# 14 Subband Based Blind Source Separation

Shoko Araki and Shoji Makino

NTT Communication Science Laboratories
Soraku-gun, Kyoto 619-0237, Japan
E-mail: {shoko, maki}@cslab.kecl.ntt.co.jp

**Abstract.** In this chapter, we address subband-based blind source separation (BSS) for convolutive mixtures of speech by reporting a large number of experimental results. The subband-based BSS approach offers a compromise between time-domain and frequency-domain techniques. The former is usually difficult and slow with many separation filter coefficients to estimate. With the latter it is difficult to estimate statistics when the adaptation data length is insufficient. With subband-based BSS, a sufficient number of samples for estimating statistics can be held in each subband by using a moderate number of subbands. Moreover, by using FIR filters in each subband, which are shorter than the filters used for time-domain BSS, we can handle long reverberation. In addition, subband-based BSS allows us to select the separation method suited to each subband. Using this advantage, we introduce efficient separation procedures that take both the frequency characteristics of the room reverberation and speech signals into consideration. In concrete terms, longer separation filters and an overlap-blockshift in BSS's batch adaptation in low frequency bands improve the separation performance. Consequently, frequency-dependent subband processing is successfully realized with subband-based BSS.

## 14.1 Introduction

Blind source separation (BSS) is an approach that estimates original source signals $\mathrm{s}_i(n)$ using only information on the mixed signals $\mathrm{x}_j(n)$ observed in each input channel. This technique can be applied for audio applications such as noise robust speech recognition, high-quality hands-free telecommunication, and hearing aid systems.

We consider the BSS of speech signals in a real environment, i.e., the BSS of convolutive mixtures of speech. In a real environment, signals are filtered by an acoustic room channel. To separate such complicated mixtures, we need to estimate the separation filters of several thousand taps. Several methods have been proposed for achieving the BSS of convolutive mixtures [1] and most of these utilize independent component analysis (ICA) [2], [3]. To solve the convolutive BSS problem, algorithms in the time and frequency domains have been proposed [4–12].

In time-domain BSS, ICA is directly applied to convolutive mixtures and separation FIR filters are directly estimated (e.g., [4–8]). Therefore, the independence of output signals can be evaluated directly. However, the convergence of time-domain BSS algorithms is generally not good. This is because

the adaptation of such a long separation filter is very complex and there are many local minima [3]. The computational complexity is also a problem. Moreover, most time-domain BSS algorithms have another problem: the whitening effect, which means the signal's spectrum becomes flat. Because most time-domain BSS algorithms were designed for i.i.d. signals, these algorithms try to make output signals both spatially and temporally independent [8]. When we apply such time-domain BSS algorithms to mixtures of speech signals, the output speech signals are whitened and sound unnatural.

By contrast, in frequency-domain BSS, mixtures are converted into the frequency domain and ICA is applied to instantaneous mixtures at each frequency (e.g., [9–12]) as shown in the previous chapter. Although we can greatly reduce the computational complexity by using frequency-domain BSS, frequency-domain BSS algorithms have inherent issues. One is that the independence is evaluated at each frequency. In a real environment, an impulse response changes momentarily. Therefore it is preferable that we estimate separation filters using adaptation data that are as short as possible, especially when we use a batch algorithm. However, when we apply a longer frame that can cover realistic reverberation for speech mixtures of a few seconds, the number of samples in each frequency bin becomes small, and therefore, we cannot correctly estimate the statistics in each frequency bin [13]. This means that, in such a case, the independence is not evaluated correctly. This is our strongest reason for utilizing subband-domain BSS method. We also face permutation and scaling problems, which result in the estimated source signal being recovered with a different permutation and gain in different frequency bins. Recently, some solutions have been provided for these problems [12], [14–17] and some of these were introduced in the previous chapter.

