EFFICIENT SOURCE LOCALIZATION AND TRACKING IN REVERBERANT ENVIRONMENTS USING MICROPHONE ARRAYS

Fabio Antonacci, Davide Lonoce, Marco Motta, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica e Informazione – Politecnico di Milano
Piazza Leonardo da Vinci 32, Milano, Italy
antonacci/sarti/tubaro@elet.polimi.it

ABSTRACT

In this paper we propose an algorithm for acoustic source localization and tracking that is suitable for reverberant environments. The approach that we propose is based on the iterative identification of the FIR channels that link source and microphones through an LMS method (Multi-Channel LMS) [2], but we propose additional solutions that significantly improve this method in terms of computational efficiency and localization reliability, without affecting its convergence properties. This is achieved through a modified block-wise implementation of the approach combined with Kalman filtering. We also show the results of extensive comparative testing using novel performance parameters for the assessment of localization reliability.

1. INTRODUCTION

Source localization algorithms based on microphone arrays can be roughly divided into two broad classes: those that are based on the identification of the source-mirophone channels, and those that are not. The former class of solutions usually relies on a channel ideality assumption. In this case the Direction Of Arrival (DOA) can be estimated from the signal delays between adjacent microphones, which can be derived by maximizing the cross-correlation between corresponding signals. Examples of this sort are *Time Delay Estimation*, *GCC-PHAT* and *SRP-PHAT* (for an overview see [1]). Other examples of methods that do not consider channel identification are *Beamforming*, *Capon Beamforming* and *MUSIC*), which are tailored on narrow-band signals (which is not usually the case of audio signals) and estimate the source location from the phase displacement between adjacent microphones. This choice turns the problem into a that of the estimation of spatial frequencies (see [3]). The performance of all such methods, however, tend to drop significantly in the presence of reverberation, as the channel ideality assumption is no longer satisfied. Methods based on channel identification are aimed at overcoming this limitation. A popular solution of this sort is the *Multi-Channel Least Mean Square* method *MCLMS*, which is based on a preliminary iterative blind identification of the transfer functions between source and microphones. The MCLMS method then determines the direct echo signal in the estimated impulse response and uses this information to estimate the DOA. Compared with the methods that do not perform channel identification, the MCLMS algorithm shows a significant resilience against reverberations, but it is quite computationally expensive and it produces results that are not as stable as we would hope.

In this paper we propose a novel method to overcome the above limitations of the MCLMS method. In fact, we achieved a significant reduction of the computational cost by working in a block-wise fashion, and for this reason we refer to our approach as the *Fast MCLMS* algorithm. This result is achieved without affecting the performance or the convergence properties of the method. In order to achieve better localization stability over time, we introduced a source tracking algorithm based on Kalman filtering, with significant improvement in the quality of the localization of moving sources.

In order to assess the performance of our solution, we conducted an extensive set of high-quality simulations with various levels of environment reverberation. The data sequences were artificially generated with a fast beam tracing algorithm of proven effectiveness and accuracy (it accounts for reflections as well as diffraction) [4]. The performance evaluation was conducted on our methos as well as various other well-known solutions such as MCLMS, GCC and MUSIC.

2. CHANNEL MODEL AND MCLMS METHOD

Our channel model is a SIMO system with one input signal (audio source) $s(n)$ and M output signals $x_i(n)$ acquired by the microphones (see Fig. 1). Each output signal is the convolution of the input signal with a different

This work was developed within the FIRB-VICOM project (www.vicom-project.it) funded by the Italian Ministry of University and Scientific Research (MIUR); and within the VISNET project, a European Network of Excellence (www.visnet-noe.org)

impulse response h. Neglecting, at this stage, the additive noise, we have $x_i(n) = s * h_i(n)$. Adopting a vector

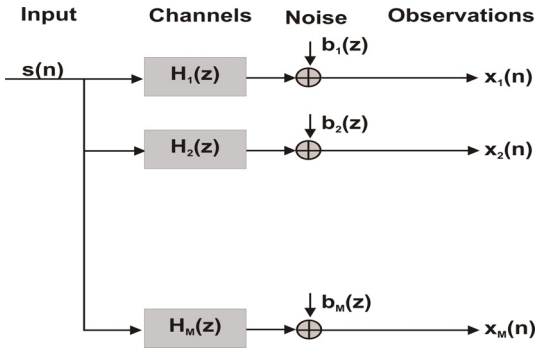

Fig. 1. SIMO channel model used in MCLMS method derivation.

