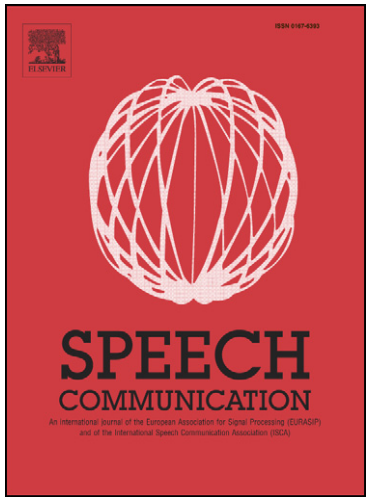
# Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms

Jedrzej Kocinski




# Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms

Jedrzej Kocinski

PII: S0167-6393(07)00117-3
DOI: 10.1016/j.specom.2007.06.003
Reference: SPECOM 1653

To appear in: *Speech Communication*

Received Date: 20 March 2006
Revised Date: 15 June 2007
Accepted Date: 15 June 2007

Please cite this article as: Kocinski, J., Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms, *Speech Communication* (2007), doi: 10.1016/j.specom.2007.06.003

# Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms[1]

**Jedrzej Kocinski**

Institute of Acoustics, Faculty of Physics, Adam Mickiewicz University, 85 Umultowska str., 61-614 Poznan, Poland, **jen@amu.edu.pl**



[1] This work has been partially presented at ForumAcusticum 2005 in Budapest.

**Abstract**

The present study is concerned with the blind source separation (BSS) of speech and speech-shaped noise sources. All recordings were carried out in an anechoic chamber using a dummy head (two microphones, one in each ear). The program which implements the algorithm for BSS of convolutive mixtures introduced by Parra and Spence (2000a) was used to separate out the signals. In the postprocessing phase two different denoising algorithms were used. The first was based on a minimum mean-square error log-spectral amplitude estimator (Ephraim and Malah, 1985), while the second one was based on Wiener filter in which the concept of an *a priori* signal-to-noise estimation presented by Ephraim (1985) was applied (Scalart and Filho, 1996). Non-sense word tests were used as a target speech in both cases while one or two disturbing sources were used as interferences. The speech intelligibility before and after the BSS was measured for three subjects with audiologically normal hearing. Next the speech signal after BSS was denoised and presented to the same listeners. The results revealed some ambiguities caused by the insufficient number of microphones compared to the number of sound sources. For one disturbance only, the intelligibility improvement was significant. However, when there were two disturbances and the target speech, the separation was much poorer. The additional denoising, as could be expected, raises the intelligibility slightly. Although the BSS method requires more research on optimization, the results of the investigation imply that it may be applied to hearing aids in the future.

**Keywords:**

blind source separation, speech enhancement, denoising, speech intelligibility.

# 1. Introduction

People with a sensorineural hearing loss often suffer from insufficient speech intelligibility. They complain about a poor understanding of an interlocutor, particularly if there exist some interferences nearby. Simple amplification of the signal is insufficient, because all the signals (desirable and undesirable) are amplified, thus a signal-to-noise ratio (SNR) remains unchanged and the intelligibility cannot be increased. Therefore, it is necessary to increase the SNR to extract the target information from noise.

There has been a lot of research on the improvement of speech-to-noise ratio using different signal processing methods. Two of them were used in the present study. Denoising

algorithms are well established, common and used for many years, while the BSS method is a recent fruitful technique for speech intelligibility improvement.

It must be emphasized that 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. Thus, it can be that 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 acoustical environment and the arrival delays. Given independent sources (e.g. target speech and maskers) $s_m(t)$, $m = 1,2,...,M$, where $t$ denotes time, the real mixing process (including delays) can be assumed as:

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

where 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.

The main goal of the convolutive BSS is to filter out the signals from a microphone array to extract original sources while reducing interfering signals. In this case recovering the independent signals $u_i(t)$ can be described as:

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

where $h_{in}$ are the unmixing filters to be estimated. As can be seen in equations (1) and (2) there exist a convolution of signals. The main goal of the BSS method 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 equations (2) so that the signals $u_i(t)$ are as independent as possible. To capture statistical independence some statistic measures are required (Cardoso, 1989; Hyvärinen *et al.*, 2001).

Four fundamental approaches to the separation problem can be enumerated. They are based on the assumption of statistical independence of the signals (Comon, 1991; Jutten and Herault, 1991; Comon, 1994; Hyvärinen *et al.*, 2001).

The first exploits some measure of statistical independence of signals as the cost function, namely non-gaussianity of sparseness. In this approach, the higher-order statistics (HOS) is essential to solve BSS problem (Cichocki and Amari, 2003; Choi *et al.*, 2005).

Another approach exploits the various diversities of signals, typically, time, frequency, (spectral or time coherence) and/or time-frequency diversities, or more generally, joint space-time-frequency (STF) diversity. This approach leads to the concept of Time-Frequency Component Analyzer (TFCA) (Belouchrani and Amin, 1996).

The third approach is to exploit temporal structures of the sources. Each source has non-vanishing temporal correlation. In such a case less restrictive conditions than statistical independence can be used such as second order statistics. Several approaches are based on this assumption (Molgedey and Schuster, 1994; Ziehe *et al.*, 2000; Cichocki and Belouchrani, 2001; Choi *et al.*, 2002; Choi *et al.*, 2003).

The last fundamental approach is based on non-stationarity properties and second order statistics (SOS). The non-stationarity was first taken into account by Matsuoka *et al.* (1995). This problem has been studied by Parra and Spence (2000a) and Pham *et al.* (2003). It was shown that decorrelation is able to perform the BSS task for wide class of source signals. This approach seems to be easier and more reliable as the higher order statistics methods work satisfactorily in computer simulations but perform poorly for recordings in real environment (Parra and Spence, 2000a).

Separating filters can be estimated in: the time domain, the frequency domain and both domains (Makino *et al.*, 2005). As there is a convolution in equation (2), the performance of the algorithms in the time domain is computationally expensive and time consuming (Amari *et al.*, 1997; Kawamoto, 1998; Douglas and Sun, 2003; Buchner *et al.*, 2004). However, they present good results for instantaneous mixtures.

A more common approach to the problem is frequency domain BSS (Smaragdis, 1997; Smaragdis, 1998; Anemueller and Kollmeier, 2000; Parra and Spence, 2000a; Zhou and Xu, 2003; Sawada *et al.*, 2005). It is possible to use an appropriate Fourier Transform to equation (2). In this case the time series become polynomials and the convolution is transformed to the element-wise multiplications:

$$U_n(f) = \sum_{m=1}^{M} X_m(f) H_{nm}(f). \qquad (3)$$

In the frequency domain, this problem becomes easier and can be solved separately at each frequency bin. However, it must be emphasized that moving to the frequency domain, makes the computation easier and faster, on the one hand, while leading to ambiguities of the solution on the other hand, namely the frequency bins at the output of the BSS can be permuted. This problem can be omitted by applying an appropriate operation to the separation

matrix that smoothes the separation matrices in the frequency domain. This can be obtained by reducing the filter length by a rectangular window as suggested by Smaragdis (1998), Parra (2000a), Schobben (2002) and Buchner (2004) or by averaging the separation matrices with neighbouring frequencies (Smaragdis, 1998). However, such an operation changes the separation matrix slightly and can also influence the final separation. There exist more approaches to the ambiguity problems. Saruwatari *et al.* (2003) suggest to use beamforming methods and analyze the directivity patterns determined by the BSS to identify the directions of arrival (DOA) of the sources.

