SIMPLIFIED OPTIMAL LINE SELECTION FOR ACOUSTIC LOCALIZATION IN THE PRESENCE OF REVERBERATION

Albenzio Cirillo, Raffaele Parisi, Michele Scarpiniti and Aurelio Uncini

INFOCOM Department
"Sapienza" University of Rome
via Eudossiana 18, 00184 Rome
Email: albenzio.cirillo@uniroma1.it

ABSTRACT

The problem of acoustic source position estimation in reverberant environments has been tackled in different ways. The authors have recently proposed a valid solution which shows good results even in very hard conditions in terms of reverberation. Although the estimator is indicated for a real-time implementation as long as it can be solved in a closed form, in this paper the authors present a simplified version of the estimator in order to significantly reduce the complexity of the previous realization. A direct comparison among the two estimators will be shown, additionally the estimator can be considered also a robust time difference of arrival estimator in the presence of reverberation.

1. INTRODUCTION

In recent years research has been centered on Intelligent Systems focused on the realization of robust interfaces meant to be context aware and to take decisions according to heuristic rules. The position of a sound source is considered as an important piece of information for the context analysis and requires a robust system with a low rate of outliers position estimations. A large part of the microphone array community has been dealing with this topic and a plethora of solutions have already been proposed [3] [14] [16] [15] [11] [2].

The Linear Intersection (**LI**) [3] is a very efficient closed form estimator for acoustic source localization. LI belongs to a family of algorithms which are usually called indirect methods or Time Delay Estimation (**TDE**) methods [16]: this means that the location estimation is consequent to a preliminary TDE, hence it can be seriously affected by errors in the case of reverberant environments.

The most commonly adopted technique for TDE is the Generalized Cross-Correlation (**GCC**) [9]. The time lag corresponding to the maximum peak of the GCC is usually adopted as an estimate of the time-delay among two microphones. GCC can be affected mainly by noise and reverberation[1] [10], causing *anomalies*, which means that the main peak of the GCC might be not associated to the actual time-delay. These effects have been taken in consideration in [8] [7] [4]. In our work we are going to consider mainly the effect of reverberation with respect to noise, that is usually negligible. Even in cases of anomaly, it can be hypothesized that there is a secondary peak of the GCC which is the one associated to the true time-delay [16]. Unfortunately there is not a direct way to retrieve further information on which is the correct peak to select.

[1]The reverberation time (T_{60}) is the time required for reflections of a direct sound to decay by 60 dB below the level of the direct sound.

Previous works have dealt with the importance of considering secondary peaks in the GCC [5] [6] [13], obtaining a new method for acoustic source localization in reverberant environment [6] which has been called Optimal Line Selection (**OLS**). OLS is an estimator that naturally extends the LI estimator to the case of reverberant environments. Results in [6] show the benefit obtained by considering multiple peaks on the GCC in order to achieve a more accurate and robust localization in the presence of reverberation.

Depending on the number of microphones used and on the number of peaks considered in the GCC, OLS can be computationally expensive. It is indeed necessary, in OLS, to search exhaustively among all the possible combination of delays to obtain the source position estimate. In the simplified version the number of comparisons is significantly reduced by giving an initial guess on the source position.

The paper outlines in the first section background knowledge on GCC and OLS, in particular OLS complexity will be taken in examination. In the second part of the paper, the proposed estimator will be illustrated and numerical results, based on Monte-Carlo trials, will show the performance varying the reverberation time.

2. BACKGROUND

2.1 The GCC Method

Generalized Cross Correlation was introduced in [9] and basically estimates the cross-correlation among two ad-hoc prefiltered signals.

Given a source signal $s(t)$ in a reverberant environment, the signals acquired by a microphone pair are modeled as

$$\begin{array}{rcl} x_1(t) & = & s(t) * h_1(t) + n_1(t) \\ x_2(t) & = & s(t) * h_2(t) + n_2(t), \end{array} \tag{1}$$

where $h_i(t)$ $(i = 1,2)$ is the room impulse response between the source and the i-th microphone and $n_i(t)$ is uncorrelated noise. It is demonstrated in [7] and [12] that the best prefilter for signals affected by reverberation is the Phase Transform (**PHAT**).
Hence GCC-PHAT is evaluated as the following

$$R_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \Psi(f) G_{x_1x_2}(f) e^{j2\pi f\tau} df, \tag{2}$$

where $G_{x_1x_2}(f)$ is the cross power spectrum of $x_1(t)$ and $x_2(t)$ and $\Psi(f)$ is the PHAT prefiltering

$$\Psi(f) = \frac{1}{|G_{x_1x_2}(f)|}. \tag{3}$$

 ISSN 2076-1465

Ideally, for low noise and reverberation, the time lag where the main peak of the GCC is located represents the estimation of the delay between the direct path in the two impulse responses, also known as Time Difference Of Arrival (**TDOA**). Both [4] and [7] show how the percentage of anomalies in time-delay estimation rises up with increasing reverberation and noise level.

