COMMON-ACOUSTICAL-POLE PREPROCESSING FOR SOURCE LOCALIZATION IN REVERBERANT ROOMS

Raffaele Parisi, Riccardo Gazzetta and Elio D. Di Claudio
INFOCOM Dept., University of Rome "La Sapienza",
Via Eudossiana 18, I00184, Rome, Italy
Tel: +39 06 44585486; fax: +39 06 4873300
e-mail: parisi@infocom.uniroma1.it

ABSTRACT

Localization of acoustic sources has become an important task in many practical applications. Typically the position of a source is determined from preliminary estimation of a set of relative time delays between couples of sensors, followed by triangulation. Unfortunately the presence of reverberation degrades the performance of most time delay estimators, thus leading to grossly incorrect location estimates. Algorithms based on the Generalized Cross Correlation are commonly used, but show clear limitations even in the presence of low reverberation levels. More complex approaches are possible, but they require a higher computational cost.
In this paper the common-acoustical-pole modeling of room transfer functions is exploited to get improved time delay estimates with respect to generalized cross-correlation approaches, with reduced computational cost. Experimental tests are described to demonstrate the effectiveness of the proposed solution in combating reverberation, while maintaining the simplicity requirements needed in many practical situations.

1 INTRODUCTION

Microphone arrays have become a powerful tool in a variety of applications [1]. In particular, microphone arrays are often employed to localize acoustic sources in a generic environment with a high degree of accuracy [1]. Usually, source localization proceeds from a preliminary estimation of the relative Time Delays of arrival (TDs) between signals acquired by pairs of microphones. The position of the source is then estimated by geometrical considerations.

Most common approaches to TD estimation make use of the Generalized Cross Correlation function (GCC) [2], which is based on a single-delay model. Unfortunately, in the presence of multiple reflections and/or primary sources [3], GCC methods often fail or have unsatisfactory performance. In particular, the single-delay model is inadequate even for low reverberation levels, so that more effective approaches are required.

It must be kept in mind that alternative solutions should keep computational and calibration requirements at a reasonably low level, since most real-time applications of microphone arrays call for low-cost systems.

In this paper a novel prefiltering technique is proposed in order to overcome the limits of conventional GCC methods. The described approach takes into account the physics of acoustic propagation in closed environments and requires only a preliminar calibration of the system. More specifically, the concept of *Common Acoustical Poles* (CAP) [4] is exploited to filter out the contributions of room resonant modes from microphone responses. Experiments are described to illustrate the improvements of the proposed method both in terms of TD and source position estimates.

2 GENERALIZED CROSS CORRELATION

GCC methods have become very popular in TD estimation problems [2] and constitute the benchmark for new algorithms. In this section we briefly recall the main concepts of GCC.

GCC is based on a single-source and single-delay model:

$$\begin{cases} x_1(t) = s(t) + n_1(t) \\ x_2(t) = s(t+D) + n_2(t) \end{cases} . \tag{1}$$

$x_i(t)$ is the output of the i-th mic, $s(t)$ is the transmitted signal, $n_i(t)$ is the noise, assumed to be white and uncorrelated between microphones and D is the relative delay.

The GCC between $x_1(t)$ and $x_2(t)$ is [2]:

$$R^{(g)}_{x_1x_2}(d) = \int_{-\infty}^{+\infty} \Psi_g(f) G_{x_1x_2}(f) e^{j2\pi f d} df, \tag{2}$$

where $G_{x_1x_2}(f)$ is the Cross-Power Spectrum of $x_1(t)$ and $x_2(t)$ and $\Psi_g(f)$ is a proper frequency-domain weighting function. D is estimated as

$$\hat{D} = arg \max_d [R^{(g)}_{x_1x_2}(d)]. \tag{3}$$

The function $\Psi_g(f)$ is introduced to ensure a large peak in the autocorrelation function in non-ideal conditions [2]. Different choices for $\Psi_g(f)$ were explored in literature, all of them being based on estimation of the spectral characteristics of received signals. Among the possible solutions, the Maximum Likelihood (ML) weighting function has been frequently employed [2]. It can be noted that $\Psi_g(f)$ in (2) is equivalent to a prefilter applied to received signals.

Unfortunately, performance of GCC-based TD estimators seriuosly degrades in the presence of reverberation [5], since the assumed model (1) becomes inadequate to describe the acoustical phenomenon. A possible improvement can be searched in adding a proper preprocessing of microphone signals, to reduce the strength of reflections, prior to application of the GCC. An interesting solution is offered by CAP modeling of room transfer functions.

