Convolutive Acoustic Mixtures Approximation to an Instantaneous Model Using a Stereo Boundary Microphone Configuration

Juan Manuel Sanchis, Francisco Castells, and José Joaquín Rieta

Universidad Politécnica de Valencia
46730 Gandia Spain
{jmsanch,jjrieta,fcastells}@eln.upv.es

Abstract. In this work it is demonstrated that, taking into account the conditions of a convolutive mixture of acoustic signals, it is possible to configure a mixture system whose separation model accomplishes the conditions of an a instantaneous mixture. This system is achievable by using stereo boundary microphones; this type of coincident microphones can be used to uniform delays of the propagation channels and to reduce the number of reflections that characterize the impulse response of the system. By means of coincident boundary microphones techniques, instantaneous BSS algorithms are applicable, thus providing optimal results with less computational cost. This system is validated in both anechoic and reverberant chambers.

1 Introduction

The main problem that presents the evaluation of the algorithms of blind separation of sources (BSS) lays on the non availability of the original signal sources and the mixture system that originate the observed available signals. This implicit problem that arises when dealing with real mixtures can be solved using synthetic signals, where the original sources are known [1], [2]. In fact, the validation of BSS algorithms with synthetic recordings may represent an initial start point towards its further application to real mixtures. Certainly, by analyzing the separation quality in the case of synthetic mixtures, the main characteristics of the algorithm can be evaluated in some basic situations, thus permitting the study of its feasibility to achieve the main objective of separating the independent sources as accurately as possible. Hence, different scenarios can be created depending on the statistical properties of the sources and on the mixture process, and the effect that these factors may introduce in the separation performance of the algorithm can be evaluated [3].

Some works have focused their study on the influence of the mixture system on the results obtained by BSS algorithms. In this sense, the effect of the acoustic environment has been analyzed [4], since the impulse response that models the mixture system is directly related to the reverberation and the density of reflections that characterize the acoustic chamber. Those studies put into evidence that chambers with long impulse responses and high reflections densities imply a decrease in the separation performance.

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 816–823, 2004.

However, there are very few studies regarding the characteristics and configuration of the transducers employed for the reception of the observations that serve as inputs for the BSS algorithm [5], [6]. At the moment the available bibliography always establishes a placement of the microphone array following the technique of separated microphones (Fig.1a). The microphones are placed according to certain criteria, normally at equispaced points [5].

In opposition to this technique, we propose the use of the coincident microphones technique, jointly with a disposition that minimizes the number of captured reflections. Using boundary stereophonic microphones will be shown that the quality of the estimated sources is improved.

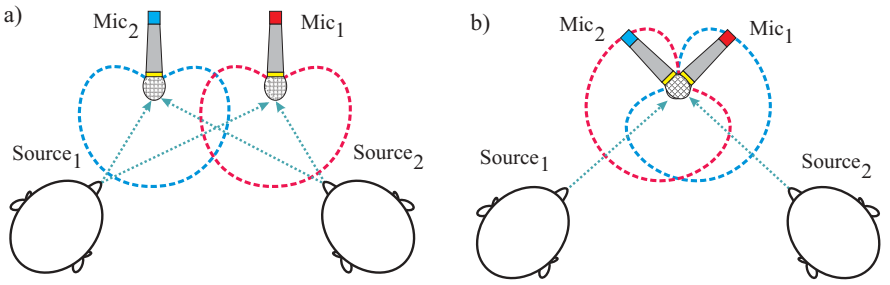

Fig. 1. Configuration for the reception of two sound sources using (a) separate microphones in front of (b) coincident microphones.

2 Approximation

In the configuration of coincident microphones (Fig. 1b), two directive microphones are perpendicularly placed such that both transducers are located at the same point. Following this configuration, the acoustic signal will ideally reach both microphones at the same time instant. Due to the directional characteristic of the microphones, the sources will be captured with different level. In the example of Fig.1b, the microphone 1 will capture more signal from source 2 than from source 1, and on the contrary, the microphone 2 will capture more signal from source 1. Consequently, two observations or mixed signals are obtained in a different manner, such that there are no time differences between the sources that are present at each mixture. That is to say, ideally we are being able to obtain an instantaneous mixture system for acoustic signals.

