A COMBINED CASCADING SUBSPACE AND ADAPTIVE SIGNAL ENHANCEMENT METHOD FOR STEREOPHONIC NOISE REDUCTION

T. Hoya, A. Cichocki, T. Tanaka, G. Hori, T. Murakami [1], *and J. A. Chambers* [2]

Laboratory for Advanced Brain Signal Processing,
BSI RIKEN, 2-1, Hirosawa, Wakoh-City, Saitama 351-0198, Japan
[1] Department of Electronics and Communication Engineering,
Meiji University, 1-1-1, Higashi-mita, Tama-ku, Kawasaki, Japan
[2] Centre for Digital Signal Processing, Division of Engineering,
King's College London, WC2R 2LS, U.K.

ABSTRACT

A novel stereophonic noise reduction method is proposed based upon a combination of cascaded subspace filters, with delay and advancing elements alternatively inserted between the adjacent cascading stages, and two-channel adaptive signal enhancers. Simulation results based upon real stereophonic speech contaminated by two correlated noise components show that the proposed method gives improved enhancement quality, as compared to conventional nonlinear spectral subtraction approaches, in terms of both segmental gain and cepstral distance performance indices.

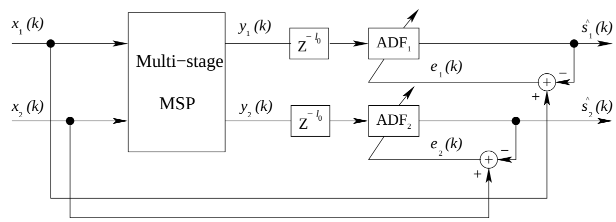


Fig. 1. Block diagram of the proposed two-channel noise reduction system.

1. INTRODUCTION

In the last few decades, noise reduction has been a topic of great interest in speech enhancement. One of the classical and most commonly used methods is based upon nonlinear spectral subtraction (NSS) [1]. In NSS methods, however, due to the block processing based approach, it is well known that such methods introduce annoying artifacts, which are often referred to as undesirable "musical tone", in the enhanced speech. Moreover, in many cases, such methods also remove some speech components in the spectra which are fundamental to the intelligibility of the speech. This is a particular problem at lower SNRs. The performance is also quite dependent on the choice of many parameters, such as, spectral subtraction floor, over-subtraction factors, or over-subtraction corner frequency parameters. The optimal choice of these parameters in practice is therefore very difficult.

Recently, in the study of blind signal processing, one of the most active potential application areas has been speech separation and several methods for blind separation/ deconvolution of speech have been developed [2] - [4]. These methods work quite well when each sensor is located close to each source. However, separation of the speech from noise is still difficult when all the sensors are located close to one dominant source but far from the others, as in cocktail party situations. This sensor configuration is typically employed in practice, for example, as in stereo conferencing systems; two microphones being placed in front of the speaker at a reasonable distance. Moreover, the existing blind separation/deconvolution methods quite often fail to work where there are more sources than sensors. On the other hand, although a number of subspace based methods have also been developed for speech enhancement [5] - [8], little attention has been paid to the extension to multichannel outputs.

In this paper, we propose a novel two-channel signal enhancement scheme using a combination of a multi-stage subspace estimation method and a two-channel adaptive signal enhancement (ASE) approach in order to tackle the aforementioned problems.

2. STEREOPHONIC NOISE REDUCTION

Fig. 1 illustrates the block diagram of the proposed two-channel noise reduction system. In the proposed method, a multi-stage moving subspace projection (M-MSP) in which each MSP acts as a filter is used in order to extract the reference signals from the primary inputs for the two adaptive signal enhancers. Moreover, between the adjacent stages, both the delay and advancing elements are alternatively inserted across the two-channels. By using the cascaded version of the MSP, it can be expected that the noise component in the signal subspace is to a certain extent decreased stage-by-stage, but some signal degradation may ensue. For the actual enhancement, two adaptive signal enhancers are used and the enhanced signal obtained from the M-MSP is used as the reference signal to the adaptive filter for each channel. The principle of this approach is that the quality of the outputs of the M-MSP will be improved by the adaptive filters.