The third approach to the BSS combines two domains. Such an approach is used in the implementation of Parra and Spence's algorithm (2000a) implemented by Harmeling (2001) where the filter coefficients are updated in the frequency domain, but the windowing of the separation filters is processed in the time domain at every iteration step.

The BSS method, however, has one disadvantage that should be mentioned here. At the simplest assumption it needs at least as many sensors (microphones) as signals (sound sources- speakers and maskers) (Hyvärinen *et al.*, 2001).

In the present study the non-on-line algorithm introduced by Parra and Spence (2000a) and implemented by Harmeling (2001) was used. It must be emphasized, that there also exist some algorithms of the convolutive BSS that are able to separate the signals on-line, e.g. (Parra and Spence, 2000b; Asano *et al.*, 2001, Aichner, 2003).

## 2. Aim

The main aim of the study was to compare the non-sense word (logatom) tests intelligibility before and after the BSS was applied. Then the denoising algorithms were applied to the speech signal separated by the BSS. The speech intelligibility after this process was measured and compared with previous results. All signals were recorded in an anechoic chamber to avoid any other effects such as ambient noise and additional reflections from the walls.

The dummy head with two microphones (one in each ear) was used as the set of sensors and there were one or two interfering sources and a target speech.

## 3. Algorithms

One of the approaches to solve the problem of convolutive BSS was presented by Parra and Spence (2000a). The algorithm is based on a calculation of the cross-correlation matrices in multiple times and minimization of a least squares cost function (based on a Forbenius

norm) that leads to the estimation of the separating filters. The program *convbss* by Harmeling (2001) that implements the algorithm for BSS of convolutive mixtures by Parra and Spence was used in the present study to proceed the BSS. This is a non-on-line program that uses least square optimization.

The main advantage of the SOS based on non-stationarity algorithms is its robustness with respect to additive noise, if the number of covariance matrices is sufficiently large. Moreover, they seem to be more reliable in terms of convergence (Parra and Spence, 2000a) and easier to implement. However, they require more time to separate the signals as they are based on statistics taken in different time intervals. The frequency-time domain changing in each iterating step can not be neglected either. However, it was shown in (Kocinski, 2005) that this algorithm performed well with the natural signals recorded in an anechoic chamber.

The parameters used in the experiment were: NFFT: $T=512$, number of matrices to diagonalize: $K=5$, number of intervals used to estimate each cross-power-matrix: $N = rx/T/K$, where $rx$ is the length of the input signals in samples. As the recordings were performed in an anechoic chamber, the length of the separating filters in time domain was set to $Q=128$ samples. The reflections of the signals could be neglected and only mutual time differences between arrival of the signals to particular microphones were important.

As it was mentioned before this algorithm does not work on-line, thus the duration of the mixture signals taken to estimate the separating filters for all SNRs separately was set arbitrarily to 10 seconds. However, the recent (unpublished) research made by the author showed that this duration can be much shorter (3 seconds or even less) to get the same efficiency. The estimated filters were taken to separate all the recorded logatoms from the noise.

Two different denoising algorithms were used in the postprocessing stage. The first was based on a minimum mean-square error log-spectral amplitude estimator (*MMSESTSA*) (Ephraim and Malah, 1985) while the second one was based on an a priori signal-to-noise estimation (*Wiener-Scalart*) (Scalart and Filho, 1996). Both of them were implemented by Zavarehei (2005b; 2005a). The time window chosen for the analysis of the signal spectra was set to 25 ms and overlapping 40% (10 ms).

In the input parameters of these implementations a time of initial noise (a time at the beginning of the input signal in which there is no speech) is required. This initial part of the signal is used to estimate an average noise spectrum and use a simple voice activity detector (VAD) for detection of speech in particular frequency bin. This VAD is based on level difference of the adequate frequency bins in the successive windowed spectra: if the level of

the particular frequency bin was higher enough then the level of the adequate frequency bin of the estimated noise, the analyzed frequency bin was marked as a "speech bin"; in the other case the bin was marked as a "noise bin" and averaged with the noise spectrum.

## 4. Subjects

Three subjects aged 23-25 with audiologically normal hearing were asked to listen to the tests and write down all understood logatoms on a special form. In all figures the subjects are depicted as S1, S2 (the author) and S3. All subjects were instructed and took part in short training session to be familiarized with the task.

## 5. Experiment and Methods

In the experiment a dummy head with two microphones (one in each ear) was used instead of the microphone array. This kind of recording was used to investigate how effective the BSS is during more natural configuration of the sources. This situation takes into account all the changes in an acoustic field connected with the head, i.e. head related transfer function (HRTF). The HRTF influences both, sound pressure level and spectra of the source signals reaching ears and can be an additional factor that influences the effectiveness of the BSS analysis.

### *5.1. Stimuli*

The research consisted in recording of the test material in an anechoic chamber using a dummy head. Target speech stimuli were Polish non-sense word (logatom) tests (Brachmanski and Staroniewicz, 1999). The interference sounds were either one or two sources of speech-shaped noise presented from a loudspeaker. The long-term spectrum of such a noise and a real speech spectrum are identical. The only difference is that noise does not convey any semantic information. The speech-shaped noise power spectrum density function is approximately constant up to the frequency of 500 Hz and above this frequency it decreases by 12 dB per octave (Duquesnoy and Plomp, 1983; Culling and Colburn, 2000). The speech-shaped noise was used to avoid the so-called deep-listening effect in real speech (Moore, 1997).

It is important to keep in mind that sound pressure levels of one noise and that produced by two sources of noise in a checkpoint (just above the dummy head) were identical and were adjusted to 75 dB SPL. It implies that the sound pressure level of each of the two

sources of noise appearing together was 3 dB lower than that in the configuration with one noise only.

### *5.2. Apparatus*

All recordings were carried out in the anechoic chamber using Tucker-Davis Technologies (TDT), System 3 device at the sampling rate of 24414.0625 Hz and the resolution of 24 bits. One speech-shaped noise was generated in a real-time by a TDT-RP2 processor, while the second one was played using a Fostex D824 digital recorder. Both sources of noise were statistically independent. Next, the signals were amplified using Pioneer A-505R to the level of 75 dB SPL and delivered to ZG-60 three-way loudspeakers placed in the anechoic chamber.

The target speech signals, previously stored on a hard drive of the PC as 24 bit binary files, were fed to a TDT-RP2 processor used as a D/A converter. Next the speech signal was fed to a TDT-PA5 programmable attenuator enabling the adjustment of SNR and amplified by a SONY STR-DE475 amplifier to the level of 78 dB SPL. Then the signal was delivered to the three-way loudspeaker ZG-60 placed in the anechoic chamber. The proper SNR was adjusted using a programmable attenuator (TDT-PA5).

The signals were recorded using a Neumann dummy head (separate channel for each ear), connected with two TDT-MA2 microphone amplifiers. Next, both signals were fed to two separate inputs of the TDT-RP2 used as a A/D converter, delivered to the PC and saved on the hard drive as 24-bit binary files (one file for each ear).

All previously recorded signals were analysed using *convbss* algorithm and presented to the subjects (i.e. before and after BSS) in double-walled, acoustically isolated chambers. The signals stored on the hard drive were fed to the TDT-RP2 processor and then amplified in a TDT-HB7 headphone buffer, to the level of 75 dB SPL at the tympanic membrane. Next, the signals were delivered to the Sennheiser HDA580 headphones and presented binaurally to the subjects. All the recordings and presentations were carried out using MatLab 6.5 computing language (MathWorks Inc.).

