J. Biomed. Eng. Res.
Vol.26, No.3, 177-184, 2005

# A New Hearing Aid Algorithm for Speech Discrimination using ICA and Multi-band Loudness Compensation

Sangmin Lee[1], Jong Ho Won[2], Hyung Min Park[3], Sung Hwa Hong[4], In Young Kim[2], Sun I. Kim[2]

[1]Division of Bionics & Bioinformatics, Chonbuk National University
[2]Dept. of Biomedical Engineering, Hanyang University
[3]Dept. of Electrical and Computer Engineering and School of Computer Science, Carnegie Mellon University
[4]Dept. of Otolaryngology-Head and Neck Surgery, Sungkyunkwan University School of medicine,Samsung Medical Center
*(Received July 12, 2004. Accepted February 15, 2005)*

**Abstract:** In this paper, we proposed a new hearing aid algorithm to improve SNR(signal to noise ratio) of noisy speech signal and speech perception. The proposed hearing aid algorithm is a multi-band loudness compensation based independent component analysis (ICA). The proposed algorithm was compared with a conventional spectral subtraction algorithm on behind-the-ear type hearing aid. The proposed algorithm successfully separated a target speech signal from background noise and from a mixture of the speech signals. The algorithms were compared each other by means of SNR. The average improvement of SNR by ICA based algorithm was 16.64dB, whereas spectral subtraction algorithm was 8.67dB. From the clinical tests, we concluded that our proposed algorithm would help hearing aid user to hear clearly a target speech in noisy conditions.

**Key words:** Hearing aid, Independent component analysis (ICA), Multi-band compensation, Speech discrimination.

## INTRODUCTION

The objective for hearing aids is to modify the acoustic signal to produce the best sound for the hearing impaired. In many cases, hearing impairment is the result of damage to the cells of the cochlea, such as inner or outer hair cells. Because of cell function loss, sharply tuned neural response characteristics become wider and sensitivity characteristics become worse. Conventional hearing aids usually compensate for this loss by compressing a sound signal into the reduced dynamic range of the impaired ear. In addition, some hearing aid algorithms try to improve the signal-to-noise ratio (SNR) in noisy environments by using adaptive filtering or spectral subtraction. The spectral subtraction method is a well-known noise reduction technique and accepted as a good method for noise reduction. However, real-world noise is usually dynamic and does not uniformly affect the speech signal over the spectrum. Therefore, there are many hearing impaired people who cannot benefit from this simple approach, and besides, the hearing aid user usually cannot hear a target speech signal if there are a number of people talking. Kompis et al [1] reported that hearing aid users commonly complain of difficulty in understanding speech in the presence of background noise or in the midst of a din of conversation. It is clear that to satisfy the needs of this group of people new and better methods are required.

Independent component analysis (ICA) is a computational and statistical technique for revealing hidden factors that underlie sets of random variables, measurements or signals. Recently, ICA has received considerable attention because of its potential applications in signal processing, such as in speech recognition systems, speech signal processing and biomedical signal processing. Bell and Sejnowski [2] have shown that information maximization theory could be used to separate many independent sources. Anemüller [3] extended this approach to a speech signal separation in various situations.

The normal ear can distinguish what one speaker is saying when another speaker is talking at the same time. This is easier when the two speakers stand apart rather than when they stand

This study was supported by the grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Korea (02-PJ3-PG6-EV10-0001).
**Corresponding Author:** In Young Kim, M.D., Ph.D.
Dept. of Biomedical Engineering, Coll. of Medicine, Hanyang University Sungdong P.O. Box 55, Seoul, KOREA ,133-605
Tel. 822-2291-1713, Fax. 822-2296-5943,
E-mail. iykim@hanyang.ac.kr

together. The ability to follow what one speaker says in the presence of the chatter of many others is called the "cocktail party effect". This is easily illustrated to an audience by recording the voice of a single speaker reading different passage of text in succession. However, the impaired ear usually does not perform well in distinguishing a voice of concern and interest compared with the normal ear because the voice of concern and interest is masked by loud background noise and the hearing impaired person usually has a narrow dynamic range. Therefore, if a hearing aid could produce a voice signal clearly from among a din of conversations and background noise, it would be most helpful for the hearing impaired.

In this paper, we developed a hearing aid algorithm to enhance speech discrimination in the presence of background noise or interruption by another voice. Using ICA technology, this algorithm can separate the target speech signal and compensate for loss of hearing sensitivity using multi-band compensation. We conducted a sequence of experiments to verify the performance of the proposed algorithm.