Motivated by these facts, we introduce a BSS method that employs subband processing [18], [19]. Hereafter, we call this method subband BSS. With subband BSS, observed mixed signals are decomposed into the subband domain with a filterbank and then separated in each subband using a time-domain BSS algorithm. Then separated signals in each subband are synthesized to obtain fullband separated signals. With this method, we can choose a moderate number of subbands, therefore, we can maintain a sufficient number of samples in each subband. The subband system also allows us to estimate FIR filters as separation filters in each subband. Moreover, as the separation filter length in each subband is shorter than that for time-domain BSS, it is easier to estimate separation filters than with time-domain BSS. Therefore, we can obtain separation filters that are long enough to cover reverberation. That is, the subband BSS approach copes with both the frequency-domain approach's difficulty in estimating statistics and the time-domain technique's difficulty in adapting many parameters.

In addition, subband BSS mitigates the permutation problem and whitening effect. Because the permutation problem does not occur within each subband, there are few permutation problems in subband BSS. Moreover, be-

cause the whitening effect can be limited in each subband, subband BSS can mitigates the whitening effect. Of course, subband BSS reduces computational complexity [20], [21]. This is an additional merit of subband BSS.

Subband BSS offers another advantage in that it allows us to select the separation method suited to each subband. By using this advantage, we can employ an efficient separation procedure taking into consideration the frequency characteristics of room reverberation and speech signals [22], [23]. Generally speaking, an impulse response is usually longer in low frequency bands than in high frequency bands. This makes the separation in low frequency bands difficult. Moreover, because speech signals have high power in low frequency bands, the separation performance in low frequency bands dominates the speech separation performance. Therefore, it is very important to improve the separation performance in the low frequency bands for speech separation. In this chapter, we utilize longer separation filters and the overlap-blockshift technique in the low frequency bands.

The organization of this chapter is as follows. Section 14.2 describes the framework for the BSS of convolutive mixtures of speech. In Section 14.3, we explain the configuration of subband BSS and mention implementation issues. We confirm the validity of subband BSS in Section 14.4 by describing experiments undertaken with reverberant data. In Section 14.5, we show some ways to improve the low frequency subband performance in which the SIR is worse than at high frequencies. Here, we take into consideration the frequency characteristics of room reverberation and speech signals. The final section concludes this chapter.

## 14.2 BSS of Convolutive Mixtures

### 14.2.1 Model Description

In real environments, the observed microphone signals are affected by reverberation. Therefore, $N_s$ signals recorded by $N_m$ microphones are modeled as convolutive mixtures

$$\mathrm{x}_j(n) = \sum_{i=1}^{N_s} \sum_{l=1}^{P} \mathrm{h}_{ji}(l)\mathrm{s}_i(n-l+1), \quad j = 1, \cdots, N_m, \tag{14.1}$$

where $\mathrm{s}_i$ is the source signal from a source $i$, $\mathrm{x}_j$ is the signal observed by a microphone $j$, and $\mathrm{h}_{ji}$ is the $P$-taps impulse response from source $i$ to microphone $j$.

In order to obtain separated signals, we estimate the separation filters $\mathrm{w}_{ij}(n)$ of $Q$-taps, and obtain the separated signals

$$\mathrm{y}_i(n) = \sum_{j=1}^{N_m} \sum_{l=1}^{Q} \mathrm{w}_{ij}(l)\mathrm{x}_j(n-l+1), \quad i = 1, \cdots, N_s. \tag{14.2}$$

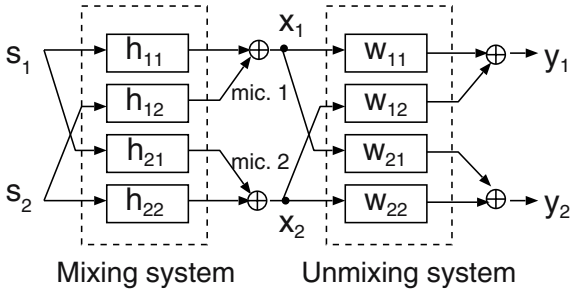


**Fig. 14.1.** BSS system configuration (when $N_s = N_m = 2$).

The separation filters are estimated so that the separated signals become mutually independent.