form for the source-microphone impulse responses $\mathbf{h}_i = [h_i(0)\ h_i(1)\ ...\ h_i(L-1)]^T$, and for the *i*-th observation

$$\mathbf{x}_i(n) = [x_i(n)\ \ x_i(n-1)\ ...\ x_i(n-L+1)]^T\ , \qquad (1)$$

we have $x_i * h_j = s * h_i * h_j = x_j * h_i$, which gives

$$\mathbf{x}_i^T(n)\mathbf{h}_j = \mathbf{x}_j^T(n)\mathbf{h}_i\ . \qquad (2)$$

By grouping all the impulse responses in a single column vector $\mathbf{h} = [\mathbf{h}_1^T\ \mathbf{h}_2^T\ ...\ \mathbf{h}_M^T]^T$ and using eq. (2), Huang and Benesty [2] obtained:

$$\mathbf{R}\,\mathbf{h} = \mathbf{0}\ , \qquad (3)$$

where

$$\mathbf{R} = \begin{bmatrix} \sum_{i\neq 1} R_{x_i x_i} & -R_{x_2 x_1} & ... & -R_{x_M x_1} \\ -R_{x_1 x_2} & \sum_{i\neq 2} R_{x_i x_i} & ... & -R_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -R_{x_1 x_M} & -R_{x_2 x_M} & ... & \sum_{i\neq M} R_{x_i x_i} \end{bmatrix} \qquad (4)$$

and $R_{x_i x_j} = E[\mathbf{x}_i(n)\,\mathbf{x}_j^T(n)]$. Eq. (4) shows that $\mathbf{h}$ can be obtained in a blind fashion (without knowledge of source statistics), using only second-order statistics of outputs. Eq. (3) implies that $\mathbf{h}$ resides in the null-space of $\mathbf{R}$ which can be used for deriving an LMS algorithm.

Let us consider, without loss of generality, a two microphone system. The cross-correlation error between channels can be defined as

$$e(n) = \mathbf{x}_1^T(n)\mathbf{h}_2 - \mathbf{x}_2^T(n)\mathbf{h}_1\ . \qquad (5)$$

By collecting the observations in a vector $\mathbf{x} = [\mathbf{x}_2^T\ \mathbf{x}_1^T]$ and the impulse responses in $\mathbf{h} = [\mathbf{h}_1^T\ -\mathbf{h}_2^T]^T$, eq. (5) can be written as

$$e(n) = \mathbf{h}^T(n)\,\mathbf{x}(n)\ . \qquad (6)$$

In order to devise an LMS algorithm that converges to the correct $\mathbf{h}$ vector, Huang identified the distance function

$$J(n) = \frac{e^2(n)}{||\,\mathbf{h}(n)\,||^2}\ . \qquad (7)$$

which determines the optimal LMS solution [6] as $\widehat{\mathbf{h}} = \mathrm{argmin}_{\mathbf{h}}\{E[J(n)]\}$. Using an iterative algorithm and updating the **h**-estimate in the opposite direction to $\nabla J(n)$, we obtain

$$\widehat{\mathbf{h}}(n+1) = \widehat{\mathbf{h}}(n) - \frac{2\mu}{||\,\widehat{\mathbf{h}}(n)\,||^2}[\widetilde{\mathbf{R}}\,\widehat{\mathbf{h}}(n) - J(n)\,\widehat{\mathbf{h}}(n)]\ , \qquad (8)$$

where

$$\widetilde{\mathbf{R}} = \begin{bmatrix} \mathbf{x}_1(n)\,\mathbf{x}_1^T(n) & \mathbf{x}_1(n)\,\mathbf{x}_2^T(n) \\ \mathbf{x}_2(n)\,\mathbf{x}_1^T(n) & \mathbf{x}_2(n)\,\mathbf{x}_2^T(n) \end{bmatrix} \qquad (9)$$

and $\widehat{\mathbf{h}}(n)$ is the $\mathbf{h}$ estimation at the *n*-th step of the LMS algorithm. The convergence of the LMS algorithm is guaranteed if $0 < \mu < \frac{1}{\lambda_{\max}}$, where $\lambda_{\max}$ is the maximum eigenvalue of $\widetilde{\mathbf{R}} - J(n)\mathbf{I}$. Once a correct **h**-estimate is obtained, we can find the two direct-echo delays from $\mathbf{h}_1$ and $\mathbf{h}_2$. Using the related delays we can determine the DOA. The final source location can be triangulated from two or more DOA estimators of this sort. The LMS algorithm can be rewritten as