After the BSS was applied, the best signal was chosen and delivered to both ears. The subject's task was to write down all heard logatoms in a specially prepared form.

### *5.3. Spatial configuration of sources in an anechoic chamber*

In each of the spatial configurations of the sound sources used in the experiment the speech (target) signal, S, was always presented from the loudspeaker placed directly in front

of the dummy head (0˚) and 3 m away from it. The dummy head was at a height of 1,5 m above steel-net floor. The noise sources were also 3 m away from the dummy head at the same height as the source of speech, but they were varying in number (one or two) and in spatial configuration. The azimuthal angle of the first noise source, $N$, varied (0˚ or -60˚ clockwise) while the azimuth of the second noise, $N$, was fixed at 45˚ (see Fig. 1).

Fig. 1

The notation of the different configurations is as follows: the upper index stands for the number of the noise source, whereas the lower one stands for the azimuth of the noise source. Four spatial configurations of the sources were considered: $SN_0$, $SN_{-60}$, $SN_0N_{45}$ and $SN_{-60}N_{45}$ (Fig. 1). It must be emphasized that in the $SN_0$ (or $SN_0N_{45}$) configuration, where both sources were supposed to be placed at the same angle, two sources (loudspeakers) were placed next to each other as close as possible (see Fig. 1).

### *5.4. Results*

Fig. 2 depicts the signals before and after BSS algorithm was applied for the lowest SNRs (-9 dB) and for all spatial configurations. As can be seen, the best results (in terms of SNR increase) were obtained in $SN_{-60}$ configuration. In other cases, the increase in SNR can not be noticed.

Fig 2.

Fig. 3. shows the set of 128 samples-length separating filters in time domain (as the elements of matrix Wt) used to estimate signals shown in Fig. 2. It can be noticed that for the $SN_0$ configuration the absolute values of the impulse response are relatively smaller than the values for the $SN_{-60}$ case (except first sample). Moreover, for all configurations the impulse response absolute values decrease as the sample index increases.

Fig 3.

The data gathered in this experiment, i.e. the speech intelligibility as a function of signal-to-noise ratio (SNR), for all subjects and all spatial configurations are depicted in Fig. 4. Filled circles show results with no BSS, empty squares show results after the BSS only, empty triangles depict the results after both BSS and *MMSESTSA* denoising algorithm while filled asterisks show results after both BSS and *Wiener-Scalart* denoising algorithm.

It is important to keep in mind that only two microphones were used during this experiment and the BSS requires at least as many microphones as sound sources to proceed the analysis correctly. Thus, the speech intelligibility improvement in this experiment depends on the number of sources used.

Fig. 4.

The cumulative distribution function (CDF) of the Gaussian distribution was fitted to mean (across subjects) values obtained in the experiment (for each paradigm separately) by means of last-mean square procedure. By this way the speech reception thresholds (SRTs) were obtained as the mean value parameter of the fitted CDF. The comparison of the values of SRTs for all paradigms is shown in Fig. 5. In the 'no data' case, the speech intelligibility after BSS was applied was too high for all SNRs, thus the unambiguous fitting was impossible.

Fig. 5.

As can be noticed in Fig. 4 the best intelligibility improvement after the BSS was obtained when there were only two sources ($SN_0$ and $SN_{-60}$). Moreover, for these two cases the best results were obtained when the interference was spatially separated form the target speech ($SN_{-60}$). In this case, for all subjects and for high signal-to-noise ratios the difference between speech intelligibility in no BSS case and after the BSS only was applied can be neglected since the subjects were able to understand correctly almost all logatoms. The situation changes with the decrease in SNR. The speech intelligibility with no BSS markedly decreases while the speech intelligibility after BSS remains almost the same and even for SNR=-9 dB reaches about 90 %. Thus the speech intelligibility improvement reaches even 50 percentage points for individual subject while the mean is about 40 percentage points.

This results can be explained on the basis of the appropriate transfer functions between the sources and the sensors or head-related transfer function (HRTF). In the $SN_{-60}$ configuration, sources were spatially separated, thus the source-sensor transfer functions were different. In terms of HRTF (see Fig. 1), the interaural phase and time differences were different from zero, thus the local SNR in the left ear was higher than in the right one. On the other hand, for $SN_0$ the source-sensor transfer functions should be the same and the configuration should be referred to the so-called ill-posed BSS problem. In terms of HRTF, it can be stated that there should be no difference in the SNRs in both ears (no difference between signals in both microphones), so the factor that could make the extraction of source signals possible should play no role at all. However, as it was mentioned above, because of

the recording method (in this case two loudspeakers were placed next to each other, not exactly at the same place), the transfer functions were somewhat different. It seems that this slight difference was enough for BSS to perform effectively. Thus, for the $SN_0$ configuration the speech intelligibility improvement was noticed (see Fig. 2 and Fig. 5), however, the difference in results before and after BSS is smaller comparing to the $SN_{-60}$ configuration. This difference can is also noticeable in Fig. 2, where for $SN_0$ configuration, the increase in SNR can not be noticed, however it the speech enhancement is noticeable in the results of the experiment.

There was much poorer speech intelligibility improvement (or even deterioration was noticed after the BSS was applied) when there were three sound sources, that is in $SN_0N_{45}$ and $SN_{-60}N_{45}$ configurations. It seems, that it was caused by the insufficient number of microphones.

It must be emphasized that the additional denoising in the postprocessing phase also brought about an increase in the speech intelligibility (see Fig. 5). This additional speech enhancement reaches from about 2 to about 4 dB in terms of SRT decrease. It seems reasonable to state that the better enhancement in SNR after the BSS, the better enhancement after denoising (it makes the substraction of the noise "easier" for the denoising algorithm).

The data gathered in this experiment were analyzed using a within-subject analysis of variance (ANOVA) with the three following factors: (1) the type of signal (i.e. No BSS, BSS only, BSS + MMSESTSA and BSS + Wiener-Scalart), (2) signal-to-noise ratio (SNR) and (3) spatial configuration. The type of signal was proved to be statistically significant [$F(3,6)=18.48$, $p<0.002$]. SNR was also proved to be statistically significant [$F(4,8)=166.17$ , $p<0.001$] as well as the spatial configuration of the sources [$F(3,6)=43.78$, $p<0.001$]. Among all the interactions the most important ones were those between type of signal and other factors. All of them were proved to be statistically significant- SNR and type of listening stratum: [$F(12,24)=4,79$, $p<0.001$]; spatial configuration and type of signal: [$F(9,18)=10.89$, $p<0.001$]. The interaction between all three factors was also proved to be statistically significant [$F(36,72)=5.1$, $p<0.001$]. This analysis proved the importance of the BSS and the denoising algorithms in the speech intelligibility improvement. However, the effectiveness of this enhancement methods depends on the configuration and speech-to-noise ratio as it could be expected.

## 6. Conclusion

A significant speech intelligibility improvement was noticed after the BSS algorithm was applied to the set of mixtures of independent signals. For individual subjects and for low signal-to-noise ratios the difference between speech intelligibility before and after the BSS method was applied reaches even more than 40 percentage points and can be increased by additional denoising in the postprocessing stage. However, the experiment brings some ambiguities connected with the insufficient numbers of sensors. When there are more sources than microphones, the algorithm is not able to proceed with the separation properly and there is poor speech intelligibility improvement or even there is no speech enhancement at all.