Analyzing the proposed mixture system for the case of two sources and two observations, and being $\mathbf{s}[n] = [s_1[n]\ s_2[n]\ ]^{\mathrm{T}}$ the sources vector, $\mathbf{x}[n] = [x_1[n]\ x_2[n]\ ]^{\mathrm{T}}$ the observations vector and $h_{ji}[n]$ the impulse response of a LTI system that connects the i^{th} observation with the j^{th} source ($i, j = 1,2$), we can express the mixture model as

$$\begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} = \begin{bmatrix} h_{11}[n] & h_{12}[n] \\ h_{21}[n] & h_{22}[n] \end{bmatrix} * \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} \tag{1}$$

Supposing that the directive property of the microphone i^{th} towards the source j^{th}, d_{ji} is frequency independent, the impulse response can be decomposed in a term that reflects the characteristics of the acoustic environment $h'_{ji}[n]$ multiplied by a directivity factor d_{ji}

$$h_{ji}[n] = \mathrm{d}_{ji} \quad h'_{12}[n], \qquad i,j = 1,2 \tag{2}$$

In addition, when two microphones are summoned in the same point, the impulse response from a certain source towards each one of the microphones can be supposed to be approximately the same:

$$\begin{aligned} \mathrm{h'}_{11}[\mathrm{n}] &= \mathrm{h'}_{12}[\mathrm{n}] = \mathrm{h'}_{1}[\mathrm{n}] \\ \mathrm{h'}_{22}[\mathrm{n}] &= \mathrm{h'}_{21}[\mathrm{n}] = \mathrm{h'}_{2}[\mathrm{n}] \end{aligned} \tag{3}$$

Taking into account these considerations, we can rewrite (1) as:

$$\begin{aligned} \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} &= \begin{bmatrix} \mathrm{d}_{11}\ h'_{11}[n] & \mathrm{d}_{21}\ h'_{21}[n] \\ \mathrm{d}_{12}\ h'_{12}[n] & \mathrm{d}_{22}\ h'_{22}[n] \end{bmatrix} * \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} \\ \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} &= \begin{bmatrix} \mathrm{d}_{11}\ h'_{1}[n] & \mathrm{d}_{21}\ h'_{2}[n] \\ \mathrm{d}_{12}\ h'_{1}[n] & \mathrm{d}_{22}\ h'_{2}[n] \end{bmatrix} * \begin{bmatrix} s_1[n] \\ s_2[n] \end{bmatrix} \\ \begin{bmatrix} x_1[n] \\ x_2[n] \end{bmatrix} &= \begin{bmatrix} \mathrm{d}_{11} & \mathrm{d}_{21} \\ \mathrm{d}_{12} & \mathrm{d}_{22} \end{bmatrix} \cdot \begin{bmatrix} s_1[n] * h'_{1}[n] \\ s_2[n] * h'_{2}[n] \end{bmatrix} \end{aligned} \tag{4}$$

These results lead us to consider that the observed signals constitute an instantaneous mixture of source signals modified by the effect of a concrete acoustic environment. The problem becomes the separation of instantaneous mixtures, where the independent components represent the signal that would have been recorded in the case that just the corresponding source had been active. If the main objective is a spatial separation of cue source and the minimisation of the interferences introduced by other sources, this perspective of the problem would be optimal.

In a practical application it will be more complicated to accomplish the requirements established in the proposed model. Firstly, a small positioning error will exist in the transducers that conform the array of coincident microphones and secondly, the reception response of the microphone, i.e. the directivity, does not remain constant with the frequency. However, in spite of these factors, with this configuration we can force that the convolutive mixture is generated according to an instantaneous model, achieving higher separation degree.

3 Real Room Experiment

To verify the previous ideas, some experiments using both configurations of microphones, coincident and separated, have been carried out. These experiments consist of a simple scenario of two sources and two microphones (2x2) summoned in two different acoustic environments: in an anechoic chamber and in a recording studio. These configurations are illustrated in Fig. 2.

For each one of the rooms the following steps were carried out:

1. Different acoustic signals are emitted by each speaker (intermittent voice + guitar), firstly simultaneously and next, separately. Hence, both mixed signals and separated signals are captured by the microphones.
2. Convolutive [7], [8] and instantaneous [9] BSS algorithms are applied to the observations. In the case of convolutive algorithms, the longitude (taps) of the separation filter was set from 1 (instantaneous mixing) until the maximum permitted by the algorithm.

