IFAC Workshop on Adaptation and Learning in Control and
Signal Processing, and IFAC Workshop on Periodic Control
Systems, Yokohama, Japan, August 30 – September 1, 2004

# STEREOPHONIC NOISE REDUCTION BY A COMBINED NON-RECURSIVE MULTI-STAGE SLIDING SUBSPACE PROJECTION AND ADAPTIVE SIGNAL ENHANCEMENT

*Tetsuya Hoya, Toshihisa Tanaka [1], Takahiro Murakami [2], and Andrzej Cichocki*

Laboratory for Advanced Brain Signal Processing, BSI RIKEN, 2-1, Hirosawa, Wakoh-City, Saitama 351-0198, Japan
[1] Dept. of Electrical and Electronics Engi., Tokyo Univ. of Agriculture and Technology (TUAT)
2-24-16, Naka-machi, Koganei-City, Tokyo, 184-8588, Japan
[2] Dept. of Electronics and Communication Engi., Meiji Univ., 1-1-1, Higashi-mita, Tama-ku, Kawasaki-City, 214-8571, Japan

## ABSTRACT

This paper proposes a novel broad-band noise reduction technique in stereophonic situations. The method utilises a combined non-recursive multi-stage sliding subspace projection and adaptive signal enhancement. Simulation results based upon real stereophonic speech contaminated by broad-band noise components show that the proposed method yields a performance improvement in the speech enhancement quality evaluated by both the cepstral distance and segmental gain in SNR objective measurements, as well as by listening, in comparison with the conventional nonlinear spectral subtraction approach.

## 1. INTRODUCTION

In many speech applications, noise reduction is a fundamental issue and the topic has intensely studied during the last few decades. One of the commonly used methods is based upon nonlinear spectral subtraction (NSS) [1]. However, it is well-known that NSS methods introduce annoying artifacts in the enhanced speech, which are often referred to as “musical tone noise”, due to the block processing based approach. Moreover, in many cases, such methods also remove some speech components in the spectra which are fundamental to the intelligibility of the speech. The performance is also quite dependent upon the selection of many parameters, such as, spectral subtraction floor, over-subtraction factors, or over-subtraction corner frequency parameters. Finding the optimal parameter setting for the NSS methods is therefore very hard in practice.

In the study of blind signal processing, one of the most active and potential application areas has recently been speech separation and a number of 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 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]-[10], little attention has been paid to the extension to multichannel outputs.

In this paper, we propose a novel signal enhancement scheme for stereophonic situations using a combination of a non-recursive multi-stage sliding subspace projection (M-SSP) and two-channel adaptive signal enhancement (ASE) approach in order to tackle the aforementioned problems.

## 2. STEREOPHONIC NOISE REDUCTION

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 $s_2(k)$ respectively corresponds to the left and right channel speech signal arriving at the respective microphones, $n_1(k)$ and $n_2(k)$ are the noise components assumed to be zero-mean and uncorrelated with the speech signals, and the constant ‘$a$’ controls the input SNR. Then, the objective is to eliminate both the noise components $n_1(k)$ and $n_2(k)$ from the corresponding observation $x_1(k)$ and $x_2(k)$.

In reality, since the stereophonic speech components $s_i(k)$ are strongly correlated with each other, it is considered that we can exploit the property of the strong correlation for noise reduction by a subspace method.

## 3. A COMBINED NON-RECURSIVE MULTI-STAGE SLIDING SUBSPACE PROJECTION AND ADAPTIVE SIGNAL ENHANCEMENT SCHEME

Fig.1 illustrates the block diagram of the stereophonic noise reduction system. In the proposed method, a non-recursive multi-stage sliding subspace projection (M-SSP) is employed to extract the source speech, whilst the two-channel adaptive signal enhancers are used to recover the stereophonic image.