# METHODS

The proposed method consists of two processing steps. The first is to separate the speech signal from the mixture of speech-noise signals using ICA technology. The second step is to apply the multi-band loudness compensation algorithm to the separated signals. The block diagram is shown in Fig. 1.

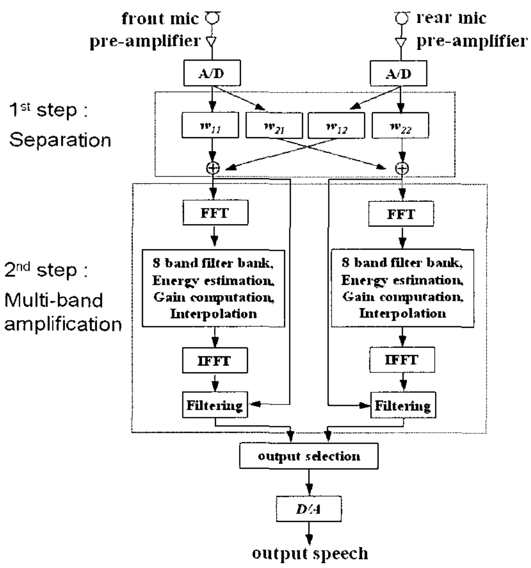



**Fig. 1.** Block diagram of the proposed method.

## Speech Signal Separation Using ICA Technology

ICA technology is used to separate the target speech signal from mixed sounds. We focus here on how ICA separates the source signal. If there are $x_1(t), x_2(t), \cdots, x_n(t)$ random variables observed by sensors, the basic premise of ICA is that these $x_n(t)$ are a mixture of independent components given by

$$x(t) = \mathbf{A} \cdot s(t)\,, \qquad (1)$$

where A is a mixing matrix and $s(t)$ are independent source components. The purpose of blind signal separation, such as ICA, is to estimate $s(t)$ using only the observed data $x(t)$. To solve this problem, the mixing matrix, $\mathbf{A}$ is estimated statistically and mathematically and then the inverse of $\mathbf{A}$, the un-mixing or de-mixing matrix, $\mathbf{W}$, is calculated. This un-mixing matrix is

$$\mathbf{W} = \mathbf{A}^{-1}\,. \qquad (2)$$

The independent components are then obtained using equation (3),

$$s(t) = \mathbf{W}x(t)\,. \qquad (3)$$

The basic idea of ICA is summarized in Fig. 2.

The source data are whitened as a pre-processing step to make the data compatible with the ICA algorithm. White data have zero mean and unit variance.

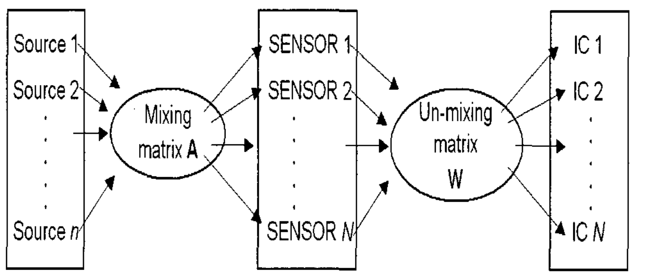



**Fig. 2.** The basic concept of independent component analysis. N source signals (S) are mixed by an unknown mixing matrix A. A set of N sensors records mixed sources and then the detected signals are separated into independent components N by ICA.

We had to decide on the method to be used to estimate the mixing matrix, **A**. There are many estimation methods, such as maximization of non-gaussianity, maximum likelihood estimation, information maximization, and minimization of mutual information. In our research, we estimated the mixing matrix based on the information maximization theory as suggested by Torkkola [4] to solve the real source separation problems.

The sound signals from the front and the rear microphones of the behind-the-ear (BTE) hearing aid are passed through an A/D-converter with a 12k [sps] sampling rate. Speech ($s_1$) and noise ($s_2$) are mixed together when they enter the front and rear microphones, and a delay time ($d$) exists between $x_1$ and $x_2$ due to the distance between the two microphones. The mixture signals entering the front and the rear microphones can be described by equation (4),

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t-d) \\ x_2(t) &= a_{22}s_2(t) + a_{21}s_1(t-d) \end{aligned} \tag{4}$$

where $a_{ij}$ denotes the mixing matrix. By using a causal FIR filter $w_{ij}$, the mixed signals can be separated as follows:

$$\begin{aligned} u_1(t) &= \sum_{d=0}^{L_{11}} w_{11}(d)x_1(t-d) + \sum_{d=0}^{L_{12}} w_{12}(d)x_2(t-d) \\ u_2(t) &= \sum_{d=0}^{L_{22}} w_{22}(d)x_2(t-d) + \sum_{d=0}^{L_{21}} w_{21}(d)x_1(t-d) \end{aligned} \tag{5}$$