The BSS block diagram is shown in Fig. 14.1 for $N_s = N_m = 2$. In this chapter, we consider the case of $N_s = N_m = N_{sm}$.

### 14.2.2 Frequency-Domain BSS and Related Issue

**Frequency-domain BSS.** The frequency-domain approach to convolutive mixtures transforms the problem into an instantaneous BSS problem at each frequency [9–12]. Using $T$-point short-time Fourier transformation for (14.1), we obtain the approximate time-frequency representation of mixtures,

$$\mathbf{x}(f,m) = \mathbf{H}(f)\mathbf{s}(f,m), \quad m = 0, \cdots, L_m - 1, \tag{14.3}$$

where $f$ denotes the frequency bin, $m$ represents the time dependence of the short-time Fourier transformation (STFT), $L_m$ is the number of data samples in each frequency bin, $\mathbf{s}(f,m) = [s_1(f,m), \cdots, s_{N_{sm}}(f,m)]^T$ is the source signal vector, and $\mathbf{x}(f,m) = [x_1(f,m), \cdots, x_{N_{sm}}(f,m)]^T$ is the observed signal vector. We assume that the $(N_{sm} \times N_{sm})$ mixing matrix $\mathbf{H}(f)$ is invertible and that its $ji$ component $h_{ji}(f) \neq 0$. The STFT is usually executed by applying a window function of length $T$. In this chapter, we call this $T$ the STFT frame size.

The separation process can be formulated in a frequency bin $f$:

$$\mathbf{y}(f,m) = \mathbf{W}(f)\mathbf{x}(f,m), \quad m = 0, \cdots, L_m - 1, \tag{14.4}$$

where $\mathbf{y}(f,m) = [y_1(f,m), \cdots, y_{N_{sm}}(f,m)]^T$ is the separated signal vector, and $\mathbf{W}(f)$ represents an $(N_{sm} \times N_{sm})$ separation matrix at frequency $f$. In this chapter, we assume that the STFT frame size $T$ is equal to the separation filter length $Q$. The separation matrix $\mathbf{W}(f)$ is determined by ICA so that the outputs $y_i(f,m)$ become mutually independent. This calculation is carried out independently at each frequency.

**Dilemma of Frequency-Domain BSS.** In order to handle long reverberation, we need to estimate long separation filters $\mathrm{w}_{ij}(n)$ of $Q$-taps. If the filters are relatively short, we cannot reduce the reverberant components of interferences that are longer than the filters and this has a detrimental effect on the separation performance [24]. On the other hand, with a batch adaptation, it is desirable that separation filters can be estimated using adaptation data that are as short as possible. This is because an impulse response changes momentarily in a real environment. We therefore have to estimate long separation filters with short length of adaptation data.

However, we have reported in [13] that when we employ a long frame $T$ with a frame shift of $T/2$ for several seconds of data in order to prepare a separation filter long enough to cover reverberation (note that we are assuming $T = Q$), the separation performance degrades. One reason for this is that it becomes difficult to maintain a sufficient number of data samples to estimate the statistics in each frequency. This makes the estimation of statistics difficult. In particular, the independence assumption between the source signals seems to collapse [13]. Therefore, we cannot obtain sufficient separation performance with a long frame with frequency-domain BSS for short adaptive data.

## 14.3 Subband Based BSS

Subband BSS discussed in this section provides a solution to the dilemma of frequency-domain BSS described in the previous section. With this method, we can choose a moderate number of subbands, and therefore maintain a sufficient number of samples in each subband. Subband BSS also allows us to estimate short FIR filters as separation filters in each subband, due to the down-sampling procedure at the subband analysis stage. Therefore, we should be able to obtain a separation filter long enough to cover reverberation. Moreover, as the separation filter length in each subband is shorter than that for time-domain BSS, it is easier to estimate separation filters than in time-domain BSS. That is, the subband BSS approach offers a compromise between a time-domain technique, which is usually difficult and slow with many parameters to estimate, and a frequency domain technique, which has difficulty estimating statistics.

