ARCHIVES OF ACOUSTICS
**32**, 4 (Supplement), 337–342 (2007)

# EFFICIENCY OF BLIND SOURCE SEPARATION IN A REAL ROOM

Paweł LIBISZEWSKI, Jędrzej KOCIŃSKI

Adam Mickiewicz University
Faculty of Physics, Institute of Acoustics
Umultowska 85, 64-614 Poznań, Poland
e-mail: bezcelu@gmail.com, jedrzej.kocinski@amu.edu.pl

(*received July 15, 2007; accepted November 12, 2007*)

The present study is concerned with the effectiveness of the blind source separation (BSS) in real acoustical environment measured by means of a speech reception threshold, SRT. BSS is a multisensoral method that leads to a signal extraction from a mixture of sounds. The performance of such algorithms are most often described by means of the increase in signal-to-noise ratio, SNR. However, the SNR is not an appropriate measure for speech enhancement, since the increase in SNR does not lead directly to the improvement of the speech intelligibility. Thus, the relationship between an increase in SNR and speech intelligibility is not straightforward.

This work shows some experiments in which the speech masked by a babble noise, music or concurrent speech was separated out using an algorithm for convolutive BSS. The SRT before and after the separation was measured for subjects with audiologically normal hearing.

All recordings were carried out in a small office room using an array of two microphones. Two spatially separated loudspeakers were used as sources of signals: target speech and disturbances.

The results of the experiment revealed a marked improvement in the speech intelligibility: the decrease in SRT reached about 7 to 24 dB in individual cases.

**Keywords:** blind source separation, speech enhancement, speech intelligibility.

## 1. Introduction

In most real situations people with hearing loss complain about a poor understanding of an interlocutor, particularly if there exist some interferences nearby. An amplification of the signal is insufficient, because all the signals (desirable and undesirable) are amplified, thus a signal-to-noise ratio (SNR) remains unchanged. Therefore, it is necessary to increase the SNR to extract the target information from mixture of all signals. There has been a lot of research on the improvement of SNR using different signal processing methods, one of the most recent is Blind Source Separation (BSS). Acoustic signals recorded simultaneously in a natural environment are usually very complex as microphones capture a mixture of sounds coming from several sources. Moreover, each signal in each microphone is delayed as it takes time to reach consecutive sensors. The recorded sound is not a simple superposition of source signals in the microphone, but a convolution of signals and the impulse response that describes the propagation process. Given independent sources (e.g. target speech and maskers) $s_m(t)$, $m = 1, 2, \ldots,$ $M$, where $t$ denotes time, the real mixing process (including delays) can be assumed as:

$$x_n(t) = \sum_{m=1}^{M} \sum_{k=0}^{K} s_m(t-k) a_{nm}(k) \tag{1}$$

$M$ is the number of the independent sources $s_m$ and $a_{nm}$ are the length $K$ mixing filters, which describe the delays at measuring points and the impulse response of the room.

The goal of the BSS is to filter the signals from a microphone array to extract source signals while reducing interfering signals:

$$u_i(t) = \sum_{n=1}^{N} \sum_{k=0}^{K} x_n(t-k) h_{in}(k), \tag{2}$$

where $h_{in}$ are the unmixing filters to be estimated. As can be seen in Eqs. (1) and (2) there exist a convolution of signals. In other words the aim is to invert the mixing process and find an unmixing matrix, so that $u_i(t) = s_i(t)$. To separate source signals from their mixtures, statistical methods are used. It means that the objective of BSS is to solve Eq. (2) so that the signals $u_i(t)$ are as independent as possible. To capture statistical independence some statistic measures are required [1–2]. This problem can be solved using the approach based on non-stationarity properties and second order statistics. This problem has been studied by MATSUOKA *et al.* [3], PARRA and SPENCE [4] and PHAM *et al.* [5]. It was shown that decorrelation is able to perform the BSS task for wide class of source signals. There is also one important disadvantage of the BSS. At the simplest assumption it needs at least as many sensors (microphones) as signals (sound sources – speakers and maskers) [2].

## 2. Aim

In the present study an attempt to investigate the speech intelligibility enhancement using convolutive BSS method in real acoustical environment, namely office room, was presented. The main aim of the study was to compare the subjective speech intelligibility before and after the BSS was proceeded. The perception of speech signal is usually measured in terms of its quality (naturalness and ease of listening) and intelligibility (percentage of words/sentences that can be correctly identified by listeners). Most of speech enhancement techniques improve speech quality. However, some of them decrease the intelligibility. It is because speech enhancement algorithms can distort the target speech signal. Thus, SNR that is usually accepted as an objective measure of efficiency [4] does not reflect the intelligibility. Hence, speech intelligibility measurements (e.g. SRT) should be proceeded to assess the "global" efficiency of speech improvement algorithms.

## 3. Experiments

Three experiments were carried out. In each of them the speech intelligibility of the Polish Sentence Test [6] in the presence of interfering signal was measured. Moreover, in each experiment different type of interference (with different statistical properties) was used: babble noise (Experiment 1), music (Experiment 2) and concurrent speech (Experiment 3). Then 10 subjects with otologically normal hearing were asked to take part in the experiments.