3 COMMON-ACOUSTICAL-POLE PREPROCESSING

Room Transfer Functions (RTFs) are dominated by a countable set of modes that exist only in the neighbourhood of certain *eigenfrequencies*. They can be approximated in the discrete-time domain by a zero/pole (e.g. ARMA) rational function [4]. In particular, the CAP model introduced in [4] is based on the assumption that poles represent the mode frequencies of the room, that do not depend on source and receiver positions under mild conditions. Zeros depend instead on time delays and anti-resonances [4].

The general ARMA model for the generic RTF (referring to the i-th microphone) is:

$$H_i(z) = \frac{B_i(z)}{A_{CAP}(z)} = \frac{\sum_{n=0}^{Q} b_i(n) z^{-n}}{1 - \sum_{n=1}^{P} a_{CAP}(n) z^{-n}}. \qquad (4)$$

Information about the TD between received signals is contained only in the numerator of (4) [4].

This model was successfully applied to the problem of RTF equalization [6] and suggests a novel prefiltering technique for TD estimation problems in the presence of reverberation. Specifically, from (4) the following filter $F(z)$ can be devised to pre-process received signals:

$$F(z) = A_{CAP}(z) = 1 - \sum_{n=1}^{P} a_{CAP}(n) z^{-n}. \qquad (5)$$

Use of (5) requires prior estimation of CAPs, that can be done from a sufficiently high number of room responses, measured in a number of *fixed* locations inside the room. In particular, estimation of AR parameters in (4) can be performed by forming an overdetermined set of *Yule-Walker* equations [7]. Moreover, use of multiple sensors in CAP estimation is suggested to avoid pole cancelation phenomena due to RTF zeros [4]. In any case, experimental tests showed that the source position during calibration has no substantial influence on the pole determination, as expected from theory.

After CAP estimation, filter in (5) is applied to all received signals. Then, conventional GCC is applied to get TD estimates to be used for source localization.

4 EXPERIMENTAL RESULTS

In this section experimental results obtained with the proposed preprocessing are described. The CAP approach (*CAP-GCC*) was compared to the conventional GCC with ML weighting (*ML-GCC*) and performance were evaluated in terms of both TD and source position estimates.

4.1 Experimental setup

Data was computer-generated by the image method [8], applied to a rectangular room with plane walls, having size ($L_x = 5.45$, $L_y = 4.15$, $L_z = 2.80$) meters. A reflection coefficient β ($0 \leq \beta \leq 1$), independent of frequency and angle of arrival, was adjusted to give the desired *reverberation time* T_R [1], according to the Eyring formula [9]. A fixed omnidirectional point source and two square arrays of four microphones each were assumed. Tests were performed considering white Gaussian sources.

Received signals were sampled at $f_s = 10$kHz and pass-band filtered in the band [450Hz $\div$ 3475Hz]. 300 frames of 2048 samples were generated at several values of T_R [5]. Additive noise was not considered, since the Signal-to-Noise ratio is usually very high and the main source of error is due to signal-induced reflections and diffusion. The length of the filter $F(z)$ in (4) was set to 76, corresponding to assuming 75 dominant CAPs [4].

TD estimates obtained by the two methods were compared. Specifically, following [5], the number of *anomalies*, bias and variance of the delay estimate were adopted as performance indexes. A delay estimate is defined *anomaly* when it exceeds $T_c/2$, being T_c a measure of the correlation time of the signal (e.g. the amplitude of the main lobe of the autocorrelation function at $-3dB$) [5].

Following TD estimation, source positions in the two cases were estimated from TD estimates by a simple geometrical algorithm [10].

4.2 Performance analysis

4.2.1 TD estimation

Figure 1 shows the statistics of the TD estimates for every microphone pair of the first array, in the case of the GCC and CAP-GCC methods. A white Gaussian, zero-mean, unitary variance omnidirectional source was considered. For each pair, the percentage of anomalies and the robust trimmed estimates of bias and variance [11] after anomaly removal are shown as a function of the

[1]The reverberation time is defined as the time needed for the sound intensity to fall below 60dBs, after the source has been stopped.

reverberation time T_R. TDs are measured in samples; intra-sample precision is achieved by interpolation.

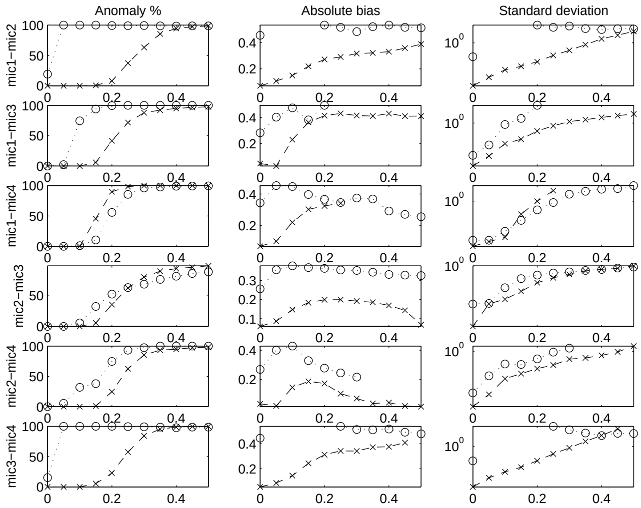

Figure 1: Percentage of anomalies, bias and standard deviation vs. T_R for the mic pairs of the first array. Circles: GCC, crosses: CAP-GCC.