In stereophonic situations, since both the speech components $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are strongly correlated with each other, even if the rank is reduced to one for the noise reduction purpose (i.e., by taking only the eigenvector corresponding to the eigenvalue with the highest energy $\sigma_1$), it is still possible to recover $\boldsymbol{s}_i$ from $\boldsymbol{y}_i$ by using adaptive filters as the post-processors.

In other words, we can reduce the additive noise by projecting

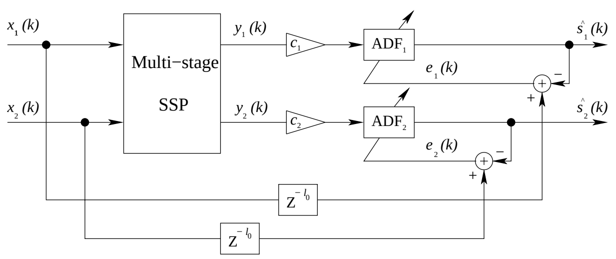



**Fig. 1**. Block diagram of the proposed stereophonic noise reduction system.

the observed signal onto the subspace around which the energy of the signal is mostly concentrated. The problem here, however, is that, since speech signals are usually non-stationary process, the correlation matrix can be time-variant. Moreover, it is considered that the subspace projection reduces the dimensionality of the signal space, e.g., a stereophonic signal pair can be reduced to a monaural signal. To solve these problems, we thus propose to use a combined subspace projection operated within a sliding-window and signal enhancers realised by adaptive filters. The former technique can estimate the correlation matrices adaptively, whereas the latter expands the degraded space into the original whole signal space again.

### 3.1. 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 [11]. 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

$$\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{2}$$

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 $\boldsymbol{X}^T\boldsymbol{X}$. 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 [12]), the matrix $\boldsymbol{X}$ can be decomposed as

$$\boldsymbol{X}^T\boldsymbol{X} = [\boldsymbol{V}_s \; \boldsymbol{V}_n] \begin{bmatrix} \boldsymbol{\Sigma}_s & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma}_n \end{bmatrix} [\boldsymbol{V}_s \; \boldsymbol{V}_n]^T, \tag{3}$$

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

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{V}_s\boldsymbol{V}_s^T \tag{4}$$

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.

### 3.2. The Non-Recursive Multi-Stage Sliding Subspace Projection

In many applications, the subspace projection above is employed in a batch mode. Here, we instead propose an on-line batch algorithm for adaptively estimating the subspaces which is operated in a cascading form.

Fig.2 shows a block diagram for the $N$-stage SSP. As in the

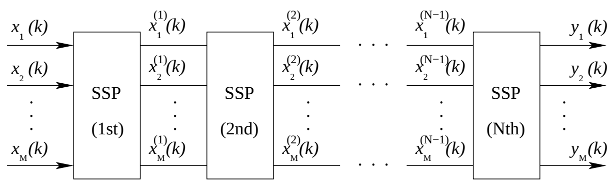



**Fig. 2**. Block diagram of the multi-stage (up to $N$-th stage) SSP for stereophonic noise reduction.

figure, the observed signals $x_i(k)$ are processed through multiple stages of SSP. The concept of this multi-stage structure was motivated from the work of Douglas and Cichocki [13], in which natural gradient [2] type algorithms are used in a cascading form for blind decorrelation/source separation. In [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-SSP, since, in principle, SSP also performs decorrelation / separation of signals [14].

Within the proposed scheme, note that since the SSP acts as a sliding-window noise reduction block, the proposed M-SSP can be viewed as an $N$-cascaded SSP processing block. To illustrate the difference between the proposed M-SSP and the conventional frame-based operation (e.g., [11]), Fig. 3 is given. In the figure, $\boldsymbol{x}^{(j)}$ denotes a sequence of the $M$-channel output vectors from the $j$-th stage SSP operation, i.e., $\boldsymbol{x}^{(j)}(0), \boldsymbol{x}^{(j)}(1), \boldsymbol{x}^{(j)}(2), \ldots$ $(j = 1, 2, \ldots, N)$, where $\boldsymbol{x}^{(j)}(k) = [x_1^{(j)}(k), x_2^{(j)}(k), \ldots, x_M^{(j)}(k)]$ $(k = 0, 1, 2, \ldots)$. As in the figure, the SSP operation is applied to a small fraction of data (i.e. the sequence of $L$ samples) using the original input at time instance $k$ in each stage and outputs only the signal counterpart for the next stage. This operation is repeated at the subsequent time instances $k + 1, k + 2, \ldots$, and thus is the name 'sliding'.