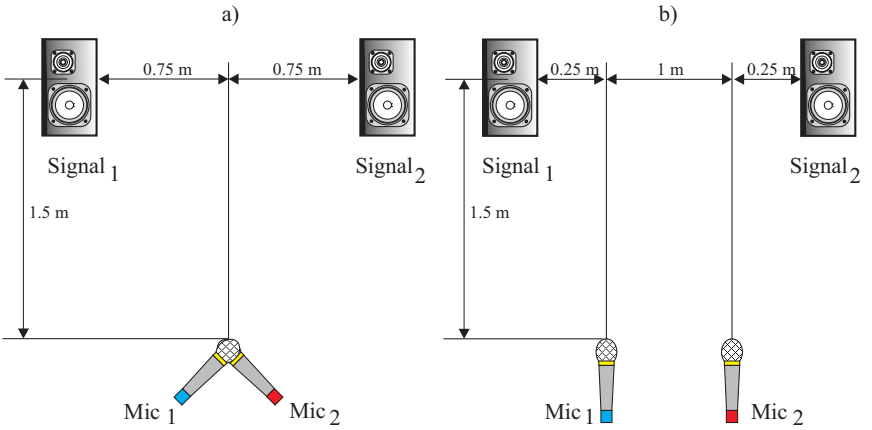

Fig. 2. Configuration of sources and microphones inside the acoustic chamber used for evaluating the influence of the microphones, according to coincident (a) and separated schemes (b).

The measurement of the separation degree is carried out by computing the parameter Signal to Interference Ratio (*SIR*), defined in [10].

The sources signals that had been registered independently (only one speaker emitting), can be used to verify that the impulse response h_{i1} is very similar to h_{i2} in the case of coincident microphones, as it had been previously hypothesized. The cross-correlation (h_{11}/h_{12}) and the autocorrelation (h_{11}/h_{11}) were computed for each configuration. The cross-correlation and autocorrelation functions corresponding to the impulse responses h_{11} and h_{12} recorded in an anechoic chamber are presented in fig. 3 and fig. 4 in the case of separated and coincident microphones, respectively. In addition, the cross-correlation and autocorrelation functions corresponding to the impulse responses h_{11} and h_{12} recorded in a recording studio are presented in fig. 5 and fig. 6 as well.

In these figures we can observe the following:

- In the configuration of coincident microphones the autocorrelation and the cross-correlation functions resemble each other, in both anechoic and reverberant chambers, with no delay differences on the registered signals.
- In the configuration of separated microphones, there exists some similarity between the correlation functions as well, although it is observed a temporal shift due to different arrival times.

- Taking into consideration the autocorrelation function as the main signal, the cross-correlation of the measures in the reverberant room presents a higher degree of distortion, being more significant in the case of separated microphones.

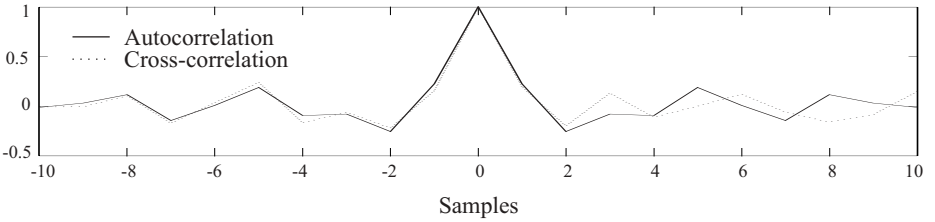

Fig. 3. Cross-correlation and autocorrelation functions of the impulse responses h_{11} and h_{12} in an anechoic chamber for coincident microphones (sample frequency = 44100 Hz).

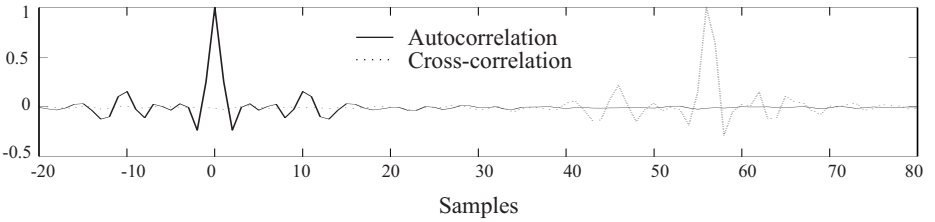

Fig. 4. Cross-correlation and autocorrelation functions of the impulse responses h_{11} and h_{12} in an anechoic chamber for separated microphones (sample frequency = 44100 Hz).

