# BLIND SOURCE SEPARATION WITH DIFFERENT SENSOR SPACING AND FILTER LENGTH FOR EACH FREQUENCY RANGE

Hiroshi Sawada, Shoko Araki, Ryo Mukai and Shoji Makino
NTT Communication Science Laboratories, NTT Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan
{sawada,shoko,ryo,maki}@cslab.kecl.ntt.co.jp
http://www.kecl.ntt.co.jp/icl/signal/

**Abstract. This paper presents a method for blind source separation using several separating subsystems whose sensor spacing and filter length can be configured individually. Each subsystem is responsible for source separation of an allocated frequency range. With this mechanism, we can use appropriate sensor spacing as well as filter length for each frequency range. We obtained better separation performance than with the conventional method by using a wide sensor spacing and a long filter for a low frequency range, and a narrow sensor spacing and a short filter for a high frequency range.**

## INTRODUCTION

Blind source separation (BSS) is a technique to estimate original source signals using only sensor observations that are mixtures of the original signals. If source signals are mutually independent and non-Gaussian (or non-stationary), we can apply techniques of independent component analysis (ICA) to solve a BSS problem. Although theoretical aspects of BSS and ICA have been well studied in the past decade [1, 4, 6–8, 11], practical issues that occur in a real-world application have not been sufficiently explored. We should take account of various things when we construct a BSS system that can separate signals mixed in a real environment. Especially, in this paper, we discuss two practical issues, (1) the spacing of sensors and (2) the length of finite impulse response (FIR) filters in a separating system, along with their relationship to frequencies.

If a sensor spacing is small enough compared with the distance between a source and a sensor, a BSS solution forms spatial nulls to the directions of jammer signals. Such behavior is similar to a beamformer [14, 19], and there are some papers [2, 5, 10] that indicate the relationship between BSS and beamforming. If the spacing is longer than half of the wave length, spatial aliasing occurs: nulls are formed to several directions. And if nulls are formed in the directions of both a jammer and a target signal, a separating system cannot utilize phase differences and must rely on amplitude differences which are very small in most cases. On the other hand, if

0-7803-7616-1/02/$17.00 (c) 2002 IEEE 

the spacing is too small, the phase difference in a low frequency becomes too small and it is difficult to achieve a good performance of source separation. Generally speaking, a low frequency prefers long spacing and a high frequency prefers short spacing. If we could configure the spacing of sensors according to frequencies, better performance of BSS would be obtained.

As for FIR filters in a separating system, they should have sufficient length to cover the reverberation of a real room environment. If the filter length is too short, the separating system cannot separate a portion of the reverberation longer than the filter length [12]. However, an excessively long filter diminishes BSS performance because the amount of data samples per filter coefficient decreases [3] and it can also cause a long echo [12]. In a typical room, the reverberation lasts longer in a low frequency than in a high frequency. Therefore, changing filter length according to frequencies would also obtain better performance of BSS.

This paper proposes a method of BSS using several separating subsystems to solve the two practical issues described above. Each subsystem is responsible for a frequency range allocated by the main system. The sensor spacing as well as the filter length of a subsystem can be configured separately according to the allocated frequency range, so that a better performance of BSS is obtained.

The next section explains the basics of BSS based on ICA. The third section proposes a method of BSS using several separating subsystems, and discusses how sensor spacing as well as filter length affect BSS performance. The fourth section shows experimental results which support our discussions. The fifth section concludes this paper.

## BLIND SOURCE SEPARATION

Let us formulate blind source separation (BSS) of convolutive mixtures. Suppose that $P$ source signals $s_p(t)$ are mixed in an environment and observed at $Q$ sensors $x_q(t) = \sum_{p=1}^{P} \sum_k h_{qp}(k) s_p(t-k)$, where $h_{qp}(k)$ represents the impulse response from source $p$ to sensor $q$. The set of impulse responses $h_{qp}(k)$ represents the mixing process. The goal of BSS is to obtain a separating system and also separated signals $y_1(t), \ldots, y_P(t)$ that are estimates of the source signals $s_1(t), \ldots, s_P(t)$. The separating system typically consists of a set of FIR filters $w_{rq}(k)$ that produces separated signals $y_r(t) = \sum_{q=1}^{Q} \sum_k w_{rq}(k) x_q(t-k)$. The separation has to be done without knowing the impulse responses $h_{qp}(k)$ nor the information of the original source signals $s_p(t)$. If the source signals $s_p(t)$ are mutually independent, we can apply independent component analysis (ICA) to construct the separating system. Figure 1 shows a BSS model where $P = Q = 2$.