In this paper, the following model representing a stereophonic environment is considered as the two channel observation $x_i(k)$ $(i = 1, 2)$:

$$\begin{aligned} x_1(k) &= a \cdot s_1(k) + n_1(k), \\ x_2(k) &= a \cdot s_2(k) + n_2(k), \end{aligned} \tag{1}$$

where $s_1(k)$ and $s_2(k)$ correspond respectively to the left and right channel speech signals, $n_1(k)$ and $n_2(k)$ are the additive noise components with zero-mean and are uncorrelated with the speech signals, and the constant 'a' controls the input SNR. In (1), the number of sources can be seen to be four; two stereophonic speech components and the two noise sources, but the number of the sensors is still assumed to be two ($M = 2$), as is representative of many stereophonic teleconferencing systems.

Hence, this seems to be really problematic since "There are more sources than sensors.". However, in stereophonic situations, the respective components $s_i(k)$ and $n_i(k)$ $(i = 1, 2)$ can be approximated by

$$\begin{aligned} s_i(k) &= \boldsymbol{h}_{i,1}^T(k)\boldsymbol{s}(k), \\ n_i(k) &= \boldsymbol{h}_{i,2}^T(k)\boldsymbol{n}(k), \end{aligned} \tag{2}$$

where $\boldsymbol{h}_{i,j}(k) = [h_{i,j}(0), h_{i,j}(1), \cdots, h_{i,j}(L_{s/n,j} - 1)]^T$ $(i, j \in 1, 2)$ are the impulse response vectors of the acoustic transfer functions between the signal/noise source and the microphones with lengths $L_{s/n,i}$ and $\boldsymbol{s}(k) = [s(k), s(k-1), \cdots, s(k - L_{s,i} + 1)]^T$ and $\boldsymbol{n}(k) = [n(k), n(k-1), \cdots, n(k - L_{n,i} + 1)]^T$ are respectively the speech and noise source signal vectors. Therefore, it is considered that the respective stereophonic components $s_i(k)/n_i(k)$ $(i = 1, 2)$ are generated from one speech/noise source using two (sufficiently long) filters $\boldsymbol{h}_{1,i}/\boldsymbol{h}_{2,i}$ and, in reality, the stereophonic speech components $s_i(k)$ are strongly correlated with each other, while in certain situations we may consider that the number of noise sources is also approximated to one with some amount of cross-correlation between $n_1(k)$ and $n_2(k)$.

2.1. The Subspace Projection for Noise Reduction

The subspace projection of a given signal data matrix contains information about the signal energy, the noise level, and the number of sources. By using a subspace projection , it is thus possible to divide approximately the observed noisy data into the subspaces of the signal of interest and the noise [10]. A summary of the noise reduction technique using the subspace projection is given as follows:

Let $\boldsymbol{X}$ be the available data in the form of an $L \times M$ matrix $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \cdots, \boldsymbol{x_M}]$, where the column vector $\boldsymbol{x_i}$ $(i = 1, 2, \cdots, M)$ is written as $\boldsymbol{x_i} = [x_i(0), x_i(1), \cdots, x_i(L-1)]^T$ (T: transpose). Then, the eigenvalue decomposition (EVD) of matrix $\boldsymbol{X}$ for $M < L$ is given by

$$\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{3}$$

where the matrix $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M] \in \Re^{M\times M}$ is orthogonal such that $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_M$ and $\boldsymbol{\Sigma} = diag(\sigma_1, \sigma_2, \ldots, \sigma_M) \in \Re^{M\times M}$, with eigenvalues $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M \geq 0$. The columns in $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{X}\boldsymbol{X}^T$. The eigenvalues in $\boldsymbol{\Sigma}$ contain some information about the number of signals, signal energy, and the noise level. It is well known that if the signal-to-noise ratio (SNR) is sufficiently high (e.g., see [9]), the matrix $\boldsymbol{X}$ can be decomposed as

$$\boldsymbol{X}\boldsymbol{X}^T = [\boldsymbol{V}_s\ \boldsymbol{V}_n] \begin{bmatrix} \boldsymbol{\Sigma}_s & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}_n \end{bmatrix} [\boldsymbol{V}_s\ \boldsymbol{V}_n]^T, \tag{4}$$

where $\boldsymbol{\Sigma}_s$ contains the s largest eigenvalues associated with s source signals and $\boldsymbol{\Sigma}_n$ contains $(M - s)$ eigenvalues associated with the noise. It is considered that $\boldsymbol{V_s}$ contains s eigenvectors associated with the signal part, whereas $\boldsymbol{V}_n$ contains $(M - s)$ eigenvectors associated with the noise. The subspace spanned by the columns of $\boldsymbol{V}_s$ is thus referred to as the signal subspace, whereas that spanned by the columns of $\boldsymbol{V}_n$ corresponds to the noise subspace.