$$\begin{aligned} \widehat{e}(n) &= \widehat{\mathbf{h}}^T(n)\,\mathbf{x}(n) \\ \widehat{\mathbf{h}}(n+1) &= \frac{\widehat{\mathbf{h}}(n) - 2\mu\widehat{e}(n)[\mathbf{x}(n) - \widehat{e}(n)\mathbf{x}(\mathbf{n})]}{||\,\widehat{\mathbf{h}}(n) - 2\mu\widehat{e}(n)[\mathbf{x}(n) - \widehat{e}(n)\mathbf{x}(\mathbf{n})]\,||} \\ n &= n+1 \end{aligned} \qquad (10)$$

If L is the length of the impulse response and F_s the sampling frequency, then the computational cost (expressed in flops/sec) of MCLMS algorithm is $O = 8L\,F_s$. This cost is usually too high for real-time applications. In the next Section we will show how to reduce it by updating the **h**-estimate with blocks of data without affecting (under a stationariety assumption) the convergence properties. For reasons of simplicity, we will consider the case of a pair of microphones, although the algorithm can be easily developed for the more general case of M sensors.

Another drawback of MCLMS method (and, in general, of all localization algorithms) is the likelihood of faulty localizations. in order to regularize the trajectory of a moving source, we thus implemented a Kalman filtering algorithm.

3. FAST-MCLMS WITH KALMAN FILTERING

Let $x_i[n, n+L+1]$ be the observations acquired between time n and time $n+L+1$. Using the same notation for the input signal s, we can write

$$x_i[n, n+L+1] = h * s[n, n+2L+1]\ . \qquad (11)$$

According to eq. (10), n is updated at every step, thus producing a new observation at every step of the algorithm. We here modify the MCLMS algorithm in such a way to use a block of f contiguous data at each step of the algorithm. This way eq.(11) becomes

$$x_i[n+f, n+f+L+1] = h * s_i[n+f, n+f+2L+1] \,, \quad (12)$$

and the new LMS algorithm becomes:

$$\begin{aligned} \widehat{e}(n) &= \widehat{\mathbf{h}}^T(n)\,\mathbf{x}(n) \\ \widehat{\mathbf{h}}(n+1) &= \frac{\widehat{\mathbf{h}}(n) - 2\mu\widehat{e}(n)[\mathbf{x}(n) - \widehat{e}(n)\mathbf{x}(\mathbf{n})]}{||\,\widehat{\mathbf{h}}(n) - 2\mu\widehat{e}(n)[\mathbf{x}(n) - \widehat{e}(n)\mathbf{x}(\mathbf{n})]\,||} \\ n &= n+f \end{aligned} \quad (13)$$

Notice that using f new observations at each step leads to errors and channel estimations that differ from those of the MCLMS algorithm.

We will now prove that, under the assumption of a stationary source, the convergence properties of the MCLMS and the Fast-MCLMS algorithms coincide. Let

$$\begin{aligned} \mathbf{A} &= E[\widetilde{\mathbf{R}}(n) - J(n)\,\mathbf{I}] \\ \mathbf{A}_F &= E[\widetilde{\mathbf{R}}(nf) - J(nf)\,\mathbf{I}] \end{aligned}$$

be the average iteration matrices of the MCLMS and the FMCLMS algorithms, respectively. Assuming that x_1 and x_2 are stationary processes, we have

$$\begin{aligned} E[\widetilde{\mathbf{R}}(nf)] &= \begin{bmatrix} E[\mathbf{x}_1(nf)\mathbf{x}_1^T(nf)] & E[\mathbf{x}_1(nf)\mathbf{x}_2^T(nf)] \\ E[\mathbf{x}_2(nf)\mathbf{x}_1^T(nf)] & E[\mathbf{x}_2(nf)\mathbf{x}_2^T(nf)] \end{bmatrix} = \\ &= \begin{bmatrix} E[\mathbf{x}_1(n)\mathbf{x}_1^T(n)] & E[\mathbf{x}_1(n)\mathbf{x}_2^T(n)] \\ E[\mathbf{x}_2(n)\mathbf{x}_1^T(n)] & E[\mathbf{x}_2(n)\mathbf{x}_2^T(n)] \end{bmatrix} = \\ &= E[\widetilde{\mathbf{R}}(n)] \,, \end{aligned}$$