However, in other cases, when the number of microphones was equal to the number of sources, the speech enhancement was significant even if the target speech and the disturbance was situated almost at the same place. Moreover, an acoustic shadow of the head seems to play an important role that helps to extract the signals. It is important to emphasize that in the experiment all information about interaural phase difference and interaural level difference after the BSS was applied was lost because the same (best) target signal was delivered to both ears. However, the binaural cues can be somewhat retained using the spatial characteristics of the estimated filter, i.e. the angle for which the separating filter spatial transmittance is maximal can be recalculated in interaural time and phase differences. This interaural information is very important for sound source localization and can also be used in the so-called spatial suppression increasing the speech intelligibility (Kocinski and Sek, 2005). Nevertheless, the speech intelligibility improvement was noticed after BSS and denoising algorithms were applied.

Moreover the combination of BSS and denoising algorithms brings about further increase in speech intelligibility. However, it must be emphasized that those algorithms perform well for stationary disturbing signals, such as those used in the present experiment. For non-stationary signals e.g. disturbing speech or music, the effectiveness of such algorithms may be poorer.

The present study has proved an important role of the BSS in the speech intelligibility improvement. It seems to be reasonable to combine the BSS method with other methods that are used in speech processing such as beamforming or automatic speech recognition. Some algorithms have been introduced by other authors, e.g. (Parra and Fancourt, 2002).

**Acknowledgements**

This work was supported by The State Committee for Scientific Research, project number N517 028 32/4327 and 4T11E 01425. The author would like to thank two anonymous reviewers for useful comments and remarks on the earlier version of this manuscript.

**References**

Aichner, R., H. Buchner, et al. (2003). On-line time-domain blind source separation of nonstationary convolved signals. 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan.

Amari, S., S. C. Douglas, et al. (1997). Multichannel blind deconvolution and equalization using the natural gradient. 1st IEEE Workshop on Signal Processing Advances in Wireless Communications.

Anemueller, J. and B. Kollmeier (2000). Amplitude madulation decorrelation for convolutive blind source separation. ICA 2000.

Asano, F., M. Goto, et al. (2001). Real-time Sound Source Localization and Separation System and Its Application to Automatic Speech Recognition. Eurospeech.

Belouchrani, A. and M. G. Amin (1996). A new approach for blind source separation using time-frequency distributions. Proc. SPIE.

Brachmanski and P. Staroniewicz (1999). Phonetic structure of a test material used in subjective measurements of speech quality (in Polish). Speech and Language Technology. Poznan. **3:** 71-80.

Buchner, H., R. Aichner, et al. (2004). Blind source separation for convolutive mixtures: A unified treatment. Audio Signal Processing for Next generation Multimedia Communication Systems. Y. Huang and J. Benesty, Kluwer Academic Publishers**:** 255-293.

Cardoso, J.-F. (1989). Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem. Proc. ICASSP 89.

Choi, S., A. Cichocki, et al. (2002). "Second order nonstationary source separation." Journal of VLSI Signal Processing **32**(1-2): 93-104.

Choi, S., A. Cichocki, et al. (2005). "Blind Source Separation and Independent Component Analysis: A Review." Neural Information Processing - Letters and Reviews **6**(1): 1-57.

Choi, S., A. Cichocki, et al. (2003). "Approximate maximum likelihood source separation using the natural gradient." IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **86**(1): 198-205.

Cichocki, A. and S. Amari (2003). Adaptive Blind Signal and Image Processing Learning Algorithms and Applications. Chichester / New York / Weinheim / Brisbane / Singapore / Toronto, John Wiley & Sons, Ltd.

Cichocki, A. and A. Belouchrani (2001). Sources separation of temporally correlated sources from noisy data using bank of band-pass filters. Third International Conference on Independent Component Analysis and Signal Separation (ICA-2001), San Diego USA.

Comon, P. (1991). "Blind Separation of sources: Problem statement." Signal Process. **24**(1): 11-20.

Comon, P. (1994). "Independent component analysis, a new concept?" Signal Process. **36**(3): 287-314.

Culling, J. F. and H. S. Colburn (2000). "Binaural sluggishness in the perception of tone sequences and speech in noise." Journal of the Acoustical Society of America **107**(1): 517-527.

Douglas, S. C. and X. Sun (2003). "Convolutive blind separation of speech mixtures using natural gradient." Speech Communication **39**: 65-78.

Duquesnoy, A. J. and R. Plomp (1983). "The effect of a hearing-aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise." Journal of the Acoustical Society of America **73**: 2166-2173.

Ephraim, E. and D. Malah (1985). "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator." IEEE Trans. on Speech and Audio Processing. **ASSP-33**(2): 443-445.

Harmeling, S. (2001). convbss. Berlin, FRAUNHOFER FIRST Berlin.

Hyvärinen, A., J. Karhunen, et al. (2001). Independent Component Analysis. New York, John Wiley & Sons, Inc.

Jutten, C. and J. Herault (1991). "Blind separation of sources part I: An adaptive algorithm based on neuromimetic architecture." Signal Process. **24**(1): 1-10.

Kawamoto, M. (1998). "A method of blind separation for convolved nonstationary signals." Neurocomputing **22**(1-3): 157-171.

Kocinski, J. (2005). Blind Source Separation (BSS) of sound sources. ForumAcusticum 2005, Budapest.

Kocinski, J. and A. P. Sek (2005). "Speech intelligibility in various spatial configurations of background noise." Archives of Acoustics **30**(2): 173-191.

Makino, S., H. Sawada, et al. (2005). "Blind Source Separation of Convolutive Mixtures of Speech in Frequency Domain." IEICE Trans. Fundamentals **E88**(7): 1640-1655.

Matsuoka, K., M. Ohya, et al. (1995). "A neural net for blind separation of nonstationary signals." Neural Networks **8**(3): 411-419.

Molgedey, L. and H. G. Schuster (1994). "Separation of mixture of independent signals using time delayed correlations,." Physical Review Letters **72**(23): 3634-3637.

Moore, B. C. J. (1997). An Introduction to the Psychology of Hearing, 4th Ed. London, Academic Press.

Parra, L. and C. Fancourt (2002). An Adaptive Beamforming Perspective on Convolutive Blind Source Separation. Noise Reduction in Speech Applications. G. Davis, CRC Press LLC.

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

Parra, L. and C. Spence (2000b). "On-line Blind Source Separation of Non-Stationary Signals." Journal of VLSI Signal Processing **26**(1/2): 39-46.

Pham, D.-T., C. Serviere, et al. (2003). Blind separation of convolutive audio mixtures using nonstationarity. ICA 2003, Nara, Japan.

Saruwatari, H., S. Kurita, et al. (2003). "Blind source separation combining independent component analysis and beamforming." EURASIP Journal on Applied Signal Processing **11**: 1135-1146.

Sawada, H., R. Mukai, et al. (2005). Frequency-domain blind source separation. Speech Enhancement. J. Benesty, S. Makino and J. Chen, Springer.

Scalart, P. and J. V. Filho (1996). "Speech enhancement based on a priori signal to noise estimation." IEEE International Conference on Acoustics, Speech, and Signal Processing **1**: 629-632.

Schobben, L. and W. Sommen (2002). "A frequncy domain blind signal separation method based on decorrelation." IEEE Trans. Signal Processing **50**(8): 1855-1865.

Smaragdis, P. (1997). Efficient Blind Separation of Convolved Sound Mixtures. EEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz NY.

Smaragdis, P. (1998). "Blind separation of convolved mixtures in the frequency domain." Neurocomputing **22**: 21-34.

Zavarehei, E. (2005a). MMSESTSA85.m.

Zavarehei, E. (2005b). WienerScalart96.m.

Zhou, Y. and B. Xu (2003). "Blind source separation in frequency domain." Signal Process. **83**(9): 2037-2046.