Many methods have been proposed to solve the convolutive BSS problem. They can be classified into two approaches. The first one is a time-domain approach, where the coefficients of the separating filters $w_{rq}(k)$ are calculated directly in the convolutive mixture model. The other one is a frequency-domain approach [2, 3, 9, 10, 12, 13, 16, 18], where frequency responses $W_{rq}(f)$ of the separating system are first calculated, and then the time-domain representation of the separating filters $w_{rq}(k)$ is obtained by applying inverse DFT (discrete Fourier transform) to them.

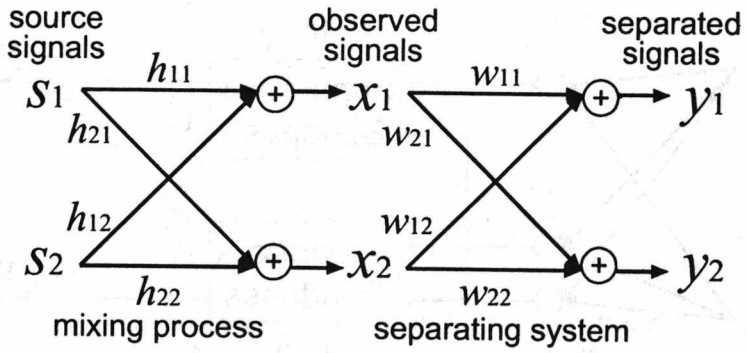


Figure 1: BSS model

The time and frequency representations of an FIR filter can be mutually converted by DFT and inverse DFT. The length $L$ of each FIR filter $w_{rq}(k)$ corresponds to the resolution of each frequency response $W_{rq}(f)$.

This paper employs the frequency-domain approach. It has an advantage in that ICA is applied for instantaneous mixtures, which are easier to solve than convolutive ones. By $L$-point short time DFT, time-domain signals $x_q(t)$ are converted into frequency-domain time-series signals $X_q(f,m)$, where $f = 0,\ f_s/L,\ \ldots,\ f_s(L-1)/L$, ($f_s$: sampling frequency). Assume that $\mathbf{X}(f,m)$ is a $Q$-dimensional vector $\mathbf{X}(f,m) = [X_1(f,m),\ \ldots,\ X_Q(f,m)]^T$. To obtain frequency responses $W_{rq}(f)$ of the separating system, we solve an ICA problem $\mathbf{Y}(f,m) = \mathbf{W}(f)\mathbf{X}(f,m)$, where $\mathbf{Y}(f,m) = [Y_1(f,m),\ldots,Y_P(f,m)]^T$ and $\mathbf{W}(f)$ is an $P \times Q$ matrix whose elements are $W_{rq}(f)$. $Y_r(f,m)$ is a frequency-domain representation of $y_r(t)$ and should be mutually independent.

The ICA algorithm we use is the information maximization approach [4] combined with the natural gradient [1]. A separating matrix $\mathbf{W}$ is gradually improved by the learning rule $\Delta\mathbf{W} = \mu\,[\mathbf{I} - \langle\Phi(\mathbf{Y})\mathbf{Y}^H\rangle]\,\mathbf{W}$. In this formula, $\mu$ is a step-size parameter that has an effect on the speed of convergence, $\langle\cdot\rangle$ denotes the averaging operator, and $\Phi(\cdot)$ is a nonlinear function. We use $\Phi(Y_r) = \tanh(\eta|Y_r|)\ e^{j\cdot phase(Y_r)}$ considering the fact that the density of $Y_r$ is independent of the phase [16]. $\eta$ is a gain parameter to control the nonlinearity. After ICA is solved in all frequency bins, we need to solve the permutation and scaling problem. We use the method in [9] to solve the permutation and scaling (amplitude) problem, and the method in [10] to solve the scaling (phase) problem.