### 14.3.1 Configuration of Subband BSS

**Basic Configuration of Subband BSS.** The subband BSS system is composed of three parts: a subband analysis stage, a separation stage, and a subband synthesis stage (Fig. 14.2) [18], [19].

First, in the subband analysis stage, input signals $\mathrm{x}_j(n)$ are divided into $N$ subband signals $x_j(k, m)$, $k = 0 \cdots, N-1$, where $k$ is the subband index, $m$ is the time index, and $N$ is the number of subbands. A polyphase

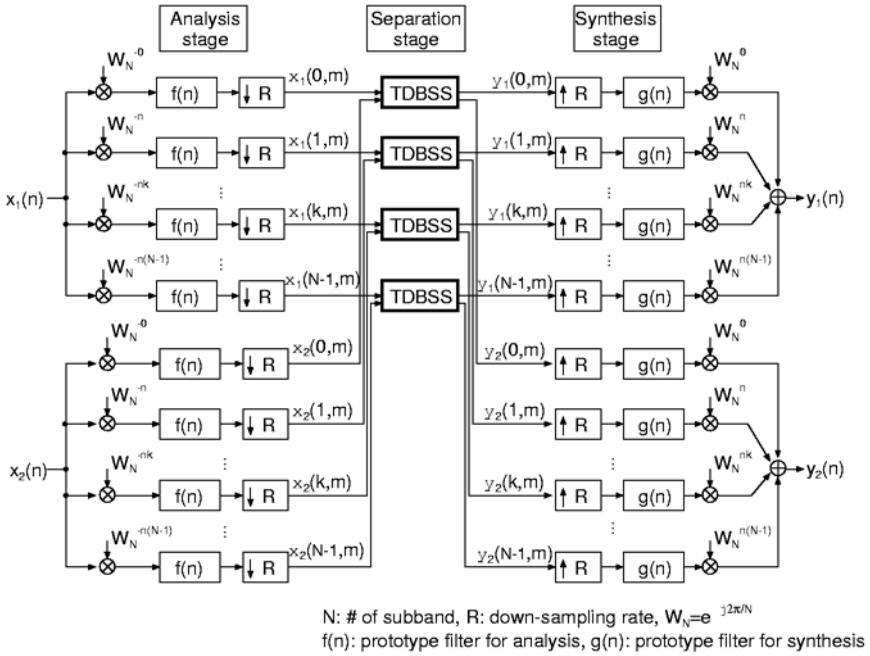


**Fig. 14.2.** Basic system configuration of subband BSS. TDBSS denotes time-domain BSS. A $2 \times 2$ case is depicted.

filterbank [25], including a cosine modulated filterbank [20] and a discrete Fourier transform (DFT) filterbank [21,26], is widely used as the subband analysis/synthesis system, because of its low computational complexity. A polyphase filterbank analyzer (synthesizer) basically consists of a modulator (demodulator), a prototype filter with a low pass characteristic, and a decimator (interpolator). The cosine modulated filterbank realizes a perfect reconstruction filterbank with real valued coefficients. The DFT filterbank can be effectively realized by using FFT, however, the subband analyzed signals $x_j(k,m)$ become complex number sequences. Since the outputs of a prototype filter are band-limited in each subband, we can employ decimation at the down-sampling rate $R$. However, as it is impossible to make an ideal low pass filter as a prototype filter, the adjacent bands overlap each other, i.e., aliasing occurs. Therefore, we should use a down-sampling rate of $R < N$ in order to reduce the aliasing distortion [27], which degrades the separation performance [20].