Ziehe, A., K. R. Muller, et al. (2000). "Artifact reduction in biomagnetic recordings based on time-delayed second order correlations." IEEE Trans. On Biomedical Engineering **47**: 75-87.

*Figure captions:*

Fig. 1. Four spatial configurations of the sources and the dummy head in an anechoic chamber.

Fig. 2. Short (10 seconds) examples of the signals before BSS (left column) and after BSS (right column) for all spatial configurations and SNR=-9.

Fig. 3. Set of 128 samples-length separating filters in time domain estimated for the signals from Fig. 2.

Fig. 4. Speech intelligibility as a function of signal-to-noise ratio (SNR) of original signal (before BSS) for three subject (S1, S2, S3) and four different spatial configurations of the sources.

Fig. 5. Speech reception thresholds (SRTs) for within-subject averaged results (taken from Fig. 4).

*FIGURES:*

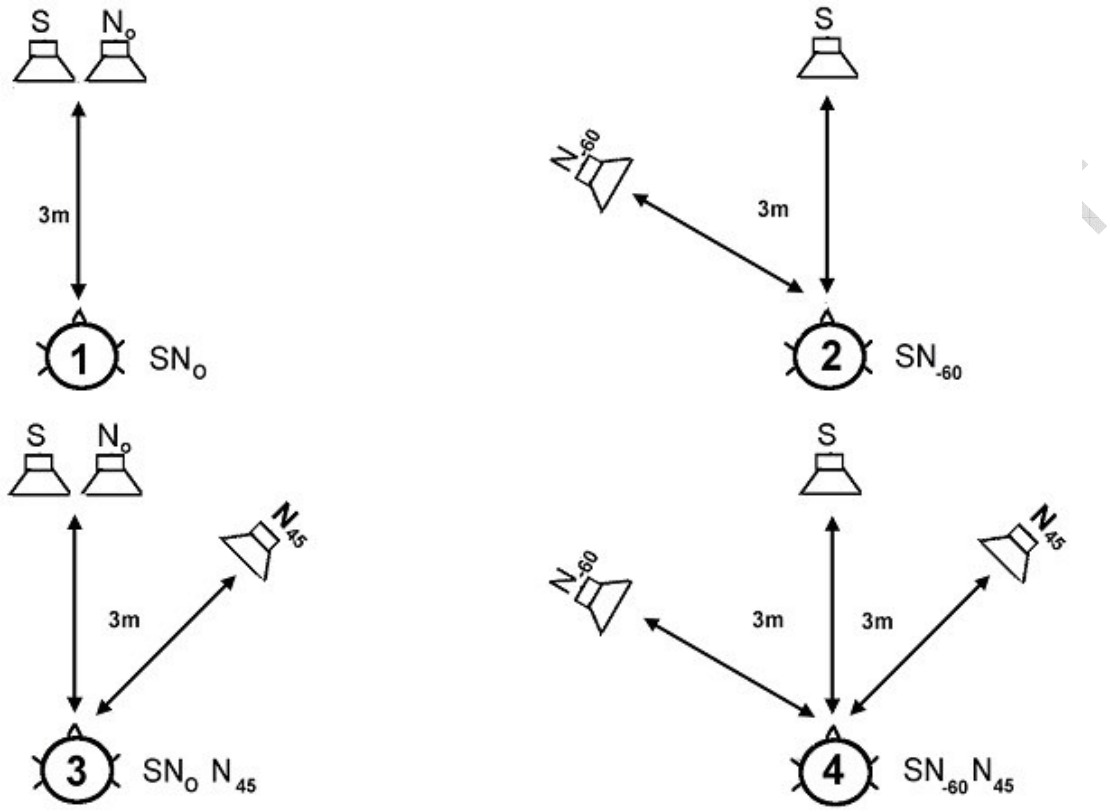
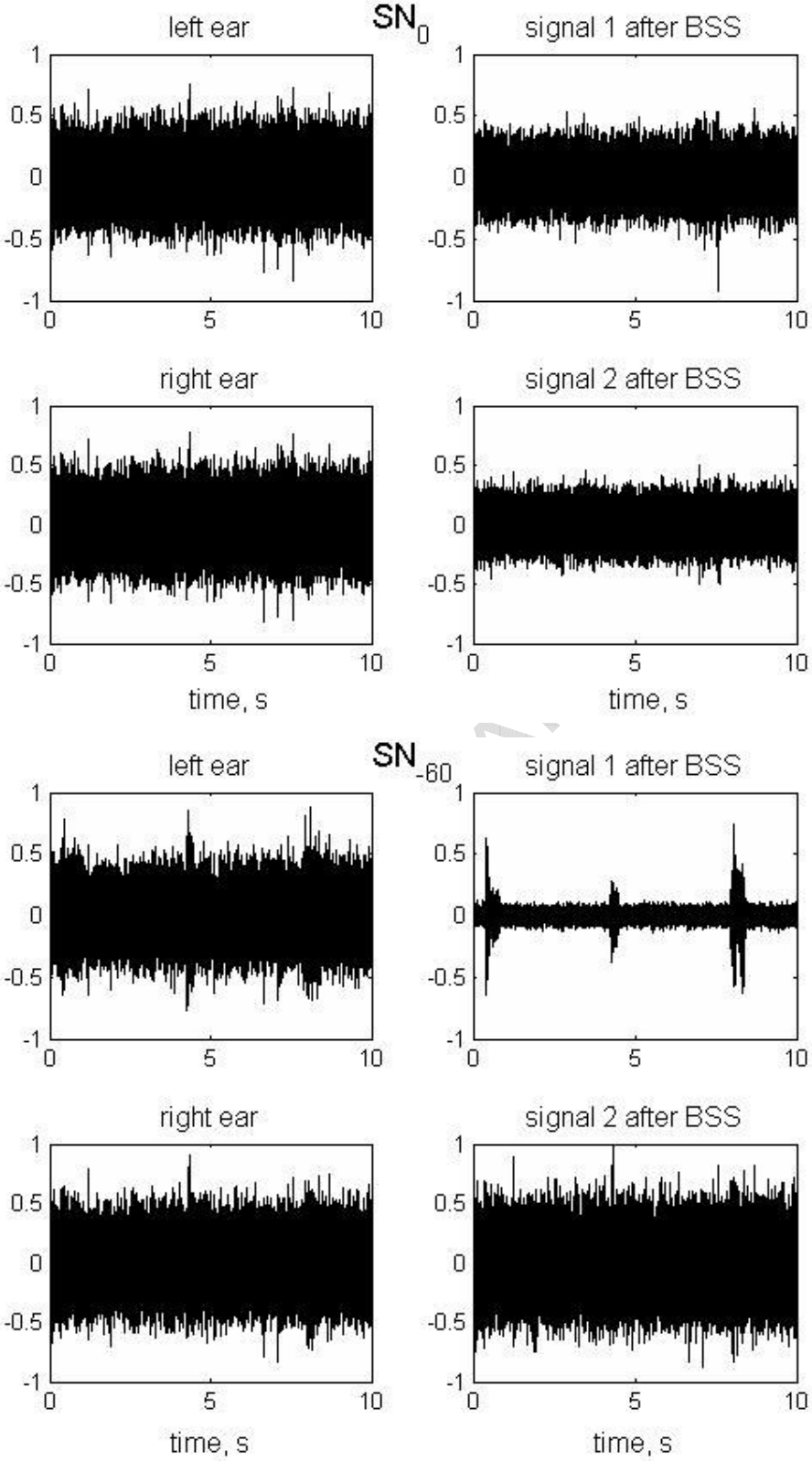



Fig. 1. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.