## BSS USING SEVERAL SEPARATING SUBSYSTEMS

This section presents a BSS system whose sensor spacing and FIR filter length can be configured separately according to a frequency range.

### System Overview

The system consists of several separating subsystems $sub_i$ ($i = 1, 2, \ldots$) and associated sensor sets $\{x_1^i, \ldots, x_Q^i\}$. Each subsystem acts as a bandpass filter for a

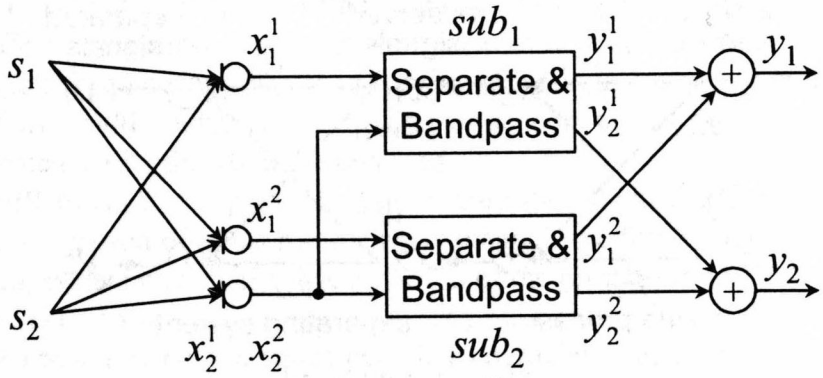


Figure 2: System structure

particular frequency range and is responsible for separating a mixed signal of that range. The FIR filter length as well as the sensor spacing can be specified separately in each subsystem.

The process flow of the system is as follows. First, we allocate a frequency range to each subsystem. An appropriate frequency range for a sensor spacing will be discussed later. Then, each subsystem $sub_i$ produces separated signals $y_r^i$, which contain the specified frequency range only. Finally, the outputs $y_r^i$ of the subsystems are integrated to form the output $y_r = \sum_i y_r^i$ of the system.

Figure 2 shows an example where two source signals are to be separated using two subsystems. The first subsystem $sub_1$ and its sensor set $\{x_1^1, x_2^1\}$ is used for a low frequency range and the second one ($sub_2$ and $\{x_1^2, x_2^2\}$) is used for a high frequency range. We can share a sensor ($x_2^1$ and $x_2^2$) between different subsystems as shown in this example.

**Subsystem**

Now let us discuss the organization of a subsystem. A straightforward way to construct a subsystem is to connect bandpass filters in serial to a basic separating system. However, this organization has a drawback in that the delay of the overall system is increased. Our approach is instead to make FIR filters $w_{rq}(k)$ that work to separate the signals and act as a bandpass filter simultaneously. Suppose that a subsystem $sub_i$ is responsible for a frequency range $f_{low}^i \leq f < f_{high}^i$. First, frequency responses $W_{rq}(f)$ of the range are calculated by ICA as explained above. Then, frequency responses $W_{rq}(f)$ out of the range, that is $f < f_{low}^i$ and $f_{high}^i \leq f$, are set to 0 for filtering out. Finally, the separating filters $w_{rq}(k)$ are obtained by applying inverse DFT to all the frequency responses $W_{rq}(f)$.

As explained in the second section, the time and frequency representations of an FIR filter can be mutually converted by DFT and inverse DFT. Thus, we can also apply the time-domain approach of BSS to obtain the frequency responses $W_{rq}(f)$ of the allocated range.