Then, the signal and noise subspace are mutually orthogonal and orthonormally projecting the observed noisy data onto the signal subspace leads to noise reduction. The data matrix after the noise reduction $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_M]^T$, where $\boldsymbol{y}_i = [y_i(0), y_i(1), \ldots, y_i(L-1)]^T$, is given by

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{V}_s\boldsymbol{V}_s^T \tag{5}$$

which describes the orthonormal projection onto the signal space. This approach is quite beneficial to practical situations, since we do not need to assume/know in advance the locations of the noise sources.

2.2. Multi-stage Moving Subspace Projection

Fig. 2 illustrates a block diagram for the N-stage MSP with p-sample delay elements and p-sample advancers. As in the figure, the observed signals $x_i(k)$ are processed through multiple stages of MSP. The concept of this multi-stage structure was motivated from the works of Douglas and Cichocki [13], in which natural gradient [14] type algorithms are used in cascading form for blind decorrelation/source separation.

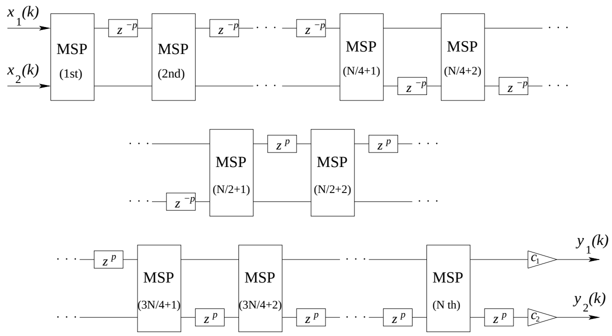


Fig. 2. Block diagram of the modified N-stage MSP with p-sample delay elements and p-sample advancers for stereophonic noise reduction.

In the analytical work [13], it is reported that the improvements in the convergence behaviors of previous decorrelation stages are compounded in subsequent decorrelation stages. A similar improvement can be obtained using M-MSP, since, in principle, MSP also performs decorrelation/separation of signals [15].

Within the proposed scheme, note that since the MSP acts as a filter, the proposed M-MSP can be viewed as an N-cascaded MSP filter. To illustrate the difference between the proposed M-MSP and the conventional frame-based operation (e.g., [10]), Fig. 3 is given. In the figure, $\boldsymbol{x}^{(j)}$ denotes

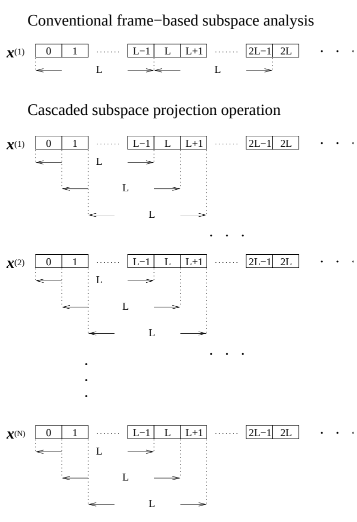


Fig. 3. Illustration of the Multi-stage MSP Operation.

a sequence of the M-channel output vectors from the j-th stage MSP operation, i.e., $\boldsymbol{x}^{(j)}(0), \boldsymbol{x}^{(j)}(1), \boldsymbol{x}^{(j)}(2), \ldots (j = 1, 2, \ldots, N)$, where $\boldsymbol{x}^{(j)}(k) = [x_1^{(j)}(k), x_2^{(j)}(k), \ldots, x_M^{(j)}(k)]$ $(k = 0, 1, 2, \ldots)$.