In the hypothesis of plane waves, it is possible to get the Direction of Arrival (**DOA**) from TDOA τ as an angle measured with respect to the direction of the line passing through the two sensors

$$\theta = \arccos\left(\frac{c\cdot\tau}{d}\right), \tag{4}$$

where c is the speed of sound and d is the distance between the microphone pair.

2.2 OLS Position Estimator

Using a particular microphone array displacement as two orthogonal couples (referred to as *quadruple*), the DOAs estimated on the two pairs can be used to describe the cosine directions of a bearing line whose origin is in the midpoint of the quadruple and whose direction aims at the source [3]. Then, using at least two quadruples, the source location might be found as the intersection of the bearing lines. Due to quantization errors and GCC-PHAT errors in TDE, lines will be skew. So the intersection problem is solved considering all the lines pairwise and computing with an over-constrained system [3] the points on them which are located in the proximity of the closest distance among the two lines. Hence, these points are called *at minimum distance*. Each point at minimum distance $\mathbf{s}_{ij}$ generated by the couple of lines i and j can be weighted as

$$w_{ij} = \sum_{p=1}^{2*Q} P(T(\mathbf{s}_{ij},\mathbf{m}_{1p},\mathbf{m}_{2p}),\tau_p,\sigma^2), \tag{5}$$

where p is the pair index and Q is the number of quadruples, $P(x,m,\sigma^2)$ is a normal distribution with the time-delay estimate on the p-th couple as the mean value and evaluated for

$$T(\mathbf{s}_{ij},\mathbf{m}_{1p},\mathbf{m}_{2p}) = \frac{\|\mathbf{m}_{1p}-\mathbf{s}_{ij}\| - \|\mathbf{m}_{2p}-\mathbf{s}_{ij}\|}{c}, \tag{6}$$

which is the geometrically evaluated delay among the point s_{ij} and the microphones of the p-th pair $\mathbf{m}_{1p}$ and $\mathbf{m}_{2p}$. The final estimation is obtained as a weighted sum of all the M points

$$\hat{s} = \frac{\sum_{i,j=1,i\neq j}^{M} w_{ij}s_{ij}}{\sum_{i,j=1,i\neq j}^{M} w_{ij}}. \tag{7}$$

Variance σ^2 is dependent on GCC variance, which is a hard parameter to estimate as it depends on the signal nature, the noise level and the reverberant conditions [7]. For this reason the variable σ can be set empirically.

In case of anomalies the lines will no longer aim at the source, hence the localization will be compromised.

OLS [6] takes into consideration the first k peaks of the GCC-PHAT and finds all the k^2 combinations of peaks of the two orthogonal pairs such to bear up to k^2 lines with one quadruple. Nevertheless there is a constraint on the feasibility of each time-delay combination and it is

$$\cos(\theta_{12})^2 + \cos(\theta_{34})^2 \leq 1, \tag{8}$$

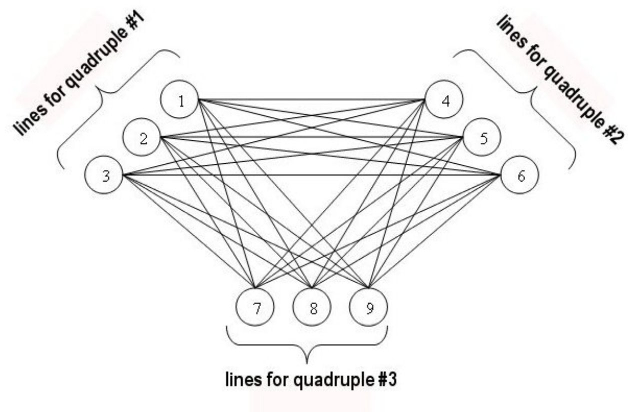

Figure 1: OLS graph of all possible combinations of lines in case of 3 bearing lines for each quadruple.

where θ_{12} and θ_{34} (see formula (4)) are the DOAs with respect to the two orthogonal pairs.