Then, time-domain BSS is executed on $x_j(k,m)$ and the separated signals $y_j(k,m)$ are obtained in each subband in the separation stage. If we utilize DFT filterbanks, we have to use a complex version of the time-domain BSS algorithm [21]. In each subband, we estimate FIR filters as separation filters so as to cover the reverberation. Since we employ down-sampling, short FIR

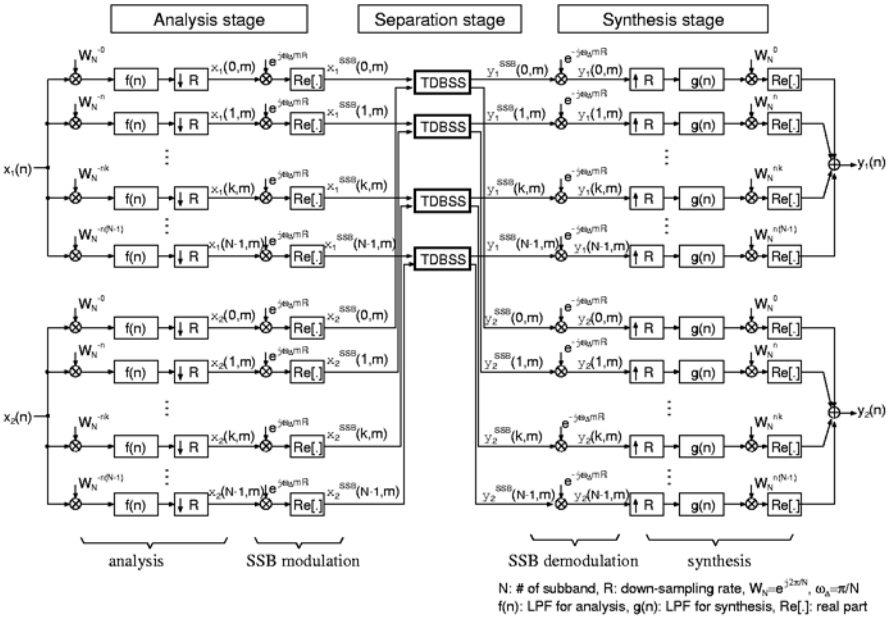


**Fig. 14.3.** Block diagram of subband BSS with an SSB filterbank. TDBSS denotes time-domain BSS. LPF denotes low pass filter. A $2 \times 2$ case is depicted.

filters of length $Q/R$ are sufficient to separate the subband signals in each subband.

Finally, in the subband synthesis stage, separated signals $\mathrm{y}_i(n)$ are obtained by synthesizing each separated signal $y_i(k,m)$.

**Subband BSS with SSB Filterbank.** In this chapter, we utilize a polyphase filterbank [25] with single sideband (SSB) modulation [28], which is widely used in the echo canceller area [29,30]. A block diagram of subband BSS with an SSB filterbank is shown in Fig. 14.3. Since this also has the form of a generalized discrete Fourier transform (GDFT) filterbank [28], the filterbank can be realized effectively by FFT. Furthermore, in order to make the analyzed signals real-valued, SSB modulation is adopted in the analysis stage (Fig. 14.3). Moreover, to avoid the aliasing problem, the SSB-modulated subband signals are not critically sampled, but oversampled, i.e., $R < N$. Here, we employ two-times oversampling $R = \frac{N}{4}$. The low-pass filter used here in the analysis filterbank as a prototype filter is $f(n) = \mathrm{sinc}(\frac{n\pi}{N/2})$ of length $6N$. By using SSB modulation, we obtain SSB modulated real-valued signals $x_j^{\mathrm{SSB}}(k,m)$ in each subband.

Thanks to the SSB modulation, in the separation stage, we can apply the time-domain BSS algorithm to $x_j^{\mathrm{SSB}}(k,m)$ without expanding it into a

complex-valued version. A detailed explanation of the time-domain BSS algorithm we employed is provided in the following subsection.

After obtaining the separated signals $y_i^{\mathrm{SSB}}(k,m)$ in each subband, we execute the SSB demodulation and synthesize them to obtain output signals $\mathrm{y}_i(n)$ in the time domain. The low-pass (prototype) filter used in the synthesis filterbank is $g(n) = \mathrm{sinc}(\frac{n\pi}{R/2})$ of length $6R$.