The determinant of the Jacobian of the network is

$$\begin{aligned} |J| &= \frac{\partial y_1}{\partial x_1}\frac{\partial y_2}{\partial x_2} - \frac{\partial y_1}{\partial x_2}\frac{\partial y_2}{\partial x_1} = y_1' y_2' D \\ \log|J| &= \log y_1' + \log y_2' + \log D \end{aligned} \tag{6}$$

where

$$D = \frac{\partial u_1}{\partial x_1}\frac{\partial u_2}{\partial x_2} - \frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} = w_{11}(0)w_{22}(0) - w_{12}(0)w_{21}(0)$$

$$y_1' = \frac{\partial y_1}{\partial u_1}$$

$$y_2' = \frac{\partial y_2}{\partial u_2}$$

The adaptation rule for each parameter of the network can be determined by computing the gradient of $\log|J|$ with respect to that parameter.

$$\Delta w_{11}(0) \propto \frac{\partial \log|J|}{\partial w_{11}(0)} = \frac{1}{y_1'}\frac{\partial y_1'}{\partial w_{11}(0)} + \frac{1}{y_2'}\frac{\partial y_2'}{\partial w_{11}(0)} + \frac{1}{D}\frac{\partial D}{\partial w_{11}(0)} \tag{7}$$

where

$$\begin{aligned} \frac{\partial y_1'}{\partial w_{11}(0)} &= \frac{\partial y_1'}{\partial y_1}\frac{\partial y_1}{\partial u_1}\frac{\partial u_1}{\partial w_{11}(0)} = -\varphi(u_1)y_1'x_1 \\ \frac{\partial y_2'}{\partial w_{11}(0)} &= \frac{\partial y_2'}{\partial y_2}\frac{\partial y_2}{\partial u_2}\frac{\partial u_2}{\partial w_{11}(0)} = 0, \frac{\partial D}{\partial w_{11}(0)} = w_{22}(0) \end{aligned}$$

Except for the case of the zero delay, the adaptation rule for each weight is:

$$\Delta w_{ij}(d) \propto -\varphi(u_i(t))x_j(t-d)\,. \tag{8}$$

## Multi-band Loudness Compensation

The hearing impaired usually have a reduced dynamic range because of the shift in auditory threshold. Another feature of hearing impairment is the non-linear relationship of the loudness, called recruitment [5]. As a result, for many years various dynamic range compression strategies, or multi-channel compression schemes, have been widely used for hearing aid users [6]. These methods have to be very delicate because the hearing loss tends to vary with frequency. A schematic diagram of the multi-band loudness compensation algorithm adopted in this study is shown in Fig. 1 and a detailed version in Fig. 3.

The separated signal resulting from the ICA processing is divided into eight frequency bands after spectrum estimation. After spectrum estimation the energy of the signal is estimated for each frequency band to compute a time-varying correction factor. We have to measure the acoustic energy to account for the perception of loudness of the complex sounds; the energy is estimated based on eight frequency bands, as suggested by Park [7].

The input signal is segmented into 128 sample window samples, and a Hamming window is applied. Then, fast Fourier transform (FFT) of each block of time domain signals gives the input energy for each frequency band. With personal hearing loss data provided, we can control the gain to restore the loudness perception of a hearing impaired person to a normal hearing level. In other words, a loudness correction table (LCT) is defined for each frequency band, based on the input-output curves of a normal and of an impaired listener, for gain control. After obtaining gain factors for the eight frequency bands, samples of the desired filter frequency function, $H(n)$, are obtained by interpolation. The eight frequency bands are B0: 0~350Hz, B1: 350~700Hz, B2: 700~1200Hz, B3: 1200~1750Hz, B4: 1750~2450Hz, B5:B6: 3500~4900Hz and B7: 4900~6000Hz. The impulse response coefficients, $h(m)$, are then computed by inverse fast Fourier transform (IFFT). Finally, the output sequence is determined by convolving $h(m)$ with the input sequence. The procedure is referred to as the frequency sampling method.

$$h(m)=\frac{1}{N}\{H(0)+2\sum_{n=1}^{\frac{N}{2}-1}H(n)\cos(2\pi(m+1+\frac{N}{2})n)\},0\leq n\leq\frac{N}{2} \tag{9}$$

where N=128.