Considering one line for each quadruple, it is then possible to consider up to k^{2Q} set of lines (see Figure 1), each one with its own set of points at minimum distance. A measure of how much is expanded the region of space containing the points at minimum distance for each set is obtained with

$$C_L = \frac{1}{M_L}\sum_{(i,j)\in L}(\|s_{ij}-b_L\|)^2, \tag{9}$$

where L is the set index and b_L is the midpoint of the set of points at minimum distance.

OLS elects as candidate for the position estimation the L-th set with the lowest C_L value. The final estimation is then refined as in (7).

3. A SIMPLIFIED VERSION OF OLS

As formerly specified, OLS might be computationally expensive as long as it requires an exhaustive search among all the feasible set of lines. Basically it requires the following operations

for all k^{2Q} *set of lines* **do**
solve $\binom{Q}{2}$ *points at minimum distance systems*;
evaluate formula (9);
end for

However it is true that if it could be possible to know with certainty one of the points of the best set, then it could be uniquely guessed the set of points. It is indeed true that as long as the GCC is a finite sequence and that the peaks belong to the range $-\tau_{MAX} \leq \frac{l}{T_c} \leq \tau_{MAX}$, with l an integer number, T_c the sampling time and

$$\tau_{MAX} = \frac{d}{c}, \tag{10}$$

then it is possible to find all the possible combinations of time-delays subject to (8). Hence it is possible to bear all the feasible lines and points at minimum distance and finally to know a priori all the best set of points, that are the ones minimizing (9).

Unfortunately it is not possible to know which point belongs to the best set. Moreover, not always it could be possible to verify a perfect match among the delay estimated on the quadruples and the ones corresponding to the best sets of points, because it should be considered a tolerance due to the noisy GCC estimation on a small percentage of the quadruple used in the OLS estimator.

After this considerations it is possible to assess that an initial guess on the source position could be used to elect a set of lines, hence time-delays, and produce a final estimation according to (7).

Hereby, based on the previous assumption, a simplified version of OLS is developed and proposed.

The first step consists of picking up just a couple of quadruples and considering the first k peaks on the respective 4 pairs. If for each couple of lines, one for each quadruple, it is considered the midpoint among the two points at minimum distance, it is possible to set them as k^4 initial guesses. For each initial guess is then possible to measure the geometrical delay with respect to the $2(Q-2)$ remaining couples according to (6).

Then, for each group of quadruples q not considered in the first step, k main peaks should be selected on the GCC related to the orthogonal microphone pairs, such to combine them into k^2 sets of time-delay couples $(\tau_{q12}^{(l_q)}, \tau_{q34}^{(l_q)})$, with $l_q = 1, \ldots, k^2$.

Finally an error measure can be introduced with respect to the initial guess $\mathbf{s}_{Ls}$

$$
\begin{aligned}
e(s_{Ls}, l_3, \ldots, l_Q) \quad &= \quad \sum_{q=3}^{Q} [\left(T(\mathbf{s}_{Ls}, \mathbf{m}_{1q}, \mathbf{m}_{2q}) - \tau_{q12}^{(l_q)} \right)^2 \\
&+ \quad \left(T(\mathbf{s}_{Ls}, \mathbf{m}_{3q}, \mathbf{m}_{4q}) - \tau_{q34}^{(l_q)} \right)^2], \quad (11)
\end{aligned}
$$

$$
s.t. \ \cos(\frac{c\tau_{q12}^{(l_q)}}{d})^2 + \cos(\frac{c\tau_{q34}^{(l_q)}}{d})^2 \leq 1.
$$

The procedure can be then described as

for all k^4 *initial guesses* **do**

$perform\,(Q-2)k^2\,operations\,to\,evaluate\,formula\,(11);$

end for

The best initial guess is the one which minimizes equation (11) with respect to all the s_{Ls} and all the lq values.

Once the best initial guess has been chosen, the position estimation can be refined according to (7) by using the delay couples $(\tau_{q12}^{(l_q)}, \tau_{q34}^{(l_q)})$ associated to the minimization of (11). It is clear that this procedure has a lower computational load than the one in OLS. To give an idea of what has been claimed, a short example will show the saving in case of realistic parameters.

Example. The computation is lower if at least

$$
(Q-2)k^2 * k^4 \leq k^{2Q}
$$

indeed, the same inequality can be written as

$$
(Q-2)k^{6-2Q} \leq 1,
$$

which is always true for $Q \geq 3$ assuming $k = 3$.

It is also possible to repeat the estimation for all the Q quadruples in order to average the Q final estimations, but the method is advantageous with respect to OLS only for $Q \geq 4$.