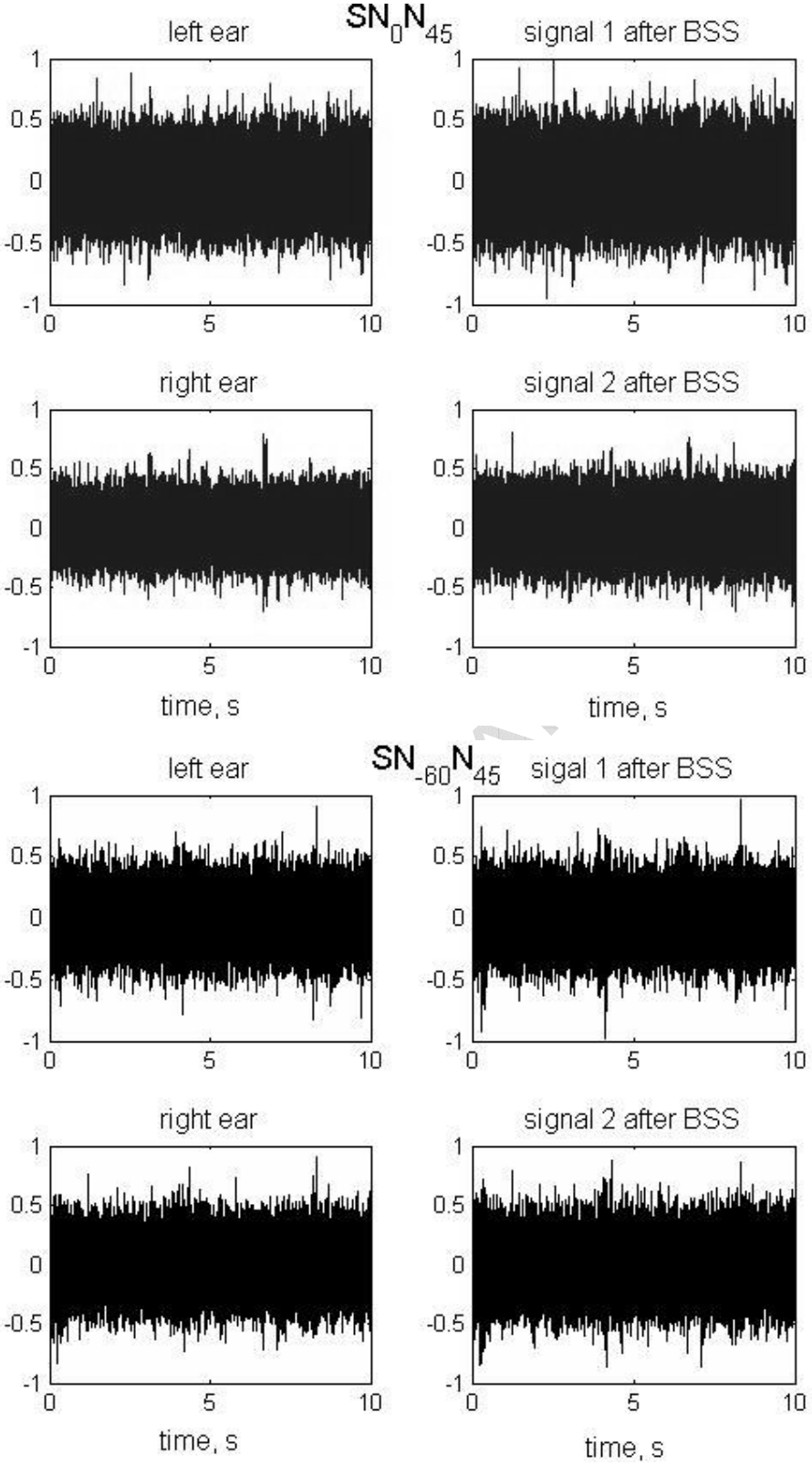
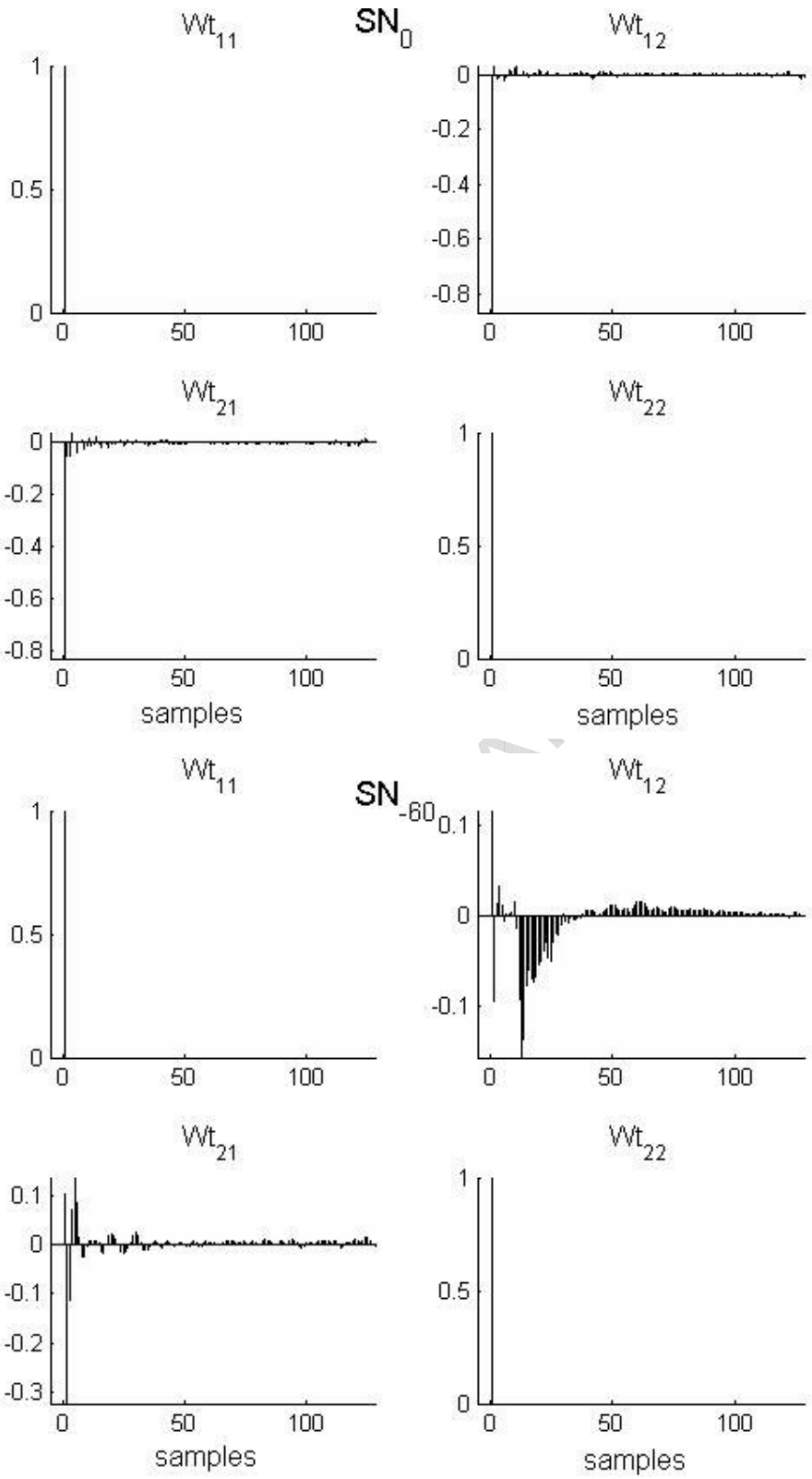



Fig. 2. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.