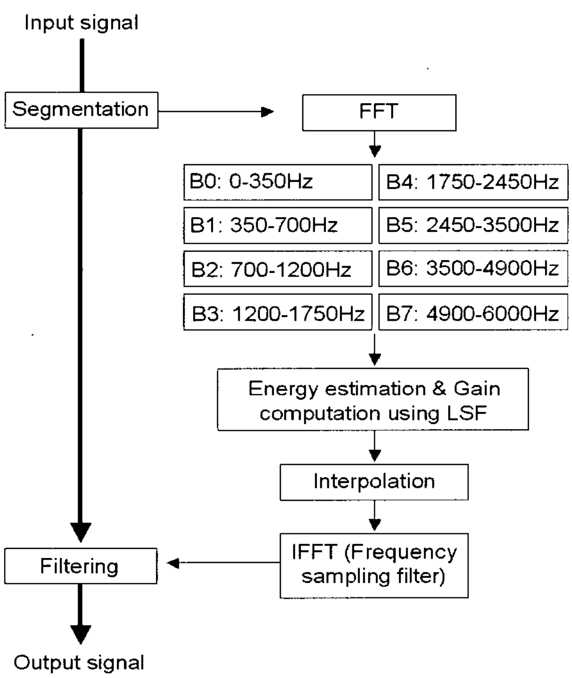



**Fig. 3.** Block diagram of the multi-band loudness compensation algorithm.

## EXPERIMENTS

We performed two kinds of experiment. One is the separation of a target speech signal from background noise and the other is the separation of two speakers. The processing was implemented by our C++ program.

### Experiment#1: Separation of Target Speech Signal from Background Noise

We recorded one speaker saying one or two syllables while loud noise was playing in the background. The speech source was located 1m in front of and the noise source 1m behind the hearing aid. The experimental arrangement is shown Fig. 4. For speech signals, we used one syllable from a male voice and two syllables from a female voice, as would be employed in a hearing examination by an audiologist, and for noise signals we used a car, babble, and factory noises from the Noisex-92 database [8]. The LCT is based on an audiogram of the hearing impaired sensory neural hearing loss by aging.

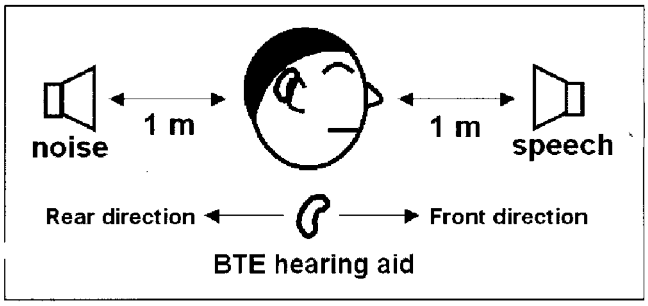



**Fig. 4.** Experimental setup of a speech/noise experiment.

### Experiment#2: Separation of the Speech Signals of Two Speakers

The voice of a man was located 1m in front of and the voice of woman was 1m behind the hearing aid, as shown in Fig. 5. We recorded a male saying "We experiment on hearing aids using ICA at the Hanyang University" in Korean and a female saying simultaneously "We study and perform experiments on hearing, hearing aids, and bio signals" in Korean. These sentences include a single vowel, diphthong, explosive sound, nasal sound, affricate, and stop sound. The LCT is the same as for the experiment in section 3.1.

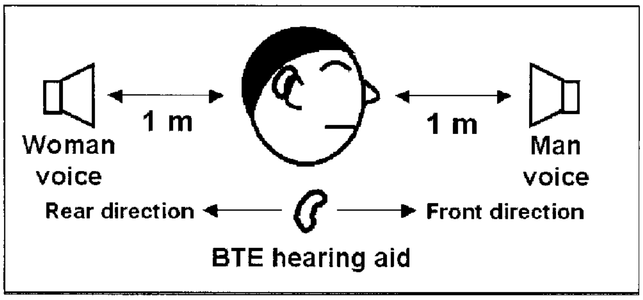



**Fig. 5.** Experimental setup of a speech/speech experiment.

### Experiment#3: Clinical Tests

The number of computation of ICA is too many to make real time processing with existing DSP chip of hearing aids at present. So we performed clinical tests based on the result of the simulation experiments. These clinical tests were performed in the laboratory of hearing test in Samsung Medical Center, Seoul, Korea. Subject was seated in the center of sound-proof room (Industrial Acoustics Company. Inc) and speaker for target speech signal was located 1m in front of the chair and another for babble noise signal was 1m behind the chair.