and

$$\begin{aligned} E[J(nf)] &= E[(\mathbf{x}_1^T(nf)\,\mathbf{h}_2 - \mathbf{x}_2^T(nf)\,\mathbf{h}_1)] = \\ &= E[(\mathbf{x}_1^T(n)\,\mathbf{h}_2 - \mathbf{x}_2^T(n)\,\mathbf{h}_1)] \,, \end{aligned}$$

therefore we can conclude that, if x_1 and x_2 are stationary processes, then $\mathbf{A}_F = \mathbf{A}$.

The case of M microphones can be developed exactly as shown above. The only differences to be noted are in the expressions of eq. (13), which needs to be modified according to eq. (8). Furthermore, the distance function (7), and therefore its gradient, takes on a more complex expression. All this will result in higher computational costs.

In order to make the localization more stable over time we introduce a source tracking process based on Kalman filtering [5]. Let $\mathbf{y}(k) = [x(k)\; y(k)]^T$ be a vector containing the source position estimated by a localization algorithm. The internal model is described by a state vector containing position and source velocity $\mathbf{x}(k) = [x(k)\; \dot{x}(k)\; y(k)\; \dot{y}(k)]^T$. Observation and state vectors are linked by

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{n}(k) \,, \quad (14)$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \,,$$

while $\mathbf{n}(k)$ is a a additive white gaussian noise with covariance matrix $\mathbf{R}$. The state-update equation is

$$\mathbf{x}(k+1) = \mathbf{A}\,\mathbf{x}(k) + \mathbf{v}(k) \,,$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \,,$$

T being the temporal distance between successive observations, and $\mathbf{v}(k)$ being a state noise with covariance matrix $\mathbf{Q}$. Let $\widehat{\mathbf{x}}^-(k)$ and $\widehat{\mathbf{x}}(k)$) be the a-priori and the a-posteriori estimations, respectivley, of the state vectors, which are obtained from the observations until $k-1$ and k. The a-priori and the a-posteriori errors are defined as $\mathbf{e}^-(k) = \mathbf{x}(k) - \widehat{\mathbf{x}}^-(k)$, and $\mathbf{e}(k) = \mathbf{x}(k) - \widehat{\mathbf{x}}(k)$, which are characterized by the following covariance matrices

$$\begin{aligned} \mathbf{P}^-(k) &= E[\mathbf{e}^-(k)\,\mathbf{e}^{-T}(k)] \,, \\ \mathbf{P}(k) &= E[\mathbf{e}(k)\,\mathbf{e}^T(k)] \,, \end{aligned}$$

respectively. The a-posteriori and a-priori estimations will be linked by

$$\widehat{\mathbf{x}}(k) = \widehat{\mathbf{x}}^-(k) + \mathbf{K}(\mathbf{y}(k) - \mathbf{H}\widehat{\mathbf{x}}^-(k)) \,,$$

where $\mathbf{y}(k) - \mathbf{H}\widehat{\mathbf{x}}^-(k)$ is the innovation and

$$\mathbf{K}(k) = \mathbf{P}^-(k)\mathbf{H}^T(\mathbf{H}\mathbf{P}^-(k)\mathbf{H}^T + \mathbf{R})^{-1}$$

is the filtering gain (see Fig. 2).

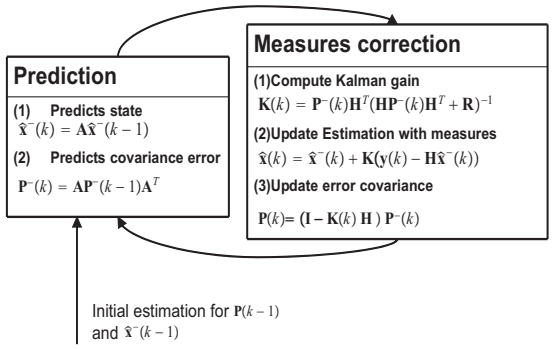

Fig. 2. Kalman algorithm.