Then, given the previous $L$ past samples for each channel at

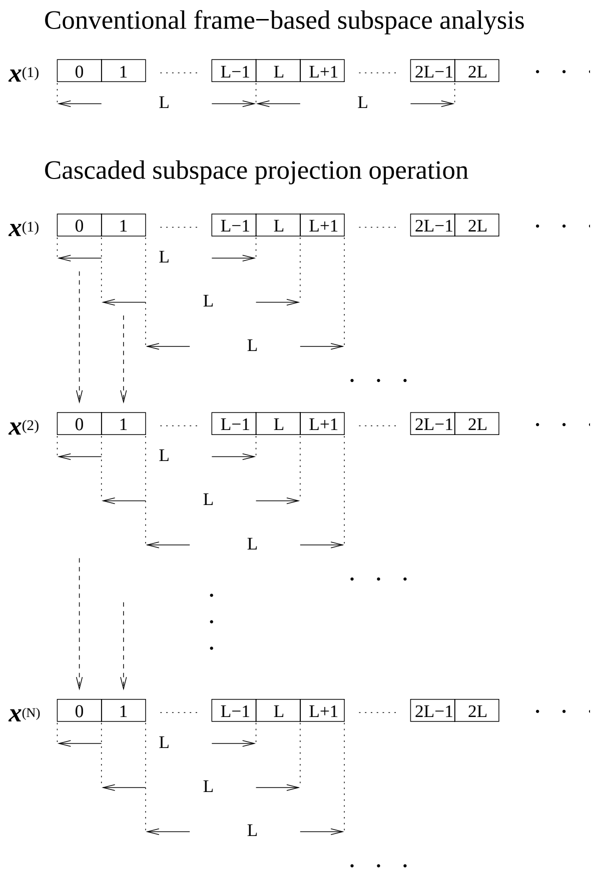



**Fig. 3**. Illustration of the multi-stage SSP Operation.

time instance $k$ ($\geq L$) and using (4), the input matrix to the $j$-th stage SSP $\boldsymbol{X}^{(j)}(k)$ ($L \times M$) can be written:

$$\boldsymbol{X}^{(j)}(k) = \boldsymbol{X}^{(j-1)}(k)\boldsymbol{V}_s^{(j-1)}(k)\boldsymbol{V}_s^{(j-1)}(k)^T \qquad (5)$$

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

$$\begin{aligned} \boldsymbol{x}^{(0)}(k) &= \boldsymbol{x}(k), \\ \boldsymbol{X}^{(j)}(0) &= \begin{bmatrix} \boldsymbol{0}_{(L-1)\times M} \\ \boldsymbol{x}^{(j-1)}(0) \end{bmatrix}. \end{aligned}$$

Note that the above new non-recursive sliding operation is considered to yield rather a *smooth* estimate of the signal subspace from the data, in comparison with the original approach in [15], in that the estimate is not based upon the data matrix $X(k)$ modified *after* recursive (or repetitive) EVD operations.

Then, the first row of the new input matrix $\boldsymbol{X}^{(j)}(k)$ given in (5) corresponds to the $M$-channel signals after the $j$-th stage SSP operation $\boldsymbol{x}^{(j)}(k) = [x_1^{(j)}(k), x_2^{(j)}(k), \ldots, x_M^{(j)}(k)]^T$:

$$\begin{aligned} \boldsymbol{x}^{(j)}(k) &= \boldsymbol{X}^{(j)}(k)^T \boldsymbol{q}\,, \\ \boldsymbol{q} &= [1, 0, 0, \ldots, 0]^T \ (L \times 1)\,. \end{aligned} \qquad (6)$$

Since the input data used for the $j$-th stage SSP 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 SSP (M-SSP) 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 SSP 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 SSP. In the single stage, we rather perform decomposition into a signal-plus-noise subspace and a noise subspace [6]. For this reason, applying M-SSP gradually reduces the noise level. Eventually, the outputs after the $N$-th stage SSP, $y_i(k)$, 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.