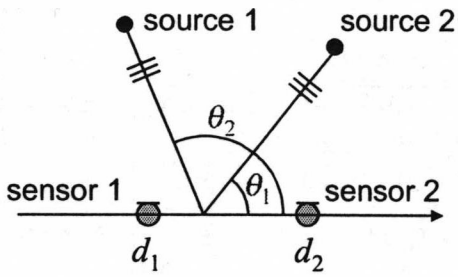


Figure 3: Source directions and sensor positions

**Appropriate Frequency Range for a Sensor Spacing**

As explained in the introduction, a BSS solution usually forms spatial nulls to the directions of jammer signals. This subsection discusses an appropriate frequency range for a sensor spacing by looking into a directivity pattern formed by a BSS solution.

The frequency response $B_{rp}(f)$ from a source $s_p$ to a separated signal $y_r$ can be decomposed as $B_{rp}(f) = \sum_{q=1}^{Q} W_{rq}(f) \cdot H_{qp}(f)$, where $W_{rq}(f)$ is the frequency response of a separating system obtained using ICA. Based on the mixing model used in the beamforming theory [14, 19], the frequency response $H_{qp}(f)$ of an impulse response $h_{qp}$ can be approximated as $H_{qp}(f) = e^{j2\pi f c^{-1} d_q \cos\theta_p}$, where $\theta_p$ is the direction of source $s_p$ and $d_q$ is the position of sensor $q$ (Fig. 3). A plane wavefront is assumed in this equation, and only the direct path of the impulse response is considered. Consequently, $B_{rp}(f)$ can be expressed as $B_{rp}(f) = \sum_{q=1}^{Q} W_{rq}(f) \cdot e^{j2\pi f c^{-1} d_q \cos\theta_p}$. If we regard $\theta_p$ as a variable $\theta$, $B_{rp}(f)$ is expressed as $B_r(f,\theta) = \sum_{q=1}^{Q} W_{rq}(f) \cdot e^{j2\pi f c^{-1} d_q \cos\theta}$. It changes according to the direction $\theta$, and thus is called a directivity pattern.

Figure 4 shows the gain of a directivity pattern (left) and its plot on a complex plane (right). This directivity pattern was calculated using an actual BSS problem with two sources and two sensors. The sensor positions and frequency were $d_1 = 0$, $d_2 = 141.5$ mm, and $f = 789$ Hz, respectively. The estimated source directions were $\theta_1 = 47°$ and $\theta_2 = 134°$ ($\Diamond$ and $\square$ in Fig. 4). We see that the gain of $\theta_2$ is almost 0 while the gain of $\theta_1$ is sufficiently large. This is a very nice situation, and the separation performance was good.

Now, let us discuss an appropriate frequency range for a sensor spacing. The directivity pattern $B_r(f,\theta)$ of the above example can be simplified as $B_r(f,\theta) = W_{r1}(f) + W_{r2}(f) \cdot e^{j2\pi f c^{-1} d \cos\theta}$ by assuming $d_1 = 0$ and $d_2 = d$. As $\cos\theta$ changes from 1 to $-1$ in accordance with the change of direction $\theta$ ($0° \leq \theta \leq 180°$), the frequency response $B_r(f,\theta)$ moves on the circle (right hand side of Fig. 4) whose center and radius are $W_{r1}$ and $W_{r2}$, respectively. Based on this, we see that spatial aliasing does not occur if $2\pi f c^{-1} d < \pi \Leftrightarrow f < c/(2d)$.

We can, however, derive a more appropriate frequency range if the directions of two sources $\theta_1$ and $\theta_2$ can be estimated. By using the MUSIC (MUltiple SIgnal Classification) algorithm [14, 17], we can estimate the source directions even with mixed signals if the number of sources is less than the number of sensors