### *3.1. Algorithm*

One of the approaches to solve the problem of convolutive Blind Source Separation was presented by PARRA and SPENCE [4] and patented (US patent US6167417). The program *convbss* [7] by HARMELING that implements the algorithm for Blind Source Separation of convolutive mixtures by Parra and Spence was used in the present study to proceed the BSS. This is non-on-line program that uses least square optimization. Two different lengths of the estimated separating filters were taken into account, namely: 256 samples (short) and 22 050 samples (long). The first one refers to the earlier researches by KOCIŃSKI [8] carried out in an anechoic chamber and the second one respects the reverberation time in an experimental room. It must be emphasized that the longer impulse response of the filter, the more time-consuming is the BSS.

### *3.2. Apparatus*

All recordings were carried out in an office room (reverberation time, $\mathrm{RT} \approx 0.5$ s) using two microphones Sennheiser e915. The signals were recorded separately (sampling rate 44 100 samples/s) and then they were mixed in the computer with the given SNR. Next, the separating filters (short and long, separately) were determined for one, randomly chosen, sentence and 51 SNRs (from 0 to $-50$ dB, with the 1 dB step). The filters were then stored on the hard drive. In the study the adaptive method (1-up/1-down with 1-dB step) was used to determine the Speech Reception Threshold (SRT) before (without the use of the separating filters) and after BSS for short and long filters. This method was introduced by OZIMEK *et al.* [6]. All presented results are the arithmetic mean for 10 subjects for each configuration, separately.

### *3.3. Experiment 1. Speech masked by babble noise*

In this experiment the sentence test signal was masked by so-called babble noise made by mixing all sentences from the test [6]. The spatial configuration used in the experiments is depicted in Fig. 1.

The results, i.e. SRT for each of the conditions (before BSS, after BSS with short filters, after BSS with long filters) used in the research are shown in Fig. 2. As can be seen the SRT after BSS method was applied is significantly decreased. Moreover, using

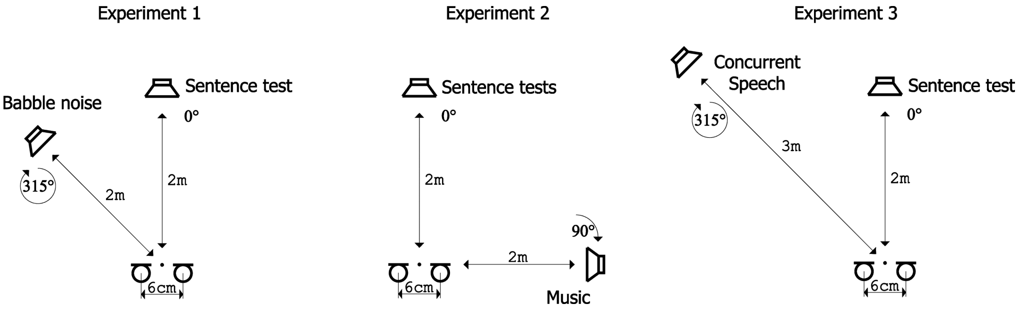

Fig. 1. Spatial configuration used in Experiment 1, 2 and 3.

long separating filters one can decrease the SRT much more (more than 20 dB), than using short ones (about 8 dB). It must be emphasized that the lower SRT, the better speech intelligibility. Thus it can be stated that BSS significantly improves the speech intelligibility.

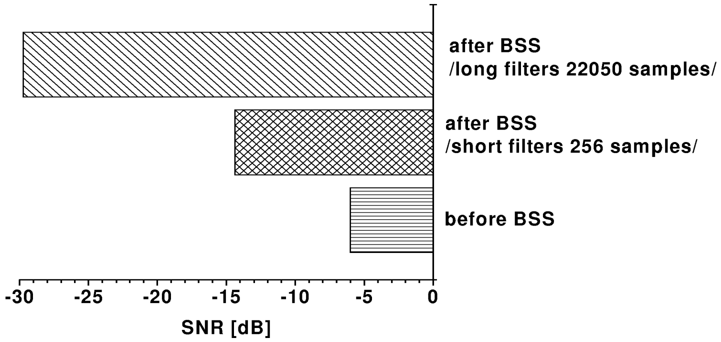

Fig. 2. Results of Experiment 1 – SRT determined for three different conditions used in the research: before BSS, after BSS with short separating filters (256 samples) and after BSS with long separating filters (22 050 samples).

### *3.4. Experiment 2. Speech masked by music*

The main goal of this experiment was similar to the Experiment 1, however music signal was used as the interference instead of babble noise. Such a signal is characterized by different statistical properties (non-stationarity) comparing to stationary noise. Moreover, the spatial configuration of the interference was different (90° clockwise) while the test signal source was placed at the same angle.

Similarly to Experiment 1, SRT for three different conditions (before BSS, after BSS with short filters and after BSS with long filters) was determined. The results are depicted in Fig. 3.