Moreover, not all the quadruples could be useful. This means that a quadruple could have GCCs with too many anomalies and generate a too large element of the sum in (11) with respect to a fixed threshold. In this case that quadruple should be discarded from the mentioned sum and formula (11) should be normalized on the number of useful quadruples for a fair minimization.

3.1 Numerical Results

A synthetic environment has been simulated according to the Image Method [1]. Room dimensions are $10 \times 6.6 \times 3[m]$ and the reverberation time has been varied from anechoic up to $2s$. A number of $Q = 4$ quadruples has been placed on 4 different walls, recording a single source emitting white noise to which other uncorrelated noise has been added such to have $SNR = 20dB$. We have used a sampling frequency of $f_S = 48KHz$ and a frame length of 4096 samples.

Localization Root Mean Square Error (RMSE) has been evaluated both for OLS [6] (Figure 3) and Simplified OLS (**S-OLS**) (Figure 2) over 100 independent Monte-Carlo trials, showing a similar performance in both cases. The probability of anomalies [7] P_A has been evaluated by averaging the anomalies on the 4 couples belonging to the quadruples used for the initial guess (Figure 4). P_A for S-OLS is evaluated only on the useful frames, that are the ones that have not been discarded because of a minimum value of (11) higher than a fixed threshold. P_A is larger for the simple GCC-PHAT with respect to S-OLS. In any case, nevertheless, when k is increased S-OLS shows a slight increasing of P_A for the same T_{60}. Although this behavior, in Figure 5 it is claimed that, increasing k, the probability of missed localization gets lower increasing T_{60}.

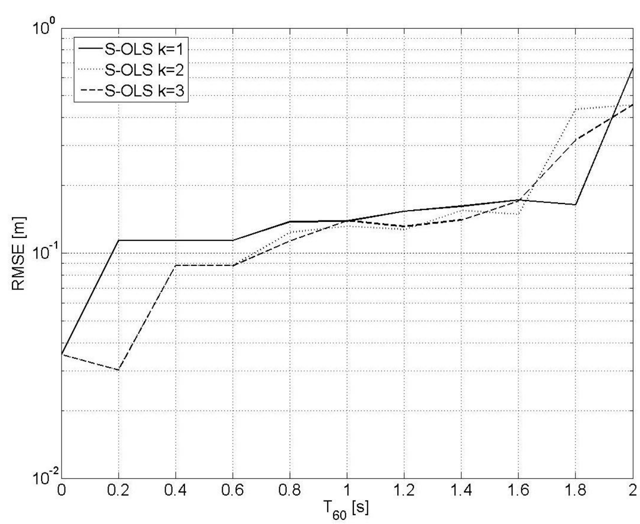

Figure 2: Position estimation RMSE versus Reverberation Time (T_{60}) with SNR=20dB, comparing the simplified OLS (S-OLS) with a different number of k peaks.

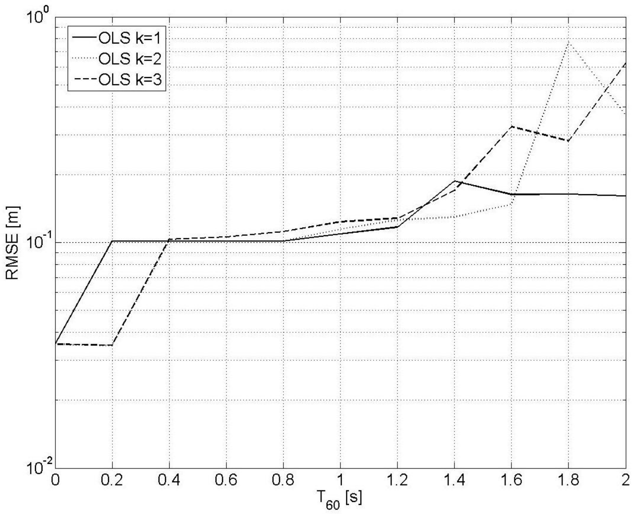

Figure 3: Position estimation RMSE versus Reverberation Time (T_{60}) with SNR=20dB, comparing OLS with a different number of k peaks.