### 14.3.2 Time-Domain BSS Implementation for a Separation Stage

Thanks to the SSB modulation, we can use any real-valued time-domain BSS algorithm for subband BSS, including a higher order statistics based algorithm [4,5] and a second order statistics based algorithm [6,31]. A generic framework is discussed in [7]. Here, we describe the algorithm we used in the experiments reported in this chapter. In addition, this section describes how we can design the initial values of the separation filters for each subband.

**Time-Domain BSS Algorithm.** Here, we employ an algorithm based on time-delayed decorrelation for non-stationary signals [31]. Relying on the non-stationarity and non-whiteness of the source signals, this algorithm minimizes the cross-correlation of output signals for some time lags for all analysis blocks, simultaneously. It is verified that this algorithm works for convolutive mixtures of speech signals [32].

We estimate FIR filters as the separation filters $w_{ij}^k(m)$ in each subband $k$. We write them in a matrix form $\mathbf{W}^k(m)$ where its $ij$ component is $w_{ij}^k(m)$ for convenience. The adaptation rule of the $i$-th iteration is

$$\mathbf{W}_{i+1}^k(m) = \mathbf{W}_i^k(m) + \frac{\alpha}{BS}\sum_{b=0}^{BS-1}\{(\mathrm{diag}\mathbf{R}_y^b(0))^{-1}(\mathrm{diag}\mathbf{R}_y^b(m)) - (\mathrm{diag}\mathbf{R}_y^b(0))^{-1}\mathbf{R}_y^b(m)\} * \mathbf{W}_i^k(m), \tag{14.5}$$

where $\mathbf{R}_y^b(\tau)$ represents the covariance matrix of outputs $\mathbf{y}(m) \equiv [y_1^{\mathrm{SSB}}(k,m),\cdots,y_{N_sm}^{\mathrm{SSB}}(k,m)]^T$ in the $b$-th ($b$=$0,\cdots,B-1$) analysis block with time delay $\tau$, [i.e., $\mathbf{R}_y^b(\tau) = \frac{1}{L}\sum_{t=1}^{L}\mathbf{y}(b\frac{L}{S}+t)\mathbf{y}^T(b\frac{L}{S}+t-\tau)$], $\alpha$ denotes a step-size parameter, $*$ denotes a convolution operator, $L$ is the block length and $S$ is the blockshift rate. Note that the algorithm we used here is a *batch* algorithm, i.e., the algorithm runs by using all the data on each iteration.

**Initial Value Design of Separation Filters.** A suitable initialization of the separation filters helps the convergence of time-domain BSS and mitigates the permutation problem in subband BSS. We can use constraint null beamformers, which makes spatial nulls towards given directions, as the initial value of the separation filters [32]. This is based on the fact that the

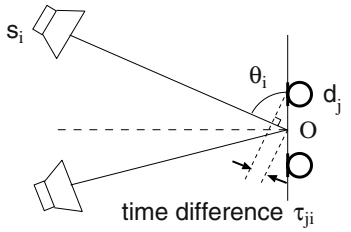


**Fig. 14.4.** Setup of null beamformer.

BSS solution behaves as adaptive beamformers, which form nulls in the jammer directions [33]. Based on this fact, we design null beamformers towards possible sound directions and utilize them as our initial values for the BSS adaptation. Here, we give an example.

Here, we assume a linear microphone array with a known microphone spacing. First, we assume that the mixing matrix $\mathbf{H}(f)$ represents only the time difference of direct sound arrival $\tau_{ji}$ with respect to the midpoint between the microphones (Fig. 14.4). This $\mathbf{H}(f)$ is written in the frequency domain as follows:

$$\mathbf{H}(f) = \begin{bmatrix} \exp(\jmath 2\pi f \tau_{11}) & \cdots & \exp(\jmath 2\pi f \tau_{1N_{sm}}) \\ \vdots & \ddots & \vdots \\ \exp(\jmath 2\pi f \tau_{N_{sm}1}) & \cdots & \exp(\jmath 2\pi f \tau_{N_{sm}N_{sm}}) \end{bmatrix}, \tag{14.6}$$

