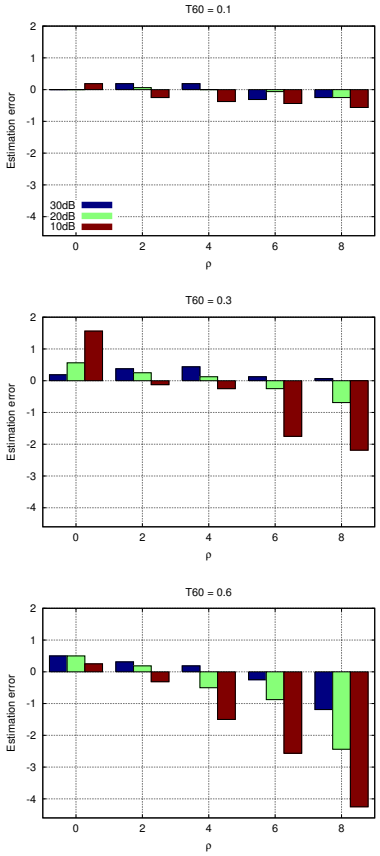


Figure 7: Estimation error when using 2nd order mirrors under different reverberation times and SNRs.

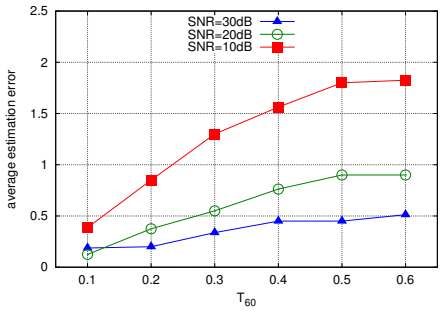
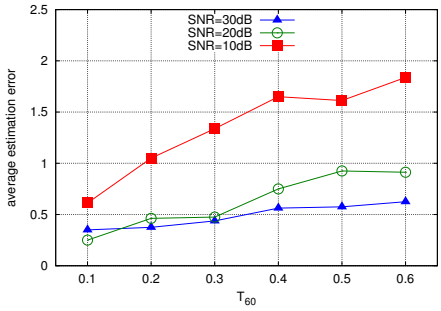
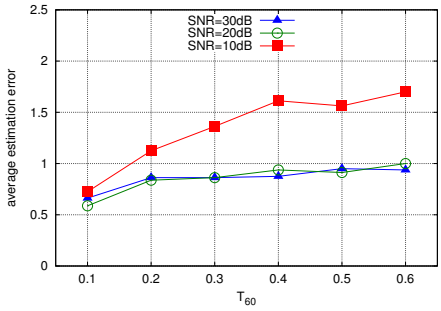


Figure 8: Average estimation error when $I = 2$.

So far the proposed method was evaluated using a very dense microphone distribution which ensures a fine sampling of the GCC-PHAT in the angular domain. Considering the specific case in which SNR=20dB, $T_{60} = 0.3s$ and $I = 2$ we evaluate now the estimator capabilities when the number of available pairs is reduced of factors 5 and 10, maintaining the same intra-pair microphone distance and a uniform distribution of the pairs in space. Figure 9 reports the average estimation error. When $M = 22$ only a slight performance reduction with respect to $M = 110$ is observed (Figure 9(a)). Conversely, as shown in Figure 9(b), using 11 pairs results in an apparent performance degradation in particular under high SNR and low reverberation conditions.



(a) M=22



(b) M=11

Figure 9: Average estimation error when $I = 2$ and a reduced number of microphone pairs is adopted.

Clearly the adopted method relies on an accurate estimation of the reverberation time from which the wall absorption coefficients can be approximated using the Sabine's formula. Figure 10 reports on the average estimation error when a mismatch occurs between the reverberation time adopted for computing the models and the actual one, which in this specific case is equal to 0.3 second. The

estimation performance does not seem to be very sensitive to overestimated T_{60} while an underestimated reverberation time is quite detrimental. Interestingly, in the most noisy case (i.e. SNR=10dB) the best performance is reached when T_{60} is slightly lower than the actual one (0.2 s).

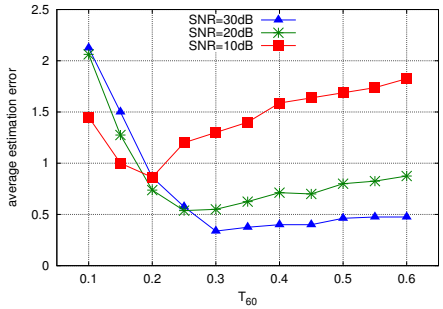


Figure 10: Average estimation error when using 2nd order reflections, with a mismatch between T_{60} used for modeling and the actual one, which in this case is 0.3s.