In the M-SSP described above, the orthonormal projection of each observation $x_i(k)$ onto the estimated signal subspace by the M-SSP 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 extracted speech signals $y_1(k)$ and $y_2(k)$. In other words, the SSP is performed only to recover the single speech source from the two observations $x_i(k)$.

### 3.3. Two-Channel Adaptive Signal Enhancement

In the proposed method, a two-channel adaptive signal enhancer (ASE) is then 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 image, and therefore are considered to compensate for the stereophonic image in each channel.

Moreover, since the respective input signals to the signal enhancer are strongly correlated with the corresponding signals of interest, the $i$-th adaptive filter functions to recover the stereophonic information in each channel from the signal $y_i(k)$ using the delayed version of the reference signal $x_i(k - l_0)$. In the diagram in Fig. 1, the delay factor $l_0$ is given by

$$l_0 = \frac{L_f - 1}{2}, \qquad (7)$$

where $L_f$ is the length of each adaptive filter. The insertion of delay factor is necessary in order to shift the centre lag of the reference 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] and similar to adaptive line enhancers (ALE, see e.g., [17]). However, unlike an ordinary ALE, the reference signal in each channel is not taken from the original input but the observation $x_i(k)$, as in Fig. 1. Moreover, in the context of stereophonic noise reduction, the role of the adaptive filters is different from the DOA, as described above.

In addition, in Fig. 1, $c_i$ are arbitrarily chosen constants and used to adjust the scaling of the corresponding input signals to the adaptive filters. These scaling factors are normally necessary, since the choice will affect the initial tracking ability of the adaptive algorithms in terms of stereophonic compensation and may be determined *a priori* with keeping a good-trade off between the initial tracking performance and the signal distortion. Finally, as in Fig. 1, the enhanced signal $\hat{s}_i(k)$ is obtained simply from the filter output.

## 4. SIMULATION STUDY

In the simulation, three stereophonically recorded speech data were used for the speech components $s_i(k)$. For the speech data, the sentence was "Pleasant zoos are rarely reached by efficient transportation" in English. Each utterance was recorded by one female and two male speakers in a non-reverberant room, sampled originally at 48(kHz) and down-sampled to 8(kHz). Each untrained speaker was asked not to move their head from the centre of the two microphones (the distance between the two mics. is 50(cm)). The speech data were then normalised to have unity variance.

In order to validate the proposed scheme, we tested the following two cases for the noise components $n_i(k)$ ($i = 1, 2$): the two noise components are 1) synthetically generated i.i.d. sequences, and 2) the real stereophonic fan noise data recorded in an ordinary room (the shape is rectangular and its size is 200(cm) wide, 350(cm) long, and 230(cm) tall) of a house near the kitchen system, without any sound shielding equipped. The two i.i.d. noise components assumed were the signals generated from 1) uniform distribution (using MATLAB function, *rand()* ) shifted to lie within the range from $-0.5$ to 0.5, and 2) Normal distribution (using MATLAB function, *randn()*).

For the SSP, the length of the analysis matrix is fixed to 32. In a separate simulation study, we confirmed that this is a reasonable choice for giving a good trade-off in terms of the performance and the computational complexity, since the SSP (i.e. the EVD) operation is the most computationally demanding part within the proposed scheme (e.g. for the actual computation, applying the Cholesky's decomposition requires $O(n^3/3)$).