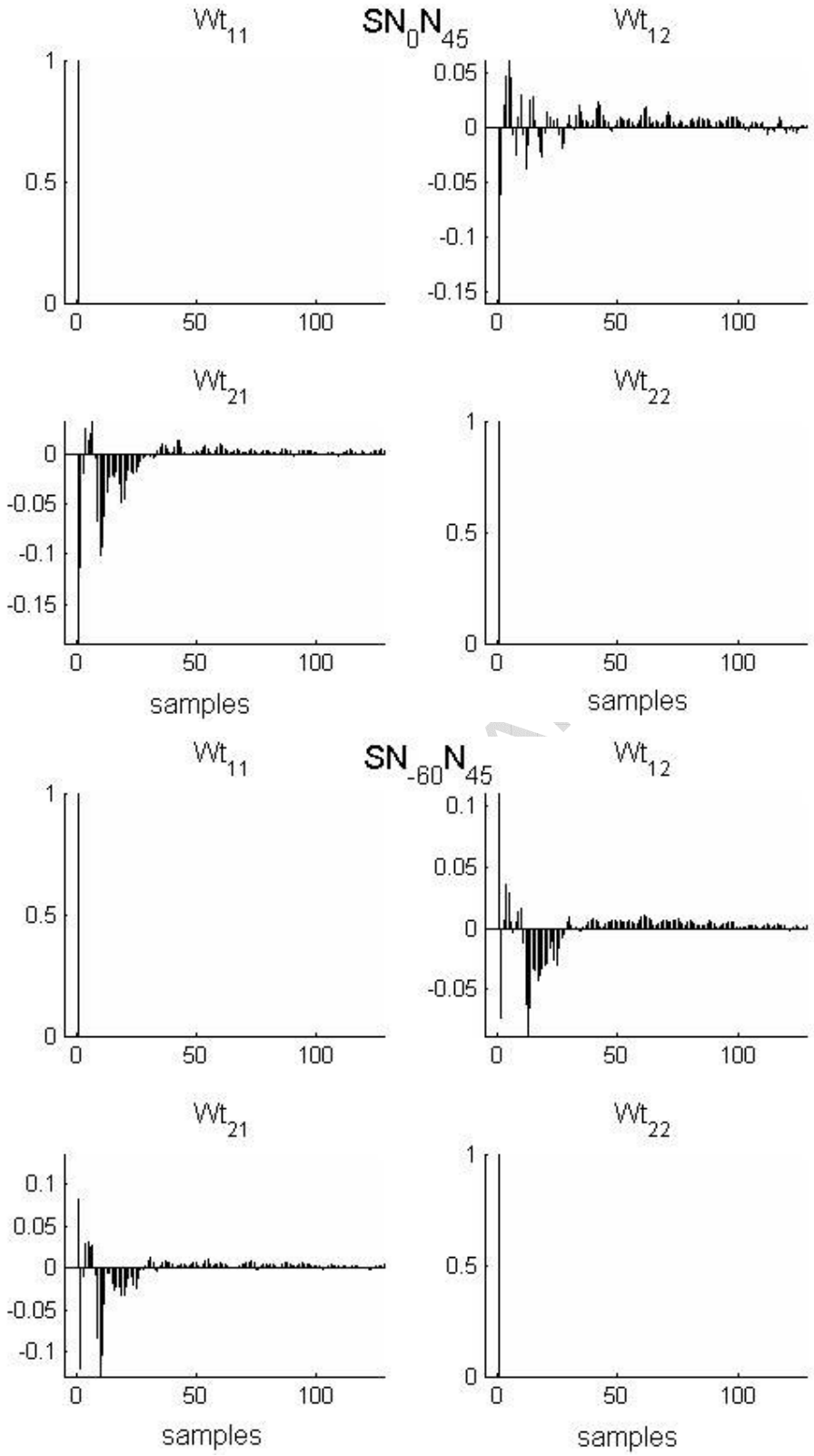



Fig. 3. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.

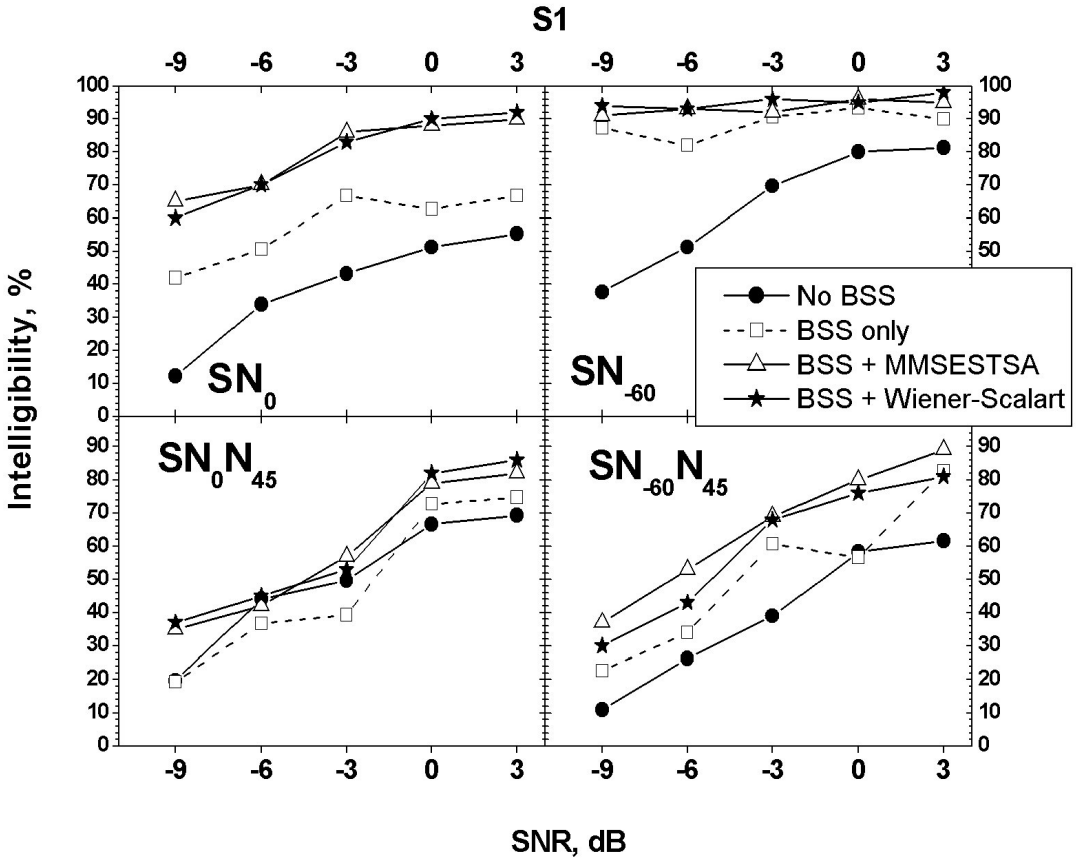



Fig. 4. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.