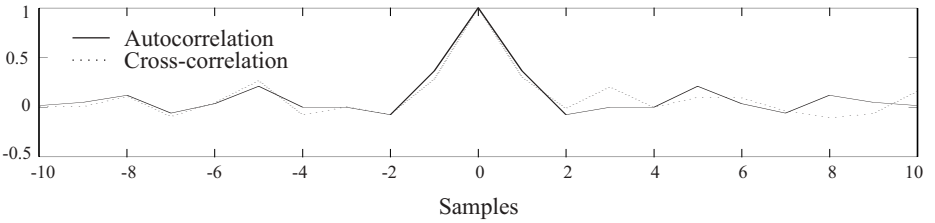

Fig. 5. Cross-correlation and autocorrelation functions of the impulse responses h_{11} and h_{12} in a recording studio for coincident microphones (sample frequency = 44100 Hz).

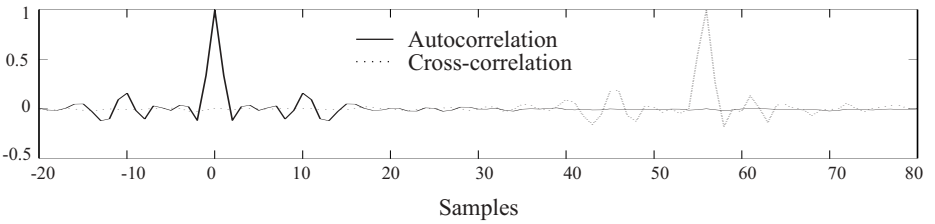

Fig. 6. Cross-correlation and autocorrelation functions of the impulse responses h_{11} and h_{12} in a recording studio for separated microphones (sample frequency = 44100 Hz).

The previous analysis regarding the correlation functions confirms that the observations obtained by coincident microphones in low-reverberant chambers can be approximated to instantaneous mixtures.

In practice, these observations do not correspond to an ideal instantaneous mixture, but at least, the number of samples of the separation filter is minimised. Indeed, this is an advantage, since convolutive algorithms provide better results with shorter filter lengths. The distortion that appears in the cross-correlation for the configuration of coincident microphones is due to the reflections in a real acoustic environment. Furthermore, any temporal delay between the cross-correlation function in reverberant and anechoic chambers is due to these reflections. However, this difference of time is much less than the existent using a configuration of separated microphones. Therefore, the separation filter length will be necessarily shorter, thus reducing the computational cost of the separation algorithm.

4 Convolutive BSS

By applying the convolutive BSS algorithms [7], [8] and varying the filter tap number, the independent sources are estimated, and the performance is measured according to the *SIR* parameter. The performance obtained with coincident and separated microphones are compared in figs. 7 and 8 for the anechoic chamber and the recording studio respectively.

From the experimental results, it can be assessed that:

- The configuration with coincident microphones permitted the best approximations, independent of the acoustic chamber. The resulting filter lengths were actually short (ideally it should be one unique sample). In particular, the algorithm contributes with a separation that is comparable with both configurations of microphones when high longitudes of the separation filter are employed.
- The results obtained with coincident microphones seem to be independent of the acoustic characteristics of the enclosure.

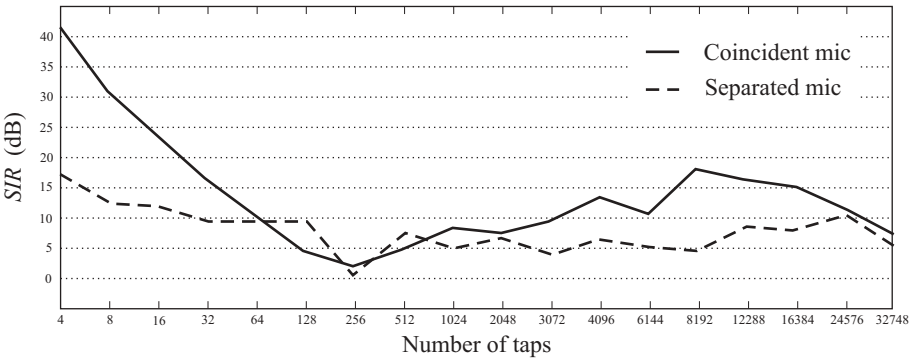

Fig. 7. *SIR* obtained with convolutive BSS algorithm for coincident and separated microphones in a anechoic chamber.