**Table 1.** Information about the subjects.

| Group | Subject Number | Age | Sex | Hearing threshold (without hearing aid) [dB] |
|---|---|---|---|---|
| Normal listener | 1 | 26 | male | Normal |
| | 2 | 26 | male | Normal |
| | 3 | 30 | male | Normal |
| | 4 | 31 | male | Normal |
| | 5 | 26 | male | Normal |
| | 6 | 25 | female | Normal |
| | 7 | 24 | female | Normal |
| | 8 | 26 | female | Normal |
| | 9 | 24 | female | Normal |
| | 10 | 31 | female | Normal |
| | 11 | 30 | female | Normal |
| | Average | 27 | | Normal |
| Hearing impaired listener | 1 | 76 | female | 25 (56) |
| | 2 | 66 | female | 25 (66) |
| | 3 | 72 | male | 25 (50) |
| | 4 | 71 | male | 25 (59) |
| | 5 | 87 | male | 25 (63) |
| | Average | 74 | | 25 (59) |

Each subject listened the target speech signal and noise simultaneously and repeated what he or she listened the target speech signal. The level of the target speech signal was most comfortable level (MCL) of each subject according to the result of his or her pure tone audiometry. The level of the noise was varied to make different SNR environments. We made the percentage of words correctly repeated in the each SNR environment with no signal processing, with spectral subtraction method, and the proposed method. Our subjects of normal hearing persons were eleven healthy young adults (ages 25~32, five males, six females) with no history of language problem, and five hearing impaired persons in ages of 65 ~ 85, three males and two females. Hearing impaired subjects wore own hearing aids bilaterally. The results of their pure tone audiometry are shown in table 1. The LCT for each subject was used according to the result of each subject's hearing test.

## RESULTS AND DISCUSSION

### Results of Experiment #1

Figure 6 shows the source signals, mixed signals and separated signals in time domain. Figure 6 (a) is original speech signals of one syllable from a male voice, (b) is car noise, (c) is recorded signal of the front microphone, (d) is recorded signal of the rearmicrophone, (e) is output speech signal extracted by the proposed method, and (f) is output speech signal extracted by the spectral subtraction method. For each trace, the SNR was computed as given in equation (10). For each trace, 4000 data samples, centered over the syllable of the speech signal, were used to compute an estimate of signal plus noise variance $\sigma_{2s+n}$, and we computed the final signal-plus-noise variance by averaging each five syllables' $\sigma_{2s+n}$. The 2000 data samples of each trace that have no speech signal were used to compute an estimate of noise variance $\sigma_{2n}$ alone. The SNR of each trace was then computed as

$$SNR = 10\log_{10}\left(\frac{\sigma_{s+n}^2 - \sigma_n^2}{\sigma_n^2}\right) \quad (10)$$

In Fig. 6, the SNR of the original speech signal (a) was 52.34 dB, (c) was 3.20dB, and (d) was2.45dB, which is lower than the signal of the front microphone. The SNR of (e) was 21.38 dB and

(f) was 12.50 dB. For the output of the proposed method, the improvement of SNR was 18.18 dB, whereas for the spectral subtraction method it was 9.30 dB. In other words, the proposed method could separate the target speech signal in the background noise by 9.88 dB more than the spectral subtraction method. We can see this result from a different standpoint, that of spectrogram analysis (Fig. 7).

Each syllable has a different frequency characteristic, as shown in (a) of Fig. 7. Car noise signal (b) has more low frequency components than high frequency components. (c) and (d) are the recorded signal of the front and the rear microphones, respectively. The speech signal was severely influenced by the noise signal, particularly in the low frequency bands. Fig. 7 (e) is output speech signal extracted by the proposed method and (f) is output speech signal extracted by the spectral subtraction method. The proposed method somewhat removes noise influence as shown in (e). We confirmed that frequency characteristics of each syllable was recovered well, especially in the low frequency bands.