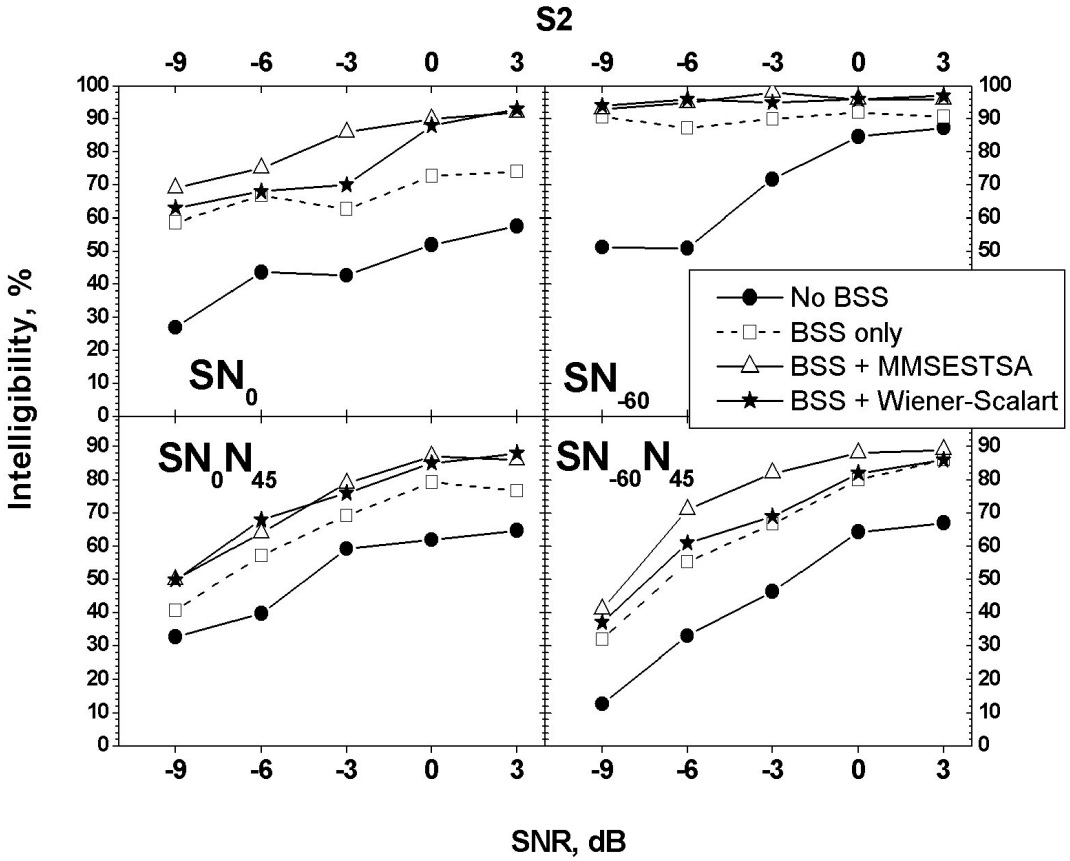



Fig. 4. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.

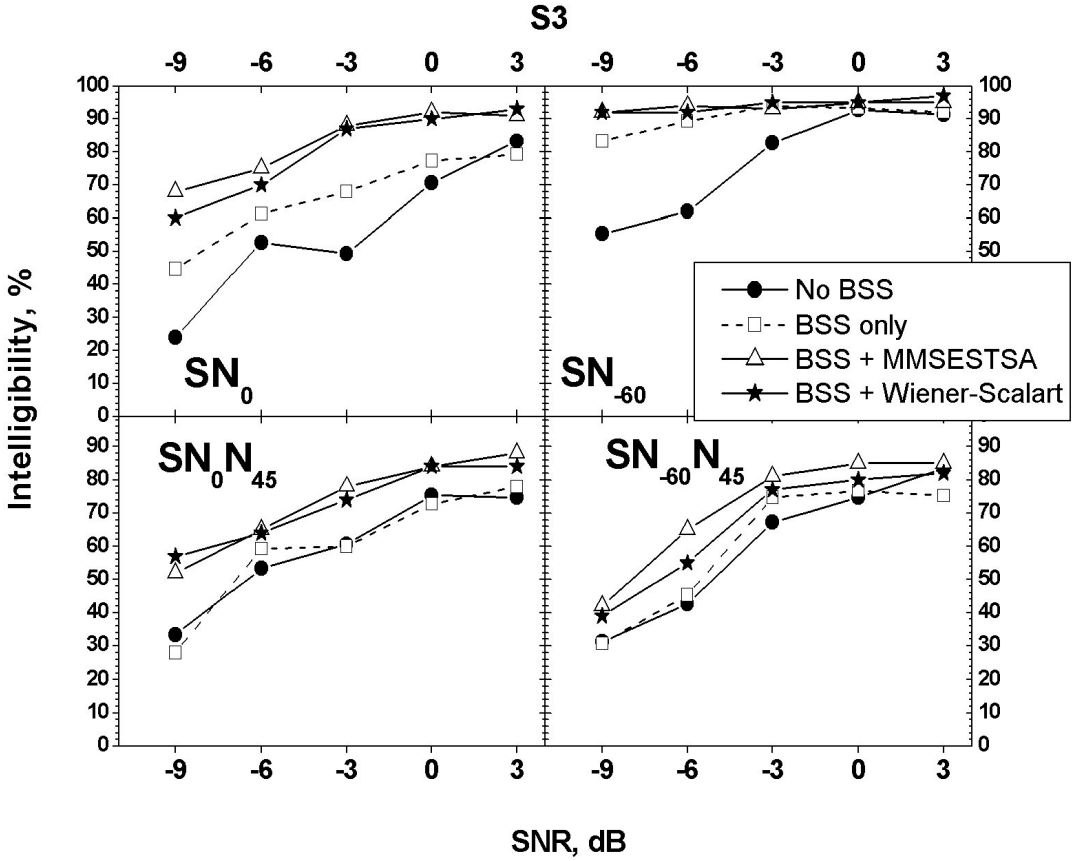



Fig. 4. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.

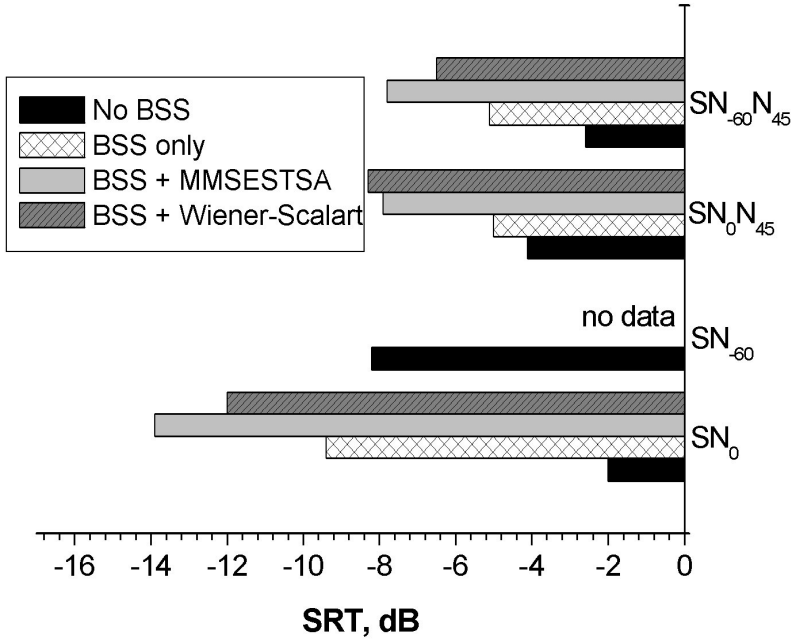



Fig. 5. Jedrzej Kocinski. **Speech Intelligibility Improvement Using Convolutive Blind Source Separation Assisted by Denoising Algorithms**.