Then, given the previous L past samples for each channel at time instance k $(\geq L)$ and using (5), the input matrix to the j-th stage MSP $\boldsymbol{X}^{(j)}(k)$ $(L \times M)$ can be written:

$$\boldsymbol{X}^{(j)}(k) = \left[\begin{array}{c} \boldsymbol{P}\boldsymbol{X}^{(j)}(k-1)\boldsymbol{V}_s^{(j)}(k-1)\boldsymbol{V}_s^{(j)}(k-1)^T \\ \boldsymbol{x}^{(j-1)}(k) \end{array} \right],$$

$$\boldsymbol{P} = [\mathbf{0}_{(L-1)\times 1}; \boldsymbol{I}_{L-1}], (L-1 \times L) \quad (6)$$

where $\boldsymbol{V}_s^{(j)}$ denotes the signal subspace matrix obtained at the j-th stage,

$$\boldsymbol{x}^{(0)}(k) = \boldsymbol{x}(k),$$

$$\boldsymbol{X}^{(j)}(0) = \left[\begin{array}{c} \mathbf{0}_{(L-1)\times M} \\ \boldsymbol{x}^{(j-1)}(0) \end{array} \right].$$

Note that in the above the first $(L-1)$ rows of $\boldsymbol{X}^{(j)}(k)$ are obtained from the previous MSP operation, whereas the last row is taken from the data obtained from the previous stage of MSP. Then, at this point, as in Fig. 3, the new data (i.e., the data from the previous stage) $\boldsymbol{x}^{(j-1)}(k)$ remains intact and the rest $(L-1)$ data vectors, i.e., those obtained by the product $\boldsymbol{P}\boldsymbol{X}^{(j)}(k)$, will be replaced by the subsequent subspace projection operations at the j-th stage. It is thus considered that this recursive operation is similar to the concept of data-reusing [11] in which the input data at the same data point is repeatedly used for, i.e., improving the convergence rate in adaptive algorithms. Moreover, since the input data used for the j-th stage MSP is different from that at the $j-1$-th stage, it is expected that the subspace spanned by $\boldsymbol{V}_s$ can contain less noise than that obtained at the previous stage and that this approach still works when the number of the sensors M is small, as in ordinary stereophonic situations.

In addition, we can intuitively justify the effectiveness of using multi-stage MSP (M-MSP) as follows: for large noise and very limited numbers of samples (this choice must, of course, relate to the stationarity of the noise), a single stage MSP (filter) may perform only rough or approximate decomposition to both the signal and noise subspace. In other words, we are not able to ideally decompose the noisy sensor vector space into a signal subspace and its noise counterpart with a single stage MSP. In the single stage, we rather perform decomposition into a signal-plus-noise subspace and a noise subspace [8]. For this reason, applying M-MSP gradually reduces the noise level.

In [7], the Hankel matrices are exploited within the single channel SVD and a further noise reduction is achieved. It is considered that one of the keys using Hankel structure within the subspace analysis resides in the 'data-shifting', which causes variations within the subspace estimation and eventually helps the signal separation, provided that the shifting is small enough not to introduce distortion. Thus, a combined Hankel-like operation with the M-MSP filter can be useful in the stereophonic situations, since with a sufficient window length L the characteristics of the two channel signals may become similar to each other in the statistical sense of block analysis. Exploiting this principle, the M-MSP with both delay and advancing elements for stereophonic noise reduction as in Fig. 2 is thus proposed.

Eventually, the outputs after the N-th stage MSP, $y_i(k)$, are considered to be less noisy than the respective inputs $x_i(k)$ and sufficient to be used for the input signal to the signal enhancers.

2.3. Adaptive Signal Enhancement

After the extraction of each signal, adaptive signal enhancers (ASEs) are used to enhance the observed two-channel signals. Since the respective input signals to the enhancer are strongly correlated with the corresponding signals of interest, the adaptive filter functions to recover the original signal in each channel from the delayed version of the signal $y_i(k-l_0)$ $(i=1,2)$ using the observation $x_i(k)$. In the diagram in Fig. 1, the delay factor l_0 is given by

$$l_0 = \frac{L_f - 1}{2}, \tag{7}$$

where L_f is the length of each adaptive filter. The insertion of the delay factor is necessary in order to shift the enhanced signals in not only the positive but also the negative time direction by the adaptive filters. This scheme is then somewhat related to direction of arrival (DOA) estimation using adaptive filters [16]. However, in the context of stereophonic noise reduction, the role of the adaptive filters is different from the DOA. Eventually, as in Fig. 1, the enhanced signal $\hat{s}_i(k)$ is obtained simply from the filter output.

In the M-MSP described earlier, the orthonormal projection of each observation $x_i(k)$ onto the estimated signal subspace by the M-MSP leads to reduction of the noise in each channel. However, since the projection is essentially performed using only a single orthonormal vector which corresponds to the speech source, this may cause the distortion of the stereophonic image in the speech reference signals $y_1(k)$ and $y_2(k)$. In other words, the MSP is performed to recover a single speech source from the two observations $x_i(k)$.