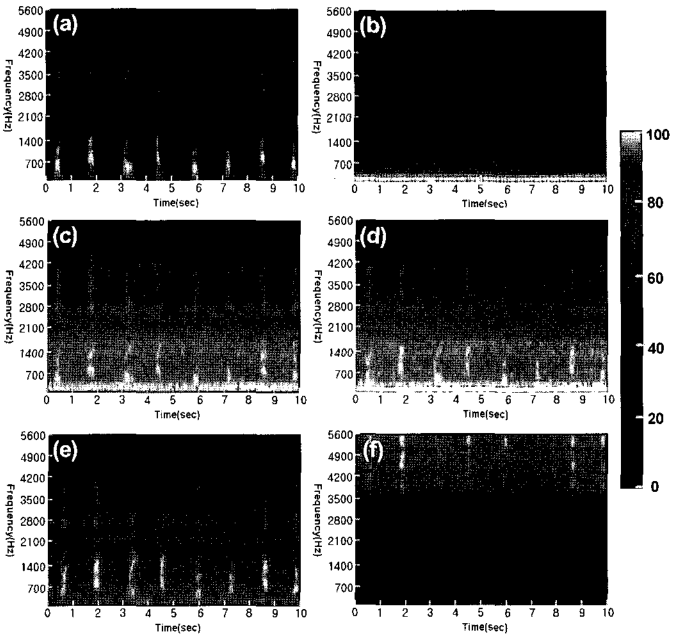



**Fig. 7.** Spectrogram of speech, noise, input and output signals. (a) original speech signals of one syllable from a male voice, (b) car noise, (c) the recorded signal of the front microphone, (d) the recorded signal of the rear microphone, (e) output speech signal extracted by the proposed method, (f) output speech signal extracted by the spectral subtraction method.

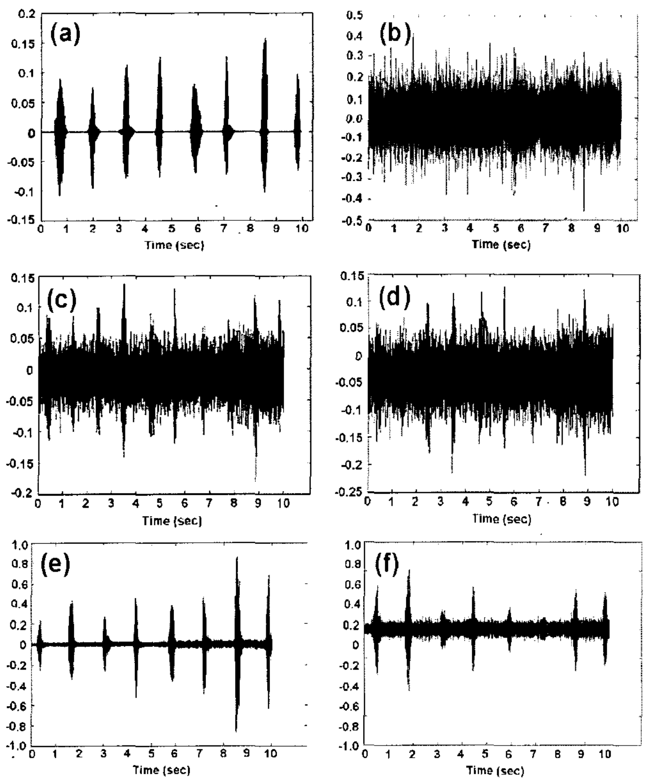



**Fig. 6.** Speech, noise, input and output signals in the time domain. (a) original speech signals of one syllable from a male voice, (b) car noise, (c) the recorded signal of the front microphone, (d) the recorded signal of the rear microphone, (e) output speech signal extracted by the proposed method, (f) output speech signal extracted by the spectral subtraction method.

Fig. 8 shows the results of the experiments with various background noises and the spectrogram of each noise signal. There is a little difference in the results of the experiments for each noise condition. (a), (b), and (c) of Fig. 8 are spectrograms of babble, car noise and factory noise, respectively. The same intensity of each noise signal was provided. Babble (a) and factory noise (c) signals are somewhat similar to white noise, which has the same intensity in all frequency bands, whereas car noise (b) has strong low frequency components, as stated above. The bar graph of Fig. 8 shows the improvement of SNR for various experimental conditions. The improvements in SNR for case (a) and (d) are approximately the same, and lower by 3.52 dB than those for cases (b)and (c). These results are caused by the characteristics of the noise signal. The noises of babble and car influenced the target speech signal in all frequency bands, whereas the noise of factory noise couldn't influences particular low frequency band. It can be said that when the noise source is the same, the improvement in the SNR is approximately the same. It doesn't matter whether one syllable or two syllables speech, or whether male or female speech. The average improvement of the SNR by the proposed method was 16.64 dB, whereas for the spectral subtraction it was 8.67 dB.