For the ASE, the standard normalised-LMS algorithm (e.g., see [17]) was used to adjust the filter coefficients in the dual adaptive signal enhancer (DASE, i.e., the case where $M = 2$ in Fig. 1). For each adaptive filter, the learning constant was chosen as 0.5. The filter length was fixed to 51, which allows approximately 3(ms) of delay in left/right channel, and, within this range, neither precedence effect (or, alternatively, Haas effect) nor echo effect will occur [18]. Moreover, the scalar constants $c_i$ were empirically fixed to 0.1 for both the left and right channels, which was empirically found to moderately suppress the distortion and satisfied a good trade-off between a reasonable stereophonic image compensation and signal distortion.

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. [19]), apart from the informal listening tests.

### 4.1. Simulation Results

In Fig. 4, (a) shows the clean speech data, (b) noisy speech (assumed the input SNR=0dB), (c) the enhanced speech by dual-mono NSS (with the same parameter setting as for NSS1 in Fig. 5), (d) the speech by only 1SSP (i.e., single stage), (e) only 8SSP (eight-stages), and (f) 8SSP+DASE. In the figure, it is clear observed that some voiced speech parts are eliminated or greatly changed in shape for the enhanced speech by NSS, whilst the overall shape is preserved by 8SSP+DASE. This coincided with the informal listening tests; the enhanced speech by 8SSP+DASE does not have such artifacts or distortion as the NSS, and the stereophonic image was recovered to a great extent, compared to the case of applying only 8SSP. Moreover, it is clear seen that the performance using 8SSP is much better than 1SSP.

In contrast, Figs. 5 and 6 show the comparisons of both the segmental gain in SNR (e.g., see [20] ) and cepstral distance obtained by averaging over the first three samples (i.e. the samples uttered in English) for the i.i.d. noise and real fan noise cases, respectively. In both the figures, three different parameter settings were attempted for the NSS (i.e., indicated as NSS1, NSS2, and NSS3) in order to see how the performance varies. As in Fig. 5 (a), at lower SNRs, the performance with NSS is better than that of M-SSP+DASE, while at higher SNRs the M-SSP+DASE algorithm is the best for the i.i.d. noise case. However, at lower SNRs, as in Fig. 5 (b), the performance in terms of cepstral distance with NSS (for all the three parameter settings) is poorest amongst all. Similar to the tendency as observed in Fig. 5, M-SSP+DASE is the best amongst all at higher SNRs for the real fan noise case, as in Fig. 6 (a), whilst, in Fig. 6 (b), the approach with 8SSP only always yields the best. However, the latter contradicts the result of the informal listening tests; the enhanced speech obtained from (8SSP only) lacks in the stereophonic image, whereas the image is mostly recovered in 8SSP+DASE. As is often the case in general speech enhancement, this indicates that the objective measurement of cepstral distance can not sufficiently describe this difference, whilst the segmental gain can do to a certain extent. Then, to investigate the stereophonic image recovery, the scatter plots in Fig. 7 are given. As in the figure, the scatter plot of the two-channel enhanced speech signals by 8SSP+DASE approaches to that of the clean speech.

## 5. CONCLUSION

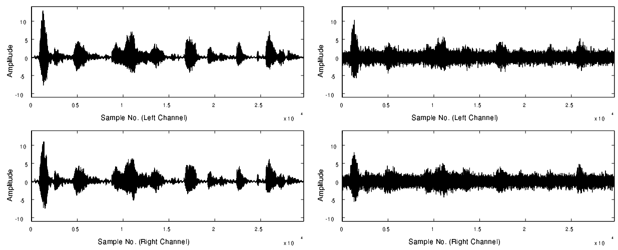
In this paper, a novel stereophonic noise reduction approach by a combined non-recursive multi-stage sliding subspace projection (M-SSP) and adaptive signal enhancement has been proposed. In the proposed method, the M-SSP is used for the extraction of speech source, whereas the dual adaptive signal enhancers are employed as the post-processors to recover the stereophonic image. In the simulation study, it has been shown that the proposed combined approach yields a superior performance to that of the conventional NSS method, in terms of both the objective measurements and the informal listening tests.