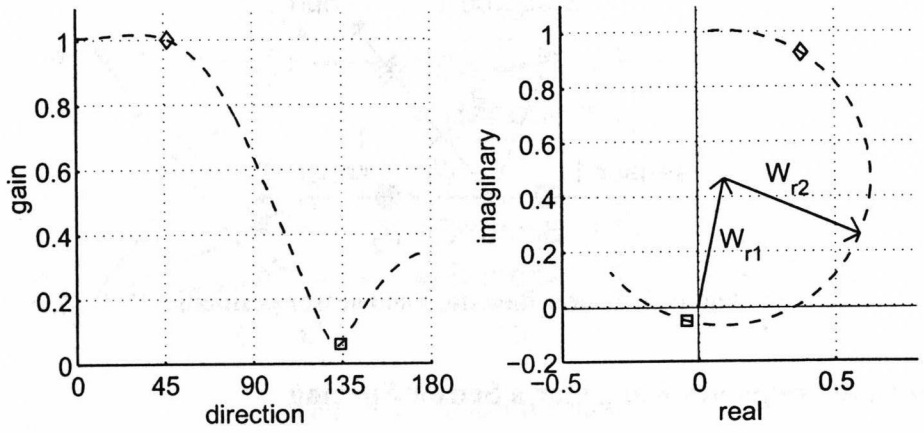


Figure 4: A directivity pattern (left) and its plot on a complex plane (right)

(e.g., three sensors for two sources). As frequency $f$ increases, the phase difference $2\pi f c^{-1} d \cos\theta_1 - 2\pi f c^{-1} d \cos\theta_2$ corresponding to the two source directions increases. And as the phase difference approaches $\pi$, the gain difference between two directions $\theta_1$ and $\theta_2$ approaches the maximum and the performance of BSS can be maximized. After the phase difference exceeds $\pi$ and approaches $2\pi$, the gain difference decreases and the performance of BSS also decreases. Therefore, we consider it an appropriate condition that the phase difference does not overly exceed $\pi$. Such a condition can be expressed as $2\pi f c^{-1} d \cos\theta_1 - 2\pi f c^{-1} d \cos\theta_2 \leq \alpha\pi \Leftrightarrow f \leq \alpha c/[2d(\cos\theta_1 - \cos\theta_2)]$, where $\alpha$ is a parameter to allow how the phase difference exceeds $\pi$.

### Appropriate Filter Length for a Frequency Range

As explained in the introduction, an appropriate filter length $L$ for a frequency range depends on the reverberation on the range. Therefore, we need to know the rough characteristics of the room where BSS is to be performed. Although such a situation may not suit the word "blind", we can get such rough information more easily than

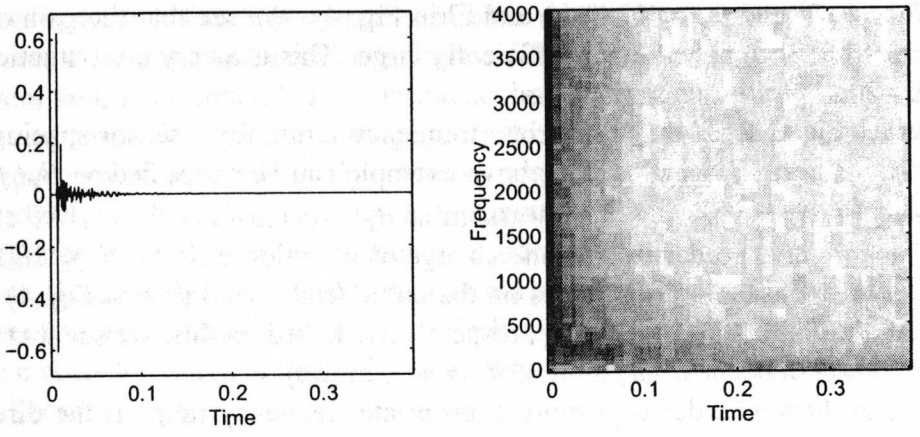


Figure 5: An impulse response and its spectrogram (the reverberation time is 190 ms measured at 500 Hz)

| $d_2 - d_1$ | Narrow: 28.3 mm | | Wide: 141.5 mm | | proposed |
|---|---|---|---|---|---|
| $L$ | 2048 | 4096 | 2048 | 4096 | method |
| 0–1029 Hz | 11.49 dB | 15.37 dB | 14.47 dB | 19.74 dB | 19.74 dB |
| 1029–4000 Hz | 29.04 dB | 23.50 dB | 26.71 dB | 23.15 dB | 29.04 dB |
| 0–4000 Hz | 12.13 dB | 15.87 dB | 15.02 dB | 20.04 dB | 20.42 dB |