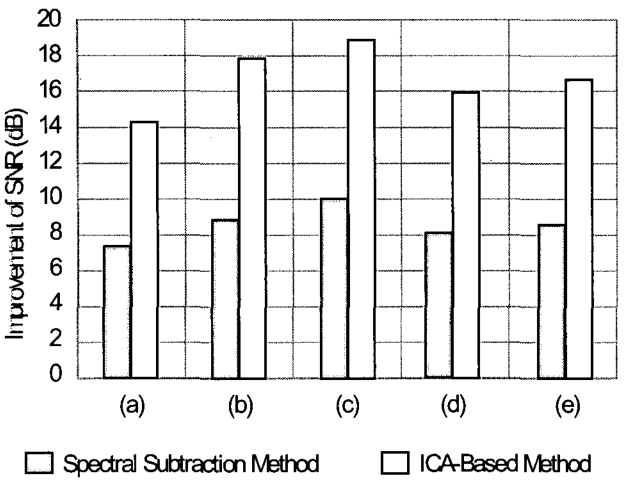
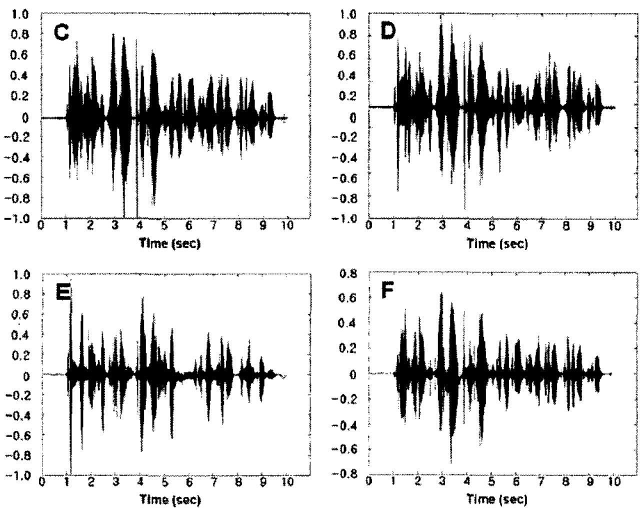



**Fig. 8.** Improvement of SNR for both methods in each noise environment: (a) male one syllable/babble noise, (b)male one syllable/car noise, (c) female two syllables/car noise, (d) female two syllables/factory noise, (e) average improvement



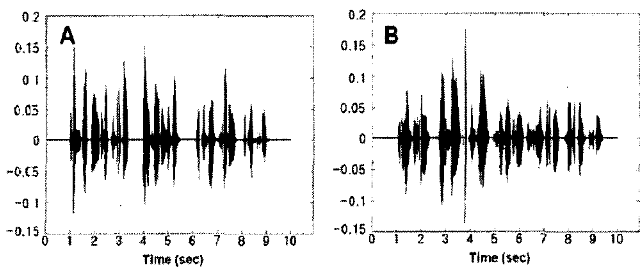




**Fig. 9.** Experiment of separation of speech signals of two speakers. (a) original speech signals of male voice, (b) original speech signals of female voice, (c) the recorded signal of the front microphone, (d) the recorded signal of the rear microphone, (e) unmixed signals (male voice) obtained by the proposed method, (f) unmixed signals (female voice) obtained by the proposed method.

### Results of Experiment #2

Fig. 9 displays the source signals, the mixed signals and the separated signals obtained by the proposed method. Fig. 9 (a) is original speech signals of male voice, (b) is original speech signals of female voice, (c) is the recorded signal of the front microphone, (d) is the recorded signal of the rear microphone, (e) is unmixed signals (male voice) obtained by the proposed method, and (f) is unmixed signals (female voice) obtained by the proposed method. As we can see, the proposed method could separate individual speech sources well, although there is a small cross-talk component. The quality of the recovered speech signal was, subjectively, quite good and 5 normal listeners who participated in our subjective speech recognition test could repeat the sentences.

### Results of Experiment #3

We believe that normal hearing person well listen speech signal than the hearing impaired no matter condition that it is quiet or noisy. In our experiments, we estimated speech perception of subject in different SNR conditions. In case of that SNR is -6dB , average speech perception of normal listeners is 61% for original speech which is non-processed speech. After spectral subtraction processing, it became 81% whereas, after ICA processing, it became 93%. In the same SNR condition, for the hearing impaired, average speech perceptions are 16%, 27% and 39% for original speech, the processed speech by spectral subtraction and ICA, respectively. When SNR was set -12dB, for normal listeners, average speech perceptions are 23%, 54% and 83% for original speech, the processed speech by spectral subtraction and ICA, whereas they are 8%, 15%, and 34% for the hearing impaired respectively. In the Fig. 10, the result of clinical tests was shown. In both case of 6dB and 12dB SNR, improvement of normal listener's speech perception was higher than the hearing impaired group. The reason may be that normal listeners are young students who might be considered to understand a fast acoustic change of speech. For both of normal listener and the hearing impaired, perception improvement in low SNR condition is higher than high SNR condition. In other words, it is very helpful to increase SNR In low SNR condition.