where $\tau_{ji} = \frac{d_j}{c}\cos\theta_i$, $d_j$ is the position of the $j$-th microphone, $\theta_i$ is the direction of the $i$-th source as an initial value, and $c$ is the speed of sound. Note that these $d_j$ values need not be precise because this $\mathbf{H}(f)$ is used only for the initialization of BSS. It should be also noted that the precise directions of sources, which are not given in a blind scenario, are not required for the initialization. That is the $\theta_i$ values can be very rough approximations, e.g., $\pm 60^\circ$ for the $2 \times 2$ case (i.e., left position or right position, for example).

Then we calculate the inverse of $\mathbf{H}(f)$ at each frequency, $\mathbf{W}(f) = \mathbf{H}^{-1}(f)$ and convert the elements $w_{ij}(f)$ of this $\mathbf{W}(f)$ into the time domain, $\mathrm{w}_{ij}(n) = \mathrm{IFFT}(w_{ij}(f))$. We can use this $\mathrm{w}_{ij}(n)$ as the initial value for time-domain BSS. Then, by applying subband analysis on these $\mathrm{w}_{ij}(n)$, we obtain the initial values of the separation filters in each subband $\mathbf{W}_0^k(m)$ for (14.5).

### 14.3.3 Solving the Permutation and Scaling Problems

Scaling and permutation problems occur in subband BSS in a way similar to that found with frequency-domain BSS, i.e., the estimated source signal components are recovered with a different order and gain in the different frequencies. Thanks to the initial value mentioned in the previous subsection, we can mitigate the permutation problem, however, it sometimes still occurs.

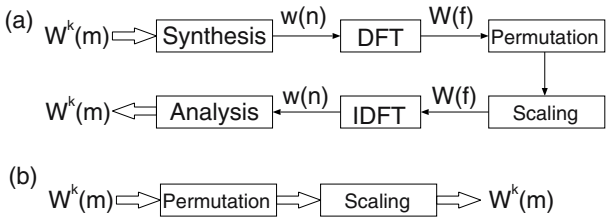


**Fig. 14.5.** Flows to solve the permutation and scaling problems (a) in the frequency domain and (b) in the subband domain.

In order to solve the permutation problem, we can also employ an adaptive-beamformer-like characteristic of the BSS solution [34]. We can solve the problem by reordering the row of estimated separation filters $\mathbf{W}^k(m)$ so that the null of the directivity pattern obtained by $\mathbf{W}^k(m)$ is sorted and forms a null toward almost the same direction in all subbands. This procedure is easily realized by looking at the directivity pattern of $\mathbf{W}(f)$ in the frequency domain [Fig. 14.5 (a)] [34]. We can also solve the permutation problem by sorting the row of the estimated separation filters $\mathbf{W}^k(m)$ so that the cross-correlation of separated signals $y_i^{\mathrm{SSB}}(k,m)$ in adjacent subbands is maximized [12], [14]. With the correlation method, we can solve the problem in the subband domain [Fig. 14.5 (b)].

For the scaling problem, we can also use the directivity pattern calculated with the separation filters [35], that is, we normalize the row of the estimated separation filters $\mathbf{W}^k(m)$ so that the gains and phases of the target directions become 0 dB and 0, respectively. It can be performed by transforming $\mathbf{W}^k(m)$ into the frequency domain [Fig. 14.5 (a)] [34]. The minimal distortion principle [17] or the projection back method [10] can also be employed for $\mathbf{W}(f)$ to solve the scaling problem [32], e.g., $\mathbf{W}(f) \leftarrow \mathrm{diag}[\mathbf{W}^{-1}(f)]\mathbf{W}(f)$. We can also solve the problem naively by normalizing the separation filters $\mathbf{W}^k(m)$ so that each component $w_{ij}^k(m)$ has the same power as the corresponding component of null beamformers $\mathbf{W}_{\mathrm{NBF}}^k(m)$, which have nulls in the jammer directions. This can be executed in the subband domain [Fig. 14.5 (b)].