## 6. REFERENCES

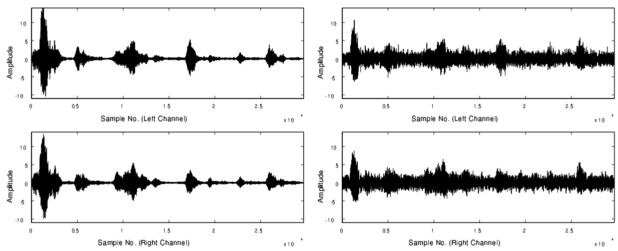

[1] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech, Audio Proc.*, vol. 9, no. 5, pp. 504-512, July 2001.

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

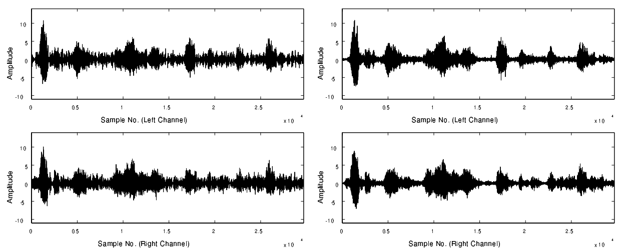
[3] K. Torkkola, "Blind separation of delayed sources based on information maximization," *Proc. ICASSP-96*, pp. 3509-3512, 1996.



(a) Noise clean speech (b) Noisy data (SNR=0dB)

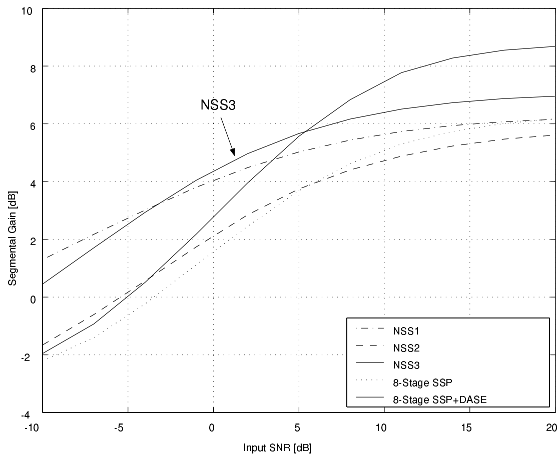


(c) NSS (d) 1SSP only

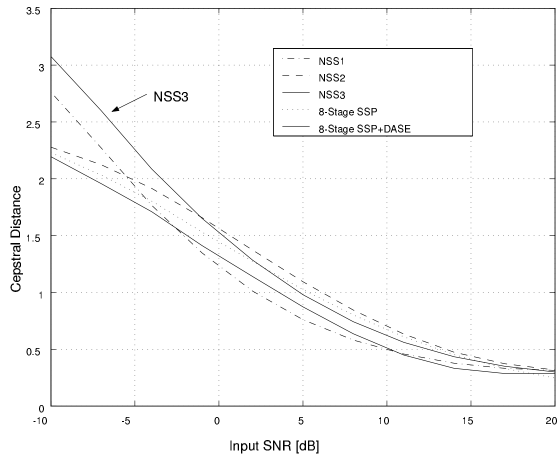


(e) 8SSP only (f) 8SSP+DASE

**Fig. 4**. Simulation results - the case where two additive i.i.d. noise components generated from Normal distributions are present.




(a) Comparison of the segmental gain




(b) Comparison of the cepstral distance

**Fig. 5**. Performance comparison - the case with two i.i.d. noise components generated from Normal distributions.

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

[6] 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.