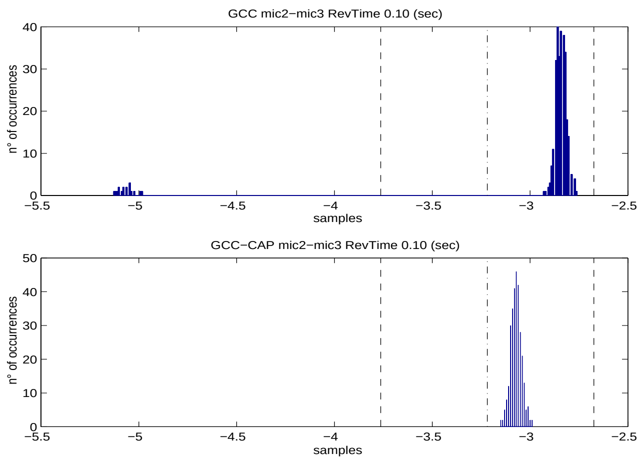

Figure 2: Histogram of TD estimates for $T_R = 0.10$ sec (mics $\mathbf{m_2}$ and $\mathbf{m_3}$ of fig. 1). Top: GCC, bottom: CAP-GCC.

Missing points in fig. 1 correspond to a percentage of anomalies higher than 95%.

Figures 2 and 3 show the histograms of the TD estimates for $T_R = 0.1$ sec and $T_R = 0.15$ sec respectively. The central vertical line indicates the true delay, while the two others are the anomaly thresholds. It is clear the performance degradation (in terms of bias and variance of the estimate) of both algorithms when the reverberation level increases. However, the CAP-GCC estimator is more robust, since the number of anomalies increase more slowly with T_R.

The last experiment illustrates the insensitivity of the CAP modeling to the source position chosen for calibration.

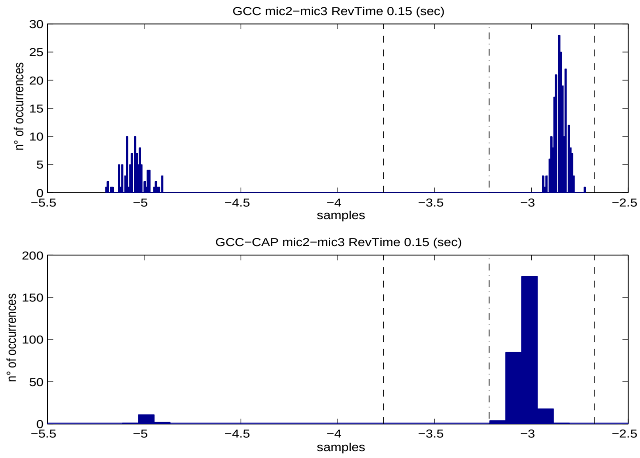

Figure 3: Histogram of TD estimates for $T_R = 0.15$ sec (mics $\mathbf{m_2}$ and $\mathbf{m_3}$ of fig. 1). Top: GCC, bottom: CAP-GCC.

Figure 4 shows the performance obtained for two different source positions [2]. Namely, during the calibration phase CAPs were estimated for source position S_1. The same filter was then used for the source position S_2. The performance in the two cases were practically identical.

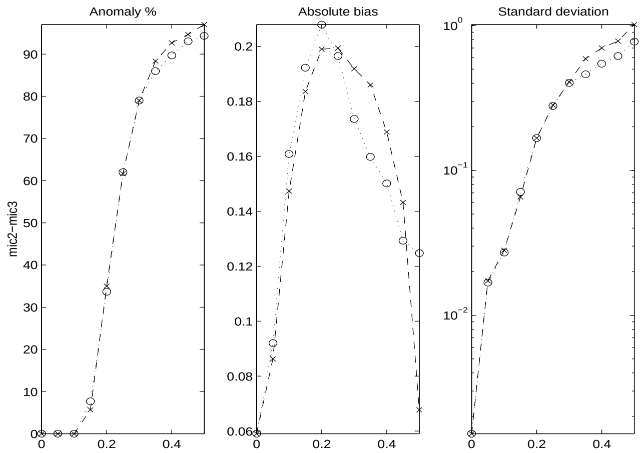

Figure 4: Performance of CAP preprocessing for different source positions. Crosses: source S_1, circles: source S_2.