In the proposed method, the adaptive signal enhancers are thus employed in order to compensate for the stereophonic image. Since, as in Fig. 1, the error signals $e_i(k)$ $(i=1,2)$ contain the information about the stereophonic image (because the observations $x_i(k)$ are true stereophonic signals), the adaptive filters (with sufficient filter lengths) essentially adjust the delay and the amplitude of the signal in each channel, both of which are of fundamental to recover the stereophonic sound, and therefore are considered to compensate for the stereophonic image in each channel.

3. SIMULATION STUDY

In the simulation, three stereophonically recorded speech datasets were used for the speech components $s_i(k)$. The speech was a sentence spoken in English, "*Pleasant zoos are rarely reached by efficient transportation*". Each utterance was recorded by one female and two male native speakers in a quiet room, sampled at 48kHz. Each untrained speaker was asked not to move their head from the centre of the two microphones (the distance between the two mic. is 50cm). The speech data were then normalised to have unity variance.

In order to validate the proposed scheme, we considered the case where the two noise components $n_i(k)$ $(i=1,2)$ are correlated by $\boldsymbol{h}_{i,2}(k)$, as in the stereophonic setup in (2). The noise source signal $n(k)$ was then modeled by

$$n(k) = cos(2\pi f k) \cdot v(k) \tag{8}$$

where $v(k)$ was assumed to be white. In the simulation, f was chosen to be 100(Hz), which yields a component like a car-engine noise and with this setting it is considered that some fundamental components of voiced speech are significantly contaminated. The setup above can hence be seen as a representation of an oscillatory random noise source. For the correlated noise components (as given by (2)), the impulse responses $\boldsymbol{h}_{2,i}(k)$ were assumed to be the models based upon zero-mean Gaussian random variables modulated by exponentially decaying envelopes, which represent the room impulse responses. Then, the coefficients $h_{2,i}(j)$ $(j=0,1,\cdots,L_{n,i}-1)$ are given by

$$h_{2,i}(j) = exp(-0.1j)\cdot n_{h,i}(j),\ (j=0,1,\cdots,L_{n,i}-1), \tag{9}$$

where $n_{h,i}(j)$ are given as two independent zero-mean Gaussian random variables. In the simulation, the lengths of $\boldsymbol{h}_{2,i}(k)$ were respectively truncated to 50.

3.1. Performance Measurements

For the evaluation of the enhancement quality, the objective measurement in terms of both the segmental gain in SNR and averaged cepstral distance was also considered (see, e.g. [17]), apart from the informal listening tests. The segmental gain in SNR (dB) is defined as

$$G(dB) = \frac{1}{Mp_1}\sum_{i=1}^{M}\sum_{j=1}^{p_1} 10\, log_{10}\frac{\|\boldsymbol{n}_i\|_2^2}{\|\boldsymbol{s}_i - \hat{\boldsymbol{s}}_i\|_2^2}, \tag{10}$$

where $M=2$ (stereophonic), $\boldsymbol{s}_i = [s_i(k), s_i(k+1), \cdots, s_i(k+N_f-1)]^T$, $\hat{\boldsymbol{s}}_i = [\hat{s}_i(k), \hat{s}_i(k+1), \cdots, \hat{s}_i(k+K-1)]^T$, $\boldsymbol{n}_i = [n_i(k), n_i(k+1), \cdots, n_i(k+N_f-1)]^T$, $(k=(j-1)N_f, (j-1)N_f+1, \cdots, jN_f-1,\ j=1,2,...,p_1)$ are respectively the noise-clean speech, enhanced speech, and the noise signal vector, and where $N_f(=256)$ is the number of samples in each frame and p_1 is the number of frames. The averaged cepstral distance is given by

$$d_{cep} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{p_{2,i}}\sum_{j=1}^{p_{2,i}}\sum_{k=1}^{2q}(c_{i,k}(j) - {c'}_{i,k}(j))^2 \quad (11)$$

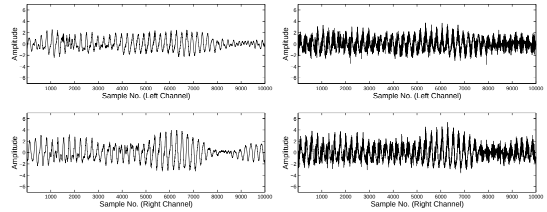
where $c_{i,k}(j)$ and ${c'}_{i,k}(j)$ are the cepstral coefficients corresponding to the clean and the enhanced signal at the left/right channel, respectively. The parameter $q(=8)$ is the order of the model, and $p_{2,i}$ $(i = 1, 2)$ is the number of frames where speech is present. The determination of speech presence was achieved by manual inspection of the clean speech signals.