TABLE 1: BSS PERFORMANCE FOR FREQUENCY RANGES

an impulse response itself. Figure 5 shows an impulse response that we used in experiments and its spectrogram. We see that the reverberation lasts longer in a low frequency than in a high frequency, and that a long filter is desirable for a low frequency range and a short filter is desirable for a high frequency range.

## EXPERIMENTAL RESULTS

To show the effectiveness of the proposed BSS system, we conducted experiments to separate two speech signals. Mixed signals were obtained by convolving impulse responses and source signals. The set of impulse responses we used is in the RWCP sound scene database in real acoustic environments [15]. As source signals, we selected male speeches from the ASJ (Acoustical Society of Japan) continuous speech corpus. The lengths of source signals were 7.4 seconds. The sampling rate was 8 kHz. We used a Hanning window in short time DFTs to obtain frequency-domain representations $\mathbf{X}$ of observed signals. The shifting interval of the window was a quarter of the window length. In the ICA algorithm, we normalized $\mathbf{X}$ to have unit variances. The step-size parameter $\mu$ and the parameter $\eta$ were set to 0.1 and 100, respectively. The number of iterations was 100 for each frequency bin.

We used three sensors to construct two separating subsystems, as shown in Fig. 2. The "Wide" spacing was 141.5 mm and the "Narrow" one was 28.3 mm. The two source directions estimated by the MUSIC algorithm were 53° and 143°. The maximum frequency that the "Wide" spacing dealt with was 1029 Hz. This was calculated based on the criteria discussed in the previous section by setting $\alpha = 1.2$. Consequently, we used "Wide" for a frequency range 0–1029 Hz, and "Narrow" for a frequency range 1029–4000 Hz. As for filter length $L$, we used 4096 for the range 0–1029 Hz and 2048 for the range 1029–4000 Hz. Table 1 shows BSS performances for the two frequency ranges. The numbers are the averages of SNRs (signal-to-noise ratio) at two outputs. We calculated the SNR at output $p$ as $10\log[\sum_t y_p^s(t)^2 / \sum_t y_p^c(t)^2]$, where $y_p^s(t)$ is a portion of $y_p(t)$ that comes from a source signal $s_p(t)$ and $y_p^c(t) = y_p(t) - y_p^s(t)$. The last column shows the result of the proposed method. The second through fifth columns show the results of conventional methods, where neither sensor spacing $d_2 - d_1$ nor filter length $L$ can be changed according to frequency ranges. We see that the proposed method improves the performance of BSS substantially.

Then, we investigated the effect of sensor spacing by setting filter lengths to the same $L = 2048$ for all frequency ranges. Figure 6 shows differences in SNR, SNR(Wide) – SNR(Narrow), measured at various frequencies. The vertical line is