Again, the BSS showed a high efficiency in terms of subjective speech intelligibility improvement that can be noticed as decrease in SRT after BSS was applied. However, this effectiveness for short filters is much lower than in Experiment 1. The higher ef-

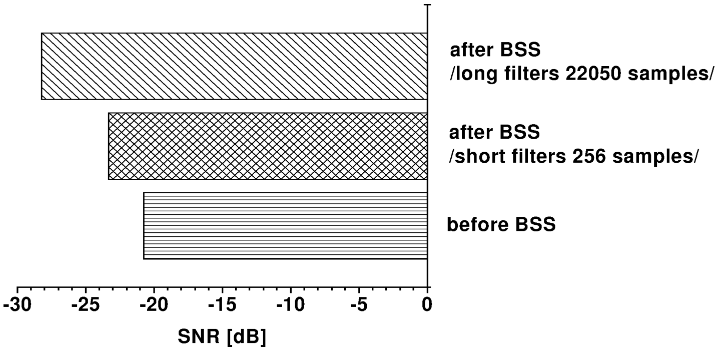

Fig. 3. Results of Experiment 2 – SRT determined for three different conditions used in the research: before BSS, after BSS with short separating filters (256 samples) and after BSS with long separating filters (22 050 samples).

ficiency of long separating filters was also confirmed. In this case the decrease in SRT reaches about 7 dB comparing to the before BSS condition.

### 3.5. *Experiment 3. Speech masked by concurrent speech*

In the last experiment the interference of the same type as test material was used, i.e. concurrent speech. The test signal source was placed as in previous experiments, while the concurrent speech source was placed at 315° clockwise.

The spatial configuration used in the experiment is shown in Fig. 1, while in Fig. 4 the results are shown in the similar way and for the same conditions as in previous experiments.

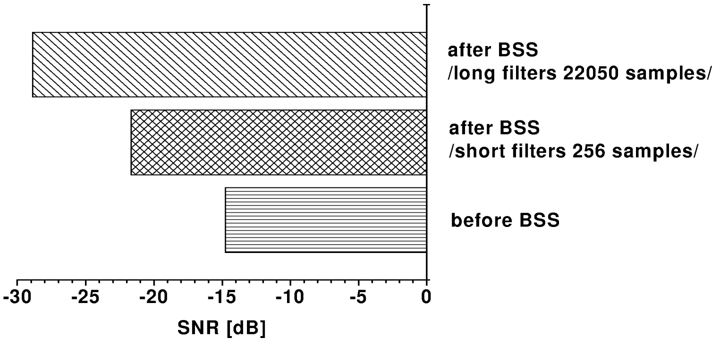

Fig. 4. Results of Experiment 3 – SRT determined for three different conditions used in the research: before BSS, after BSS with short separating filters (256 samples) and after BSS with long separating filters (22 050 samples).

As can be noticed, for such signals, the effectiveness of the BSS method is lower than the effectiveness in Experiment 1. However, comparing to Experiment 2, the effectiveness of the BSS method is relatively high.

## 4. Conclusion

The present research has proved a high efficiency of convolutive BSS in subjective speech intelligibility enhancement in real acoustical environment for different types of interferences and spatial configuration. The decrease in SRT reaches even more than 20 dB for stationary noise and long separating filters while for non-stationary signals the decrease in SRT after BSS was applied comparing to the before BSS case reaches from 7 dB (for music) to about 15 dB (for concurrent speech). It must be emphasized that long separating filters appeared more effective in all cases.

Moreover, there exist a very significant difference in SRTs for the before BSS case in all three experiments. It seems reasonable to state that this is the result of the different properties of the interferences used in particular experiment. However, the different spatial configuration should be also taken into consideration.

Variableness of two factors (type of masker and spatial configuration) determined the character of the results received in presented experiments. It may be not the exact dependence of convolutive BSS efficiency on the type of masker, but the qualitative assessment, which emphasized versatility of described method in a reverberant environment.

## Acknowledgments

This work was supported by The Ministry of Science and Higher Education, project number N517 028 32/432. The authors would like to thank two anonymous reviewers for useful comments and remarks on the earlier version of this manuscript.

## References

[1] Cardoso J.-F., *Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem*, Proceedings of the ICASSP 89, pp. 2109–2112, (1989).

[2] Hyvärinen A., Karhunen J., *et al.*, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.

[3] Matsuoka K., Ohya M., *et al.*, *A neural net for blind separation of nonstationary signals*, Neural Networks, **8**, 3, 411–420 (1995).

[4] Parra L., Spence C., *Convolutive blind source separation of non-stationary sources. US Patent US6167417*, IEEE Trans. on Speech and Audio Processing, **8**, 3, 320–327 (2000).

[5] Pham D.-T., Serviere C., *et al.*, *Blind separation of convolutive audio mixtures using nonstationarity*, ICA 2003, Nara, Japan 2003.

[6] Ozimek E., Kutzner D., *et al.*, *The Polish sentence test for speech intelligibility measurements*, Archives of Acoustics, **31**, 4S, 435–442 (2006).

[7] Harmeling S., *convbss.* FRAUNHOFER FIRST Berlin, Berlin 2001.

[8] Kociński J., *Influence of blind source separation on speech intelligibility*, Archives of Acoustics, **30**, 4S, 147–150 (2005).