3.2. Simulation Results

In Fig. 4, (a) shows the part of the noise clean speech data (sampled at 48kHz, using Speech Sample No. 2, note that the sample number for display is limited from sample no. 83001 to 93000 for a clear presentation of the results), (b) the noisy speech (assuming the input SNR = 2dB), (c) the enhanced speech by dual-mono nonlinear spectral subtraction (NSS) algorithm, (d) the enhanced speech by 4MSP with delay and advancers ($p = 1$) + Dual (i.e., two-channel) ASE (DASE), respectively. Fig. 5 shows a comparison of the segmental gain and the cepstral distance. The results shown in the figure were obtained by averaging over the three speech samples. In the figure, the performance of the five different noise reduction methods, i.e., 1) NSS, 2) 1MSP (i.e., a single-stage MSP) + DASE, 3) 4MSP + DASE, 4MSP with delay and advancers + DASE (4MSP + DA + DASE) with 4) $p = 1$ and 5) $p = 6$, is compared. In Fig. 5 (a), whilst the performance with NSS is better than the other four at lower SNRs, with 4MSP + DA + DASE (p=1), performance improvement of around 1-2dB is observed (at input SNR from -2 to 5dB) in comparison with other MSP based methods. In contrast, as in Fig. 5 (b), at lower SNRs, the case with 4MSP + DA + DASE (p=1) is superior to the other four methods. This coincided with the informal listening tests. In the listening tests, it was also confirmed that to a great extent the stereophonic image can be recovered by the proposed method.

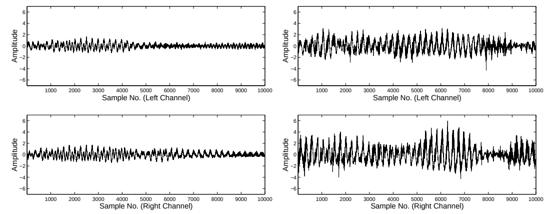
4. CONCLUSION

In this paper, a novel two-channel noise reduction method has been proposed as a combination of a multi-stage MSP and adaptive signal enhancement technique. In the proposed method, the M-MSP is used for the extraction of the signal of interest to the adaptive filter in each channel, and actual signal enhancement is performed by the adaptive approach. The proposed methods have been applied to stereophonic noise reduction, where the number of sensors is two. In the simulation study, it has been shown that the performance with the proposed methods is superior to the conventional



(a) Noise clean speech

(b) Noisy data (SNR=2dB)



(c) NSS

(d) 4MSP + DA + DASE

Fig. 4. Simulation results (speech sample No. 2) - the case where two additive correlated noise components generated from an oscillatory noise source are present.

NSS approach, where two correlated noise components generated from an oscillatory noise source are present in the respective channels. It has also been confirmed that the adaptive filters can compensate for the stereophonic image.

5. REFERENCES

[1] R. Martin, “Spectral subtraction based on minimum statistics,” *Proc. EUSIPCO-94*, pp. 1182-1185, Edinburgh, 1994.

[2] S. Amari and A. Cichocki, “Adaptive blind signal processing – neural network approaches”, *Proc. IEEE*, Vol. 86, No. 10, pp. 2026-2048, Oct. 1998.

[3] K. Torkkola, “Blind separation of delayed sources based on information maximization,” *Proc. ICASSP-96*, pp. 3509-3512, 1996.

[4] H. L. Nguyen Thi and C. Jutten, “Blind source separation for convolutive mixtures,” *Signal Processing*, Vol. 45, No. 2, pp. 209-229, 1995.

[5] F. Asano, S, Hayamizu, T. Yamada, and S. Nakamura, “Speech enhancement based on the subspace method,” *IEEE Trans. Speech, Audio Proc.*, Vol. 8, No. 5, pp. 497-507, Sept. 2000.

[6] M. Dendrinos, S. Bakamidis, and G. Carayannis, "Speech enhancement from noise: a regenerative approach," *Speech Communication*, Vol. 10, pp. 45-57, Feb. 1991.