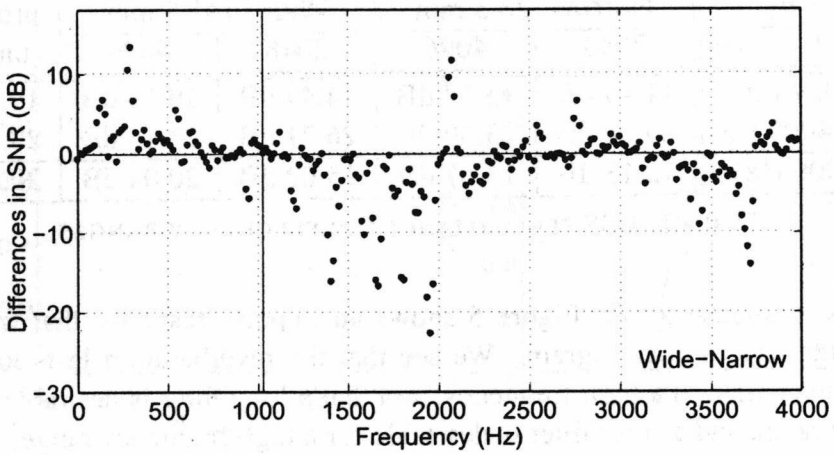


Figure 6: Differences in SNR at various frequencies

on 1029 Hz. There are many positive values in frequencies below 1029 Hz and many negative values in frequencies above 1029 Hz. The data explain the effectiveness of the proposed method. Figure 7 shows the directivity patterns. The upper and the lower half correspond to a low and a high frequency $f = 141$ Hz and $f = 1762$ Hz, respectively. $\Diamond$ and $\square$ show the estimated source directions. We see that it is difficult to form a directional null with "Narrow" in a low frequency. On the contrary, spatial aliasing occurs with "Wide" in a high frequency, and it is hard to distinguish the gain of the two source directions.

Next, we examined the effect of filter length by setting sensor spacings to the same $d_2 - d_1 = 28.3$ mm. Figure 8 shows SNRs measured in several frequency

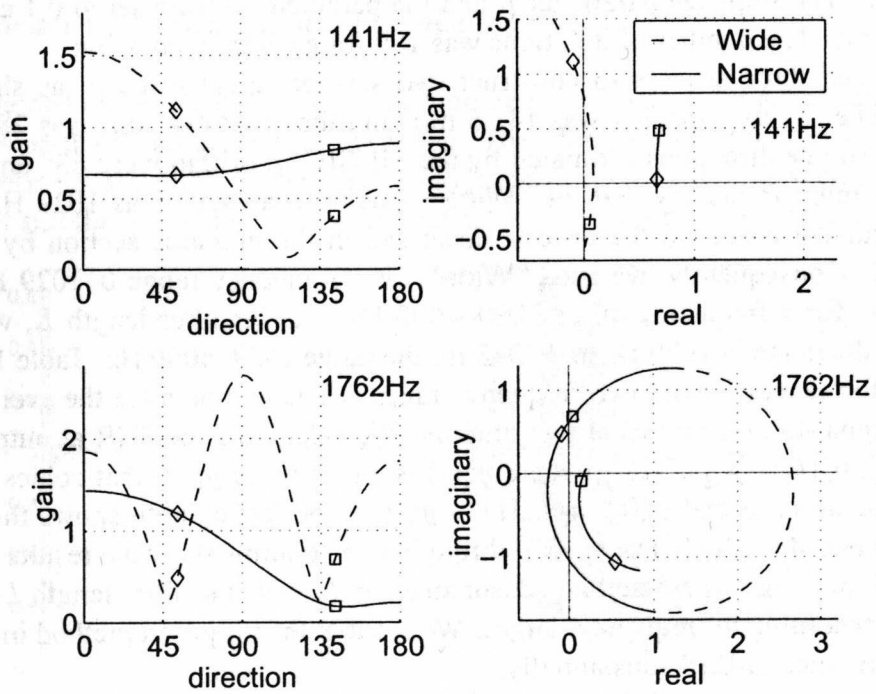


Figure 7: Directivity patterns (left) and their plot on complex planes (right)