4. RESULTS AND CONCLUSIONS

We compared the performance of FMCLMS with MCLMS and other non identificative-methods (GCC and MUSIC). In order to do so we simulated the acquisitions of an array of 8 microphones in a rectangular room using an advanced and accurate simulation method based on beam tracing [4]. This allowed us to track a moving speaker as we varied the reverberation level by changing the reflection coefficients.

Fig.3 shows which fraction of all localizations turned out to be correct (error below 2 degrees), in the various locations of the room. In this experiment the reflection coefficients was set to 0.9 (strongly reverberating room). As we can see, the Fast MCLMS method (with $f = 16$ and $f = 32$) localizes the audio source accurately in a wider area with respect to GCC and MUSIC. Fig. 4 shows the average

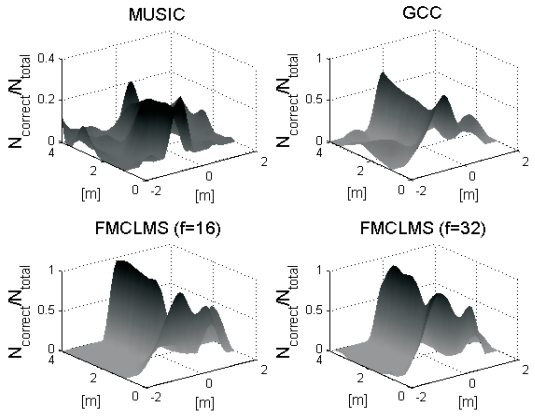

Fig. 3. Fraction of correct localizations for various algorithms in the test room.

localization error of several algorithms as reflection coefficient goes from 0.1 to 0.9. As we can see, the new algorithm performs comparably to MCLMS and better than non-identificative localization methods. Fig.5 shows the performance improvement due to Kalman filtering. A speaker following a circular trajectory of 2 m of radius around the microphone array is localized and tracked by the system. The range of DOA considered for the tracking is $-45°, \ldots, +45°$. As we can see, the impact of the Kalman filtering to localization stability is quite relevant, as it allows us to reduce the update rate of the filter a great deal without introducing tracking inconsistencies. The factor f, in fact gives us an approximate idea of the complexity reduction factor that we introduce. As a general rule, values of f between 16 and 32 turn out to be a good choice for most situations of interest, as they provide a good compromise between estimation consistency and computational efficiency.

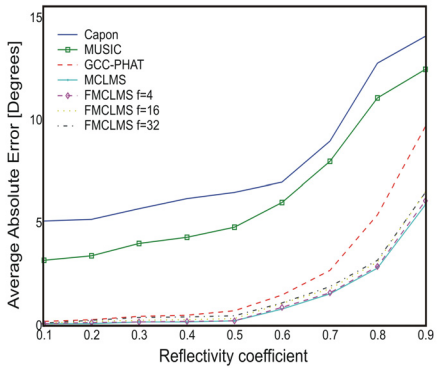

Fig. 4. A comparison of average localization error obtained with various methods.

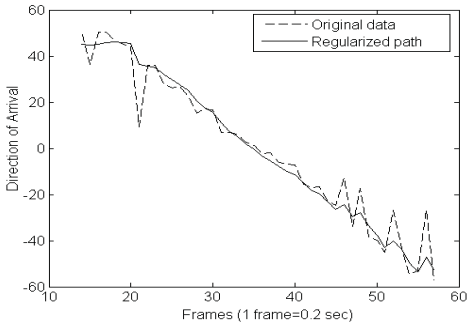

Fig. 5. Performance improvement obtained with adoption of Kalman filtering

5. REFERENCES

[1] M. Brandstein, D. Ward, *Microphone arrays*. Springer-Verlag, 2001.

[2] Y. Huang, J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification", *Signal Processing*, Vol. 82, pp. 1127-1138, Aug. 2002.

[3] P. Stoica, R.L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997

[4] F. Antonacci, M. Foco, A. Sarti, S. Tubaro, "Fast modeling of acoustic reflections and diffraction in complex environments using visibility diagrams ", *11th European Signal Proc. Conf. (EUSIPCO-2004)*, Vienna, Austria, Sept. 6-10, 2004, pp. 1773-1776.

[5] P.S. Maybeck, "The Kalman Filter-An introduction for potential users", *TM-72-3, Air Force Flight Dynamics Laboratory*, Wright Patterson AFB, Ohio, June 1972

[6] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1996.