5. CONCLUSION

This paper presented a novel method for deriving the radiation pattern of an acoustic source given a set of GCC-PHAT measurements obtained from several microphone pairs. The estimation technique works in a maximum likelihood fashion using GCC-PHAT models which are obtained by approximating the RIRs with the image method accounting for low order reflections. Environment awareness in terms of room geometry and reverberation time is employed in the RIR approximation.

Extensive numerical simulations show that the proposed method can estimate the source directivity with satisfactory accuracy in low to mid reverberation and low SNR conditions. Errors concentrate in most of the cases in highly directive patterns that tends to be underestimated. To some extent, the estimation is also robust to the use of erroneous reverberation times.

Concerning the required microphone density, it is worth mentioning that when few microphones are available their angular density can be artificially increased by accumulating over time the GCC-PHAT observations associated to a moving source.

Future work will address the robustness of the average radiation pattern model in case of real sources, either loudspeakers or humans. The robustness of the estimator will be tested in real environments. The possibility to jointly estimate the radiation pattern and the source orientation is another challenging research direction that will be addressed.

REFERENCES

[1] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of Acoustic Society of America*, 65(4):943–950, April 1979.

[2] P. Bergamo, S. Asgari, H. Wang, D. Maniezzo, L. Yip, R. E. Hudson, K. Yao, and D. Estrin. Collaborative sensor networking towards real-time acoustical beamforming in free-space and limited reverberance. *IEEE Transactions on Mobile Computing*, 3:211–224, July 2004.

[3] T. Betlehem and R. Williamson. Acoustic beamforming exploiting directionality of human speech sources. In *Proc. of ICASSP*, 2003.

[4] A. Brutti, M. Omologo, and P. Svaizer. Oriented Global Coherence Field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays. In *Proceedings of Interspeech*, 2005.

[5] W. Chu and A. C. Warnock. Detailed directivity of sound fields around a human talkers. Technical report, National Research Council Canada, December 2002. URL: http://irc.nrc-cnrc.gc.ca/pubs/rr/rr104/.

[6] J. Filos, E. Habets, and P. A. Naylor. A two-step approach to blindly infer room geometries. In *Proc. International Workshop on Acoustic Echo and Noise Control*, 2010.

[7] J. L. Flanagan. Analog measurements of sound radiation from the mouth. *The Journal of the Acoustical Society of America*, 32(12):1613–1620, December 1960.

[8] E. Habets and P. A. Naylor. An online quasi-newton algorithm for blind simo identification. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2010.

[9] C. H. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 24(4):320–327, August 1976.

[10] H. Kuttruff. *Room Acoustics*. Elsevier Applied Science, 1991.

[11] P. C. Meuse and H. F. Silverman. Characterization of talker radiation pattern using a microphone array. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 257–260, 1994.

[12] K. Nakadai, H. Nakajima, K. Yamada, Y. Hasegawa, T. Nakamura, and H. Tsujino. Sound source tracking with directivity pattern estimation using a 64 ch microphone array. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005*, pages 1690 – 1696, 2005.

[13] A. Y. Nakano, S. Nakagawa, and K. Yamamoto. Automatic estimation of position and orientation of an acoustic source by a microphone array network. *The Journal of the Acoustical Society of America*, 126(6):3084–3094, 2009.

[14] K. Niwa, Y. Hioka, S. Sakauchi, K. Furuya, and Y. Haneda. Estimation of sound source orientation using eigenspace of spatial correlation matrix. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 129 –132, 2010.

[15] R. Ratnam, D. Jones, B. W. andd W. O'Brien, C. Lansing, and A. Feng. Blind estimation of reverberation time. *The Journal of the Acoustical Society of America*, 2003.

[16] S. T. Shivappa, B. D. Rao, and M. M. Trivedi. Role of head pose estimation in speech acquisition from distant microphones. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 3557–3560, 2009.

[17] P. Svaizer, A. Brutti, and M. Omologo. Analysis of reflected wavefronts by means of a line microphone array. In *Proc. International Workshop on Acoustic Echo and Noise Control*, 2010.

[18] M. Togami and Y. Kawaguchi. Head orientation estimation of a speaker by utilizing kurtosis of a doa histogram with restoration of distance effect. In *IEEE International Conference on Acoustics Speech and Signal Processing*, pages 133–136, 2010.

[19] M. Wölfel and J. McDonough. *Distant Speech Recognition*. John Wiley and Sons, 2009.