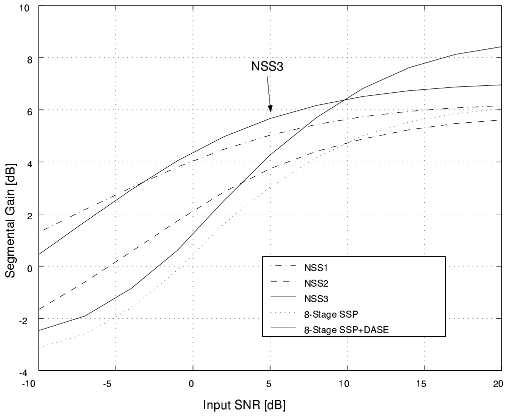
[7] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sorensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech, Audio Proc.*, Vol. 3, pp. 439-448, Nov. 1995.

[8] 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.

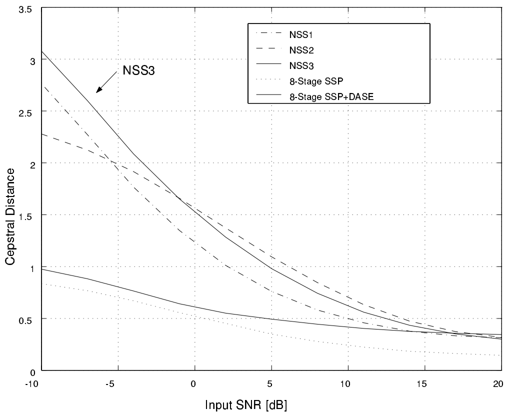
[9] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Sig. Proc.*, vol. 50, no. 9, pp. 2230-2244, Sept. 2002.

[10] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing - Applications to Real-World Problems* (eds. J. Benesty and Y. Huang), pp. 155-194, Springer, Berlin, Germany, 2003.

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

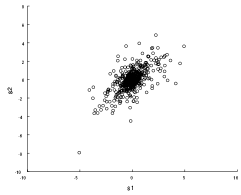



(a) Comparison of the segmental gain

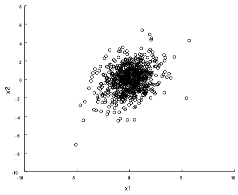



(b) Comparison of the cepstral distance

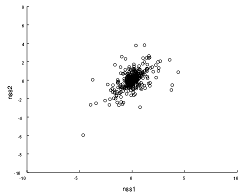
**Fig. 6**. Performance comparison - the case with two real fan noise components.



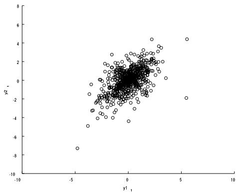
(a) Clean speech.



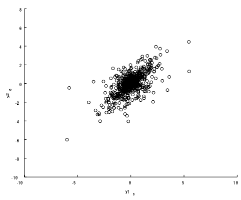
(b) Noisy data (SNR=0dB).



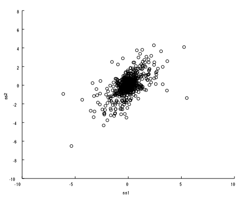
(c) Enhanced speech by NSS.



(d) Enhanced speech by 1SSP.



(e) Enhanced speech by 8SSP.



(f) Enhanced speech by 8SSP+DASE.

**Fig. 7**. The scatter plots - using speech sample No. 1 (with the two additive i.i.d noise components, at input SNR=0(dB))


[12] 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.

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

[15] T. Hoya, T. Tanaka, A. Cichocki, T. Murakami, and J. A. Chambers, "Stereophonic noise reduction using a combined sliding subspace projection and adaptive signal enhancement," to appear in the *IEEE Trans. Speech and Audio Proc.* in 2005.

[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] S. Haykin, *Adaptive Filter Theory*, 2nd Ed., Englewood Cliffs, NJ: Prentice-Hall, 1991.

[18] C. Hugonnet and P. Walder, *Stereophonic sound recording - theory and practice*, John Wiley & Sons, NY, 1998.

[19] J. R. Deller, Jr., J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.

[20] R. Le Bouquin-Jennes, A. Akbari Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator", *IEEE Trans. on Speech, Audio Proc.*, Vol. 5, No. 5, pp. 484-487, Sept. 1997.