[7] S. Doclo and M. Moonen, "A novel iterative signal enhancement algorithm for noise reduction in speech," Proc. *5th Int. Conf. Spoken Language Processing*, Sydney, Australia, pp. 1435-1438, Dec. 1998, also in internal report, K. U. Leuven, Apr. 1998.

[8] Y. Ephraim and H. L. V. Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech, Audio Proc.*, Vol. 3, No. 4, pp. 251-266, July 1995.

[9] T. Kobayashi and S. Kuriki, "Principle component elimination method for the improvement of S/N in evoked neuromagnetic field measurements," *IEEE Trans. Biomed. Eng.*, Vol. 46, pp. 951-958, Aug. 1999.

[10] P. K. Sadasivan and D. N. Dutt, "MSP based technique for noise reduction in electroencephalographic signals," *Signal Processing*, Vol. 55, No.2, pp. 179-189, 1996.

[11] J. A. Apolinario, M. L. R. de Campos, and P. S. R. Diniz, "Convergence analysis of the binormalized data-reusing LMS algorithm," *Proc. of the European Conference on Circuit Theory and Design*, Budapest, Hungary, pp. 972-977, Sept. 1997.

[12] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, Vol. AC-19, pp. 716-723, Dec. 1974.

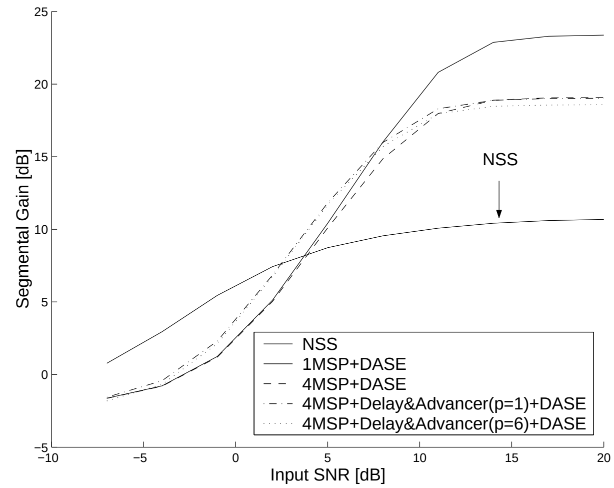
[13] S. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. on Sig. Proc.*, Vol. 45, No. 11, pp. 2829-2842, Nov. 1997.

[14] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, Vol. 10, pp. 251-276, Jan. 1998.

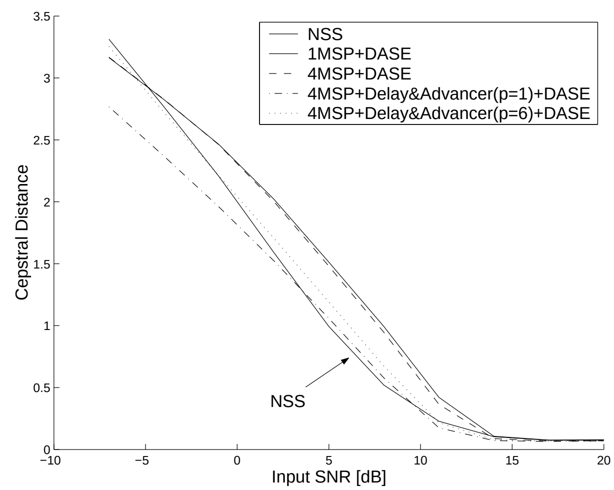
[15] D. Callaerts, J. Vandewalle, W. Sansen, and M. Moonen, "On-line algorithm for signal separation based on MSP," in *MSP and Signal Processing - Algorithms, Applications and Architectures*, Vol. 1, pp. 269-275, Elsevier Science Publishers B. V. North-Holland, 1988.

[16] C. C. Ko and C. S. Siddharth, "Rejection and tracking of an unknown broadband source in a two-element array through least square approximation of inter-element delay, " *IEEE Sig. Proc. Let.*, Vol. 6, No. 5, pp.122-125, May, 1999.

[17] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-time processing of speech signals*, Macmillan Publishing Company, 1993.



(a) Comparison of the segmental gain



(b) Comparison of the cepstral distance

Fig. 5. Performance comparison - the case with two correlated noise components generated from an oscillatory noise source.