It is important to remark that the CAP prefilter is estimated off-line during the room calibration phase, so that GCC is not influenced in terms of computational cost. In fact, while environment changes (source and furniture movements, entering or exiting of people) may easily mask some of the common acoustical modes, their frequency is expected to be rather unperturbed.

[2] The (x, y, z) coordinates of the two sources S_1 and S_2 were: S_1=[1.8 1.4 1.7] and S_2=[3.6 2.8 1.7], units are meters.

4.2.2 Localization

The actual relative delay between signals radiated by a single source and acquired by two microphones $(\mathbf{m}_{i1}, \mathbf{m}_{i2})$ is

$$D_i(\mathbf{m}_{i1}, \mathbf{m}_{i2}, \mathbf{s}) = \frac{|\mathbf{s} - \mathbf{m}_{i1}| - |\mathbf{s} - \mathbf{m}_{i2}|}{c} \qquad (6)$$

where $\mathbf{s}$ represents the (unknown) source position.

Given a set of estimated relative time delays $\{\hat{D}_i\}$, $\mathbf{s}$ can be estimated by minimization of a proper error functional E:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{arg\ min}[E(\hat{D}_i, \mathbf{s})] \qquad (7)$$

Several choices for E are available [10]. In this paper a simple geometrical solution based on the Euclidean norm was adopted. We remark that only non anomalous TD estimates were used for triangulation.

Figures 5 and 6 show the results obtained in terms of location estimates. The cluster of estimates around the true position obtained by the CAP method is clearly more dense, confirming the effectiveness of the proposed approach.

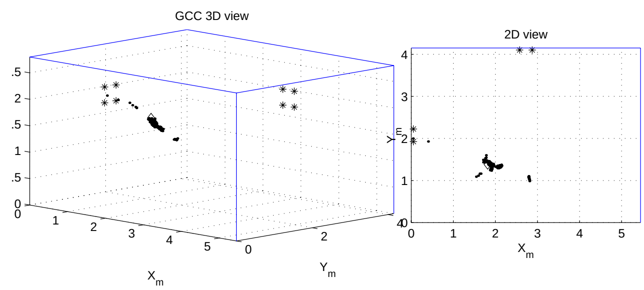

Figure 5: Position estimates: GCC.

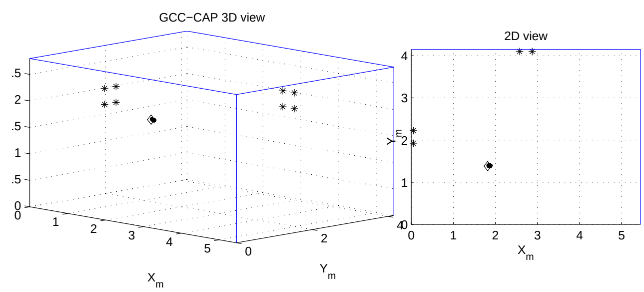

Figure 6: Position estimates: CAP-GCC.

5 CONCLUSION

A new preprocessing technique for time-delay estimation in reverberant rooms has been proposed. This novel approach is based on the common-acoustical-pole modeling of room transfer functions and is able to improve the performance of conventional generalized cross-correlation methods with a reduced computational cost in the room calibration phase. In addition, since CAP filtering is performed in the time domain, the system remains linear and the approach can be extended to multi-source environments [3]. This topic is currenlty under investigation.

References

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Techniques and Applications*, Springer-Verlag, 2001.

[2] C. H. Knapp and G. Clifford Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoust., Speech and Signal Processing*, vol. 24, no. 4, Aug. 1976.

[3] E.D. Di Claudio and R.Parisi, "Multi-source localization strategies," in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. Ward, Eds. Springer-Verlag, 2001.

[4] Y.Haneda, S.Makino, and Y.Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. on Speech and Audio Processing*, vol. 2, no. 2, pp. 320–328, April 1994.

[5] B. Champagne, S. Bedard, and A. Stephenne, "Performance of time-delay estimation in the presence of room reverberation," *IEEE Trans. on Speech and Audio Processing*, vol. 4, no. 2, pp. 148–152, Mar. 1996.

[6] Y.Haneda, S.Makino, and Y.Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. on Speech and Audio Processing*, vol. 5, no. 4, pp. 325–333, July 1997.

[7] S.L. Marple Jr., *Digital Spectral Analysis*, Prentice Hall Signal Processing Series, 1987.

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

[9] H. Kuttruff, *Room Acoustics*, Elsevier, London, 3rd edition, 1991.

[10] M.S Brandstein and H.F. Silverman, "A practical methodology for speech source localization with microphone arrays," *Computer, Speech, and Language*, vol. 11, no. 2, pp. 91–126, April 1997.

[11] P.J. Huber, *Robust Statistics*, John Wiley, New York, 1981.