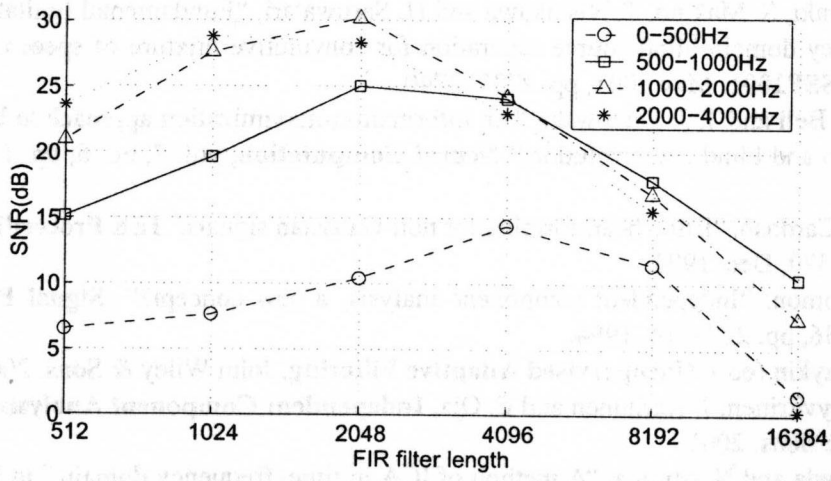


Figure 8: Effect of FIR filter length in frequency ranges

ranges by changing the length of FIR filters. We see that a long filter (4096: 512 ms) works well for a low frequency range (0–500 Hz). On the other hand, a shorter filter (1024: 128 ms or 2048: 256 ms) is better in high frequency ranges (1000–2000 Hz and 2000–4000 Hz).

## CONCLUSION

We proposed a BSS system whose sensor spacing and FIR filter length could be configured separately according to a frequency range. Generally, a wide spacing of sensors and a long filter length are effective for low frequencies, and a narrow spacing of sensors and a short filter length are effective for high frequencies. The proposed method is capable of being configured appropriately for a specific frequency range, and therefore achieves better performance than conventional methods.

## ACKNOWLEDGEMENT

We would like to thank Mr. Joshua Jacobs who conducted preliminary experiments. Thanks are also expressed to Dr. Hiroshi Saruwatari for valuable discussions and to Dr. Shigeru Katagiri for continuous encouragement.

## REFERENCES

[1] S. Amari, "Natural gradient works efficiently in learning," **Neural Computation**, vol. 10, no. 2, pp. 251–276, 1998.

[2] S. Araki, S. Makino, R. Mukai and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in **Proc. Eurospeech2001**, Sept. 2001, pp. 2595–2598.

[3] S. Araki, S. Makino, T. Nishikawa and H. Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," in **Proc. ICASSP2001**, May 2001, pp. 2737–2740.

[4] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," **Neural Computation**, vol. 7, no. 6, pp. 1129–1159, 1995.

[5] J. F. Cardoso, "Blind beamforming for non-Gaussian signals," **IEE Proceedings-F**, pp. 362–370, Dec. 1993.

[6] P. Comon, "Independent component analysis, a new concept?" **Signal Processing**, vol. 36, pp. 287–314, 1994.

[7] S. Haykin (ed.), **Unsupervised Adaptive Filtering**, John Wiley & Sons, 2000.

[8] A. Hyvärinen, J. Karhunen and E. Oja, **Independent Component Analysis**, John Wiley & Sons, 2001.

[9] S. Ikeda and N. Murata, "A method of ICA in time–frequency domain," in **Proc. ICA '99**, Jan., pp. 365–370.

[10] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in **Proc. ICASSP2000**, pp. 3140–3143.

[11] T. W. Lee, **Independent Component Analysis - Theory and Applications**, Kluwer academic publishers, 1998.

[12] R. Mukai, S. Araki and S. Makino, "Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment," in **Proc. Eurospeech2001**, Sept. 2001, pp. 2599–2603.

[13] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," **IEEE Trans. Speech Audio Processing**, vol. 8, no. 3, pp. 320–327, May 2000.

[14] S. U. Pillai, **Array Signal Processing**, Springer-Verlog, 1989.

[15] Real World Computing Partnership, "RWCP sound scene database in real acoustic environments," `http://tosa.mri.co.jp/sounddb/indexe.htm`.

[16] H. Sawada, R. Muaki, S. Araki and S. Makino, "Polar coordinate based nonlinear function for frequency-domain blind source separation," in **Proc. ICASSP2002**, May 2002, pp. 1001–1004.

[17] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," **IEEE Trans on Antennas and Propagation**, vol. 34, pp. 276–280, March 1986.

[18] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," **Neurocomputing**, vol. 22, pp. 21–34, 1998.

[19] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," **IEEE ASSP Magazine**, pp. 2–24, April 1988.