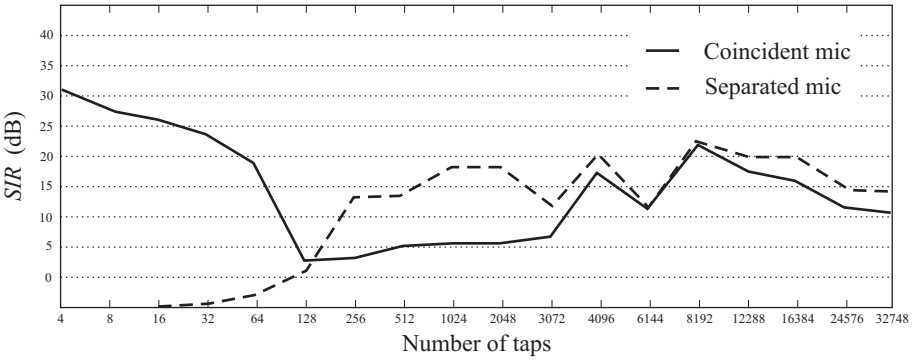

Fig. 8. *SIR* obtained with convolutive BSS algorithm for coincident and separated microphones in a recording studio.

- The convolutive BSS algorithm obtained better results for short filter lengths. Due to convergence limitations of the algorithm, it was not possible to test filter lengths close to one sample. However, extrapolating the tendency of the results, it is reasonable to think that the algorithm would have provided good results if an instantaneous mixture model had been applied.

5 Instantaneous BSS

In order to validate whether the mixtures registered by coincident microphones can be approximated to instantaneous mixtures, an instantaneous BSS algorithm has been applied to the observations [9]. The results in terms of *SIR* are reflected in Table 1. As it can be observed, the separation degree of the estimated sources is very satisfactory. Hence, the theoretical approximation to instantaneous mixtures has been corroborated empirically.

The main difference of the estimated sources via instantaneous BSS with respect to convolutive algorithms is that in the first case, the estimated sources conserve the acoustic effect due to the chamber, whereas in the second case this effect is minimized. However, it does not generally constitute a limitation of the separation algorithm, since the main objective is usually to minimize the interferences introduced by other acoustic sources.

Table 1. Performance measurement in terms of *SIR* using a configuration of two sources registered by coincident microphones in an anechoic chamber and a radio studio. The sources are separated using an instantaneous BSS approach.

	Voice	Guitar	Average
Anechoic chamber	40.9 dB	20.1 dB	30.5 dB
Recording studio	24.7 dB	18.8 dB	21.7 dB

6 Conclusion

With this work it has been demonstrated that in simple configurations of two sources and two microphones, the separation results using convolutive BSS algorithms are optimal taking into consideration a lower taps number. It is even possible to apply instantaneous BSS algorithms obtaining satisfactory results. With the approximation to the instantaneous case, it is possible to recover the source signals convolved with the impulse response of the chamber with a considerable decrease in the computational cost. In opposition to these methods, convolutive BSS algorithms aim to recover the sources exempt of the acoustic effect of the chamber.

References

1. H. Sahlin and H. Broman. "Signal separation applied to real world signals," *Proceedings of Int. Workshop on Acoustic Echo and Noise Control*, London, UK, September, 1997.
2. D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," *1st International Conference on Independent Component Analysis and Blind Signal Separation (ICA'99)*, Aussois, France, pp. 261-266, January 1999.
3. J. M. Sanchis, "*Evaluation of mixture conditions in convolutive blind source separation for audio applications*," Ph. D. Thesis, Universidad Politecnica de Valencia, December 2003.
4. R. Mukai, S. Araki and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation," *3rd International Conference on Independent Component Analysis and Blind Signal Separation*, San Diego, California, USA, pp. 230-235, December 2001.
5. D. V. Rabinkin et al., "Optimun microphone placement for array sound capture," *Proceedings of 133rd Meeting of the Acoustical Society of America*, State College, Pennsylvania, USA, pp. 227-239, June 1997.
6. J. R. Hopgood, P. J. W. Rayner and P. W. T. Yuen, "The effect of sensor placement in blind source separation," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, October 2001.
7. D. Schobben, *Real-time Adaptative Concepts in Acoustics*, KluwerAcademic Publishers, 2001.
8. R. H. Lambert, *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, Ph. D. Thesis, University of Southern California, USA, 1996.
9. V. Zarzoso. *Closed-form higher-order estimators for blind separation of independent source signals in instantaneous linear mixtures*. PH. D. Thesis, University of Liverpool, UK, October 1999.
10. S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no 2, pp. 109-116, March 2003.