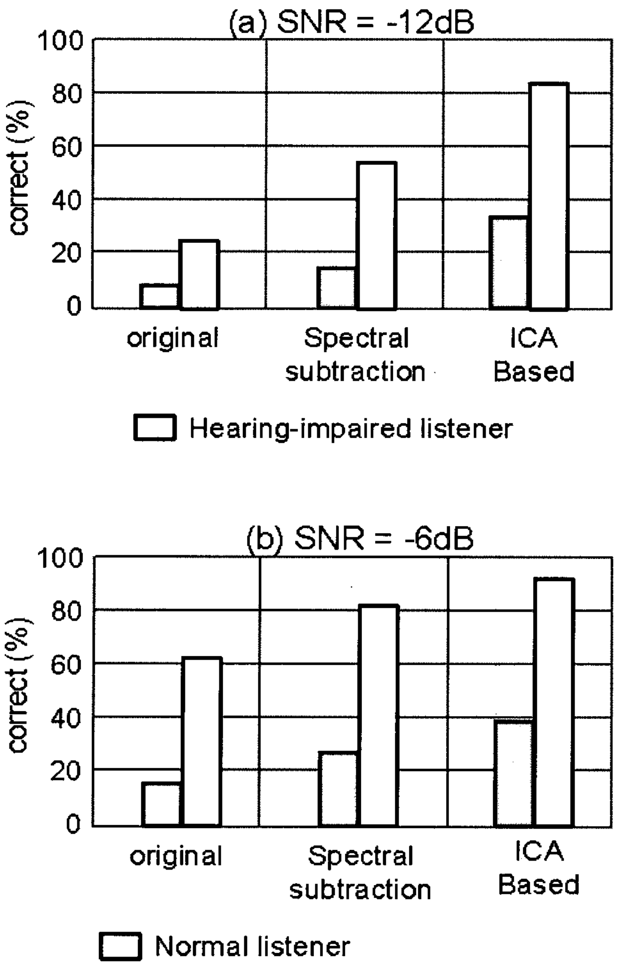



**Fig. 10.** Average improvement of speech perception: (a) SNR is 6dB, (b) SNR is 12dB.

## CONCLUSION

In this paper, a new approach to improve the speech discrimination performance of a hearing aid with background noise was proposed. The proposed new algorithm can separate target speech signal from various noise. The results revealed that the proposed method improved SNR as much as 16.64 dB, comparing with 8.67 dB for spectral subtraction method. The result of the separation of speech signals of two speakers shows the possibility of applying the proposed method to hearing aids in actual conversation situations. In particular, when hearing aids users have a conversation with someone in noisy environment, such as in factory, market, department store or restaurant, the proposed method could give clearer speech than conventional hearing aids.

The source separation algorithm can be applied to cochlear implant as well as hearing aid, in which case the electrical stimulation of the cochlear implant for the speech signal would be more delicate in the noisy environment.

## REFERENCE

[1] M. Kompis, P. Feuz, G. Valentini, and M. Pelizzone, *"A combined fixed/adaptive beam forming noise-reduction system for hearing aids"*, Engineering in Medicine and Biology Society, Proceedings of the 20th Annual International Conference of the IEEE, Vol. 6, pp.3136-3139, 1998.

[2] A. Bell and T. Sejnowski, *"An Information-Maximization Approach to Blind Separation and Blind Deconvolution"*, Neural Computation, Vol. 7, pp.1129-1159, 1995.

[3] J. Anemüller and B. Kollmeier, *"Adaptive separation of acoustic source for anechoic conditions: A constrained frequency approach"*, Speech Communications, Vol. 39, pp.7995, 2003.

[4] K. Torkkola, *"Blind separation of delayed sources based on information maximization"*, Acoustics, Speech, and Signal Processing, ICASSP '96, Conference Proceedings, 1996 IEEE International Conference, Vol. 6, pp.3509-3512, 1996.

[5] W.A. Yost, Fundamentals of hearing: an introduction, Academic Press, pp. 193-203, 2000.

[6] H. Dillon, *"Tutorial: Compression? Yes, but for low or high frequencies, for low or high intensities and with what response times"*, Ear and Hearing, pp. 287-307, 1996.

[7] Y. C. Park, *"High performance digital hearing aid processor with psychoacoustic loudness correction"*, ICCE '97, pp.100-101, 1997.

[8] Noisex-92 database is from 'http://spib.rice.edu/spib/select_noise.html'.