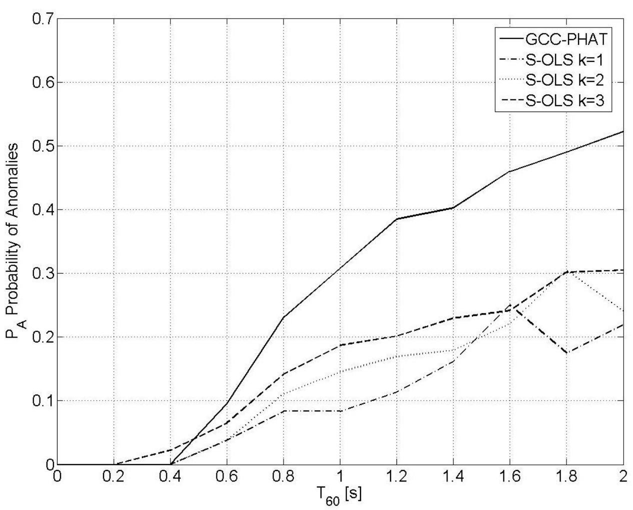

Figure 4: Probability of Anomalies versus Reverberation Time (T_{60}) with SNR=20dB, comparing the simple GCC-PHAT with the simplified OLS (S-OLS) with a different number of k peaks with SNR=20dB.

4. CONCLUSION

A new simplified and more efficient version of the OLS algorithm, named simplified OLS (S-OLS), has been introduced. The proposed algorithm is able to estimate acoustic source position and shows successful results in terms of computational cost and localization performance. Some numerical results show the effectiveness of the proposed approach and demonstrate that it can solve the problem with a minor amount of computational load. The accuracy of the localization has been evaluated in terms of RMSE which is widely used in literature.

Future works on this topic will deal with directive sources and microphones displacement.

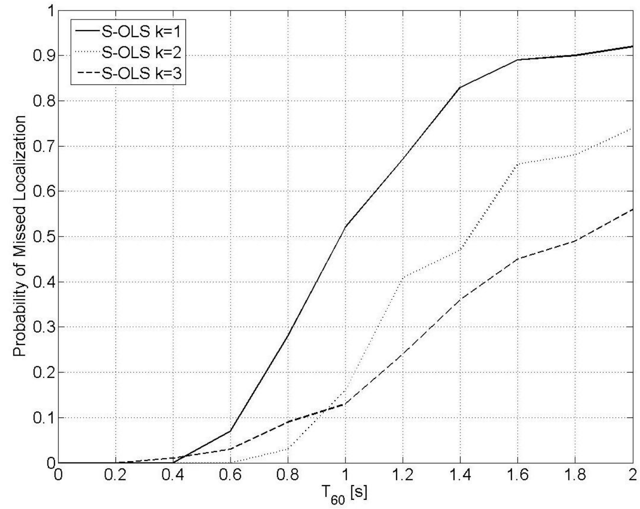

Figure 5: Probability of Missed Localization versus Reverberation Time (T_{60}) with SNR=20dB, comparing the simplified OLS (S-OLS) with a different number of k peaks.

Acknowledgment

This work has been partially supported by the Italian National Project: Wireless multiplatfOrm mimo active access netwoRks for QoS-demanding muLtimedia Delivery (WORLD), under grant number 2007R989S.

REFERENCES

[1] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. vol. 65, pp. pp. 943–950, 1979.

[2] M. Brandstein and D. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*, 1st ed. Springer, June 2001.

[3] M. Brandstein, J. Adcock, and H. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, Jan. 1997.

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

[5] A. Cirillo, R. Parisi, and A. Uncini, "Sound mapping in reverberant rooms by a robust direct method," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, March 31 2008–April 4 2008, pp. 285–288.

[6] ——, "A new consistency measure for localization of sound sources in the presence of reverberation," in *Proc. of the 16th International Conference on Digital Signal Processing*, 5–7 July 2009, pp. 1–5.

[7] T. Gustafsson, B. Rao, and M. Trivedi, "Source localization in reverberant environments: modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov. 2003.

[8] J. Ianniello, "Time delay estimation via cross-

correlation in the presence of large estimation errors," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 6, pp. 998–1003, Dec. 1982.

[9] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[10] H. Kuttruff, *Room Acoustics*, 4th ed. Taylor & Francis, 2000.

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

[12] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 3, pp. 288–292, May 1997.

[13] R. Parisi, A. Cirillo, M. Panella, and A. Uncini, "Source localization in reverberant environments by consistent peak selection," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, Honolulu, Hawaii, 2007.

[14] R. Parisi, P. Croene, and A. Uncini, "Particle swarm localization of acoustic sources in the presence of reverberation," in *Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS 2006)*, 2006, pp. 4pp.–4742.

[15] H. Schau and A. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 8, pp. 1223–1225, Aug. 1987.

[16] D. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, Nov. 2003.