We can combine some solutions mentioned above. Here is one of the solutions to the permutation and scaling problems which we employed:

i) Synthesize $\mathbf{W}^k(m)$ to obtain $\mathbf{W}(n)$ in the time-domain, then obtain $\mathbf{W}(f)$ using a discrete Fourier transform (DFT).
ii) Estimate signal directions $\theta_i$ $(i = 1, \cdots, N_{sm})$ from the directivity gain pattern of $\mathbf{W}(f)$ [35]. When $N_{sm} \geq 3$, it is recommended that signal directions be estimated analytically from $\mathbf{W}(f)$ [36].
iii) Solve the permutation problem by reordering the $\mathbf{W}(f)$ row so that the $\theta_i$ values are sorted.

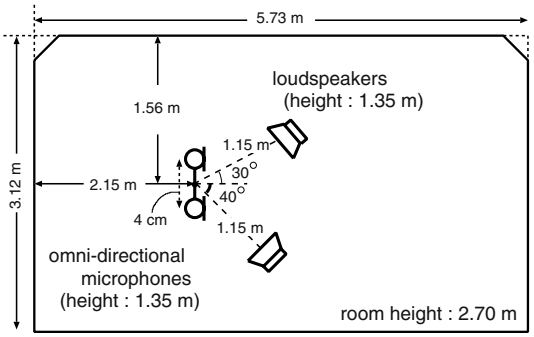


**Fig. 14.6.** Layout of room used in experiments. $T_{\rm R}$ =300 ms.

iv) Make null beamformers by using (14.6) with the estimated $\theta_i$ in step ii), and by calculating $\mathbf{W}(f) = \mathbf{H}^{-1}(f)$. We call this null beamformer $\mathbf{W}_{\rm NBF}(f)$ and use it to solve the scaling problem.
v) Calculate the inverse DFT of $\mathbf{W}_{\rm NBF}(f)$ and perform subband analysis to obtain $\mathbf{W}^k_{\rm NBF}(m)$.
vi) Rescale $\mathbf{W}^k(m)$ so that $||w^k_{ij}(m)|| = ||w^k_{{\rm NBF}ij}(m)||$, where $||x(m)||$ means $\sum_m^{Q_k} x^2(m)$ and $Q_k$ is the separation filter length in the $k$-th subband.

## 14.4 Basic Experiments for Subband BSS

### 14.4.1 Experimental Setup

In order to confirm the performance of subband BSS, we undertook separation experiments using speech data convolved with impulse responses measured in a real environment for a $2 \times 2$ case. The impulse responses were measured in the room shown in Fig. 14.6. The reverberation time $T_{\rm R}$ was 300 ms. Since the sampling rate was 8 kHz, 300 ms corresponds to 2400 taps. As the original speech, we used two sentences spoken by two male and two female speakers. Investigations were carried out for six combinations of speakers. Each mixed speech signal was about eight seconds long. We used the first three seconds of the mixed data for learning, and we separated the entire eight second data.

To evaluate the performance, we used the signal-to-interference ratio (SIR), defined as

$$\mathrm{SIR}_i = \mathrm{SIR}_{{\rm O}i} - \mathrm{SIR}_{{\rm I}i}, \tag{14.7}$$

$$\mathrm{SIR}_{{\rm O}i} = 10 \log \frac{\sum_n \mathrm{y}^2_{i{\rm s}_i}(n)}{\sum_n (\sum_{j \neq i} \mathrm{y}_{i{\rm s}_j}(n))^2},$$

$$\mathrm{SIR}_{{\rm I}i} = 10 \log \frac{\sum_n \mathrm{x}^2_{k{\rm s}_i}(n)}{\sum_n (\sum_{j \neq i} \mathrm{x}_{k{\rm s}_j}(n))^2},$$

where $\mathrm{y}_{i\mathrm{s}_j}$ is the output of the whole system at $\mathrm{y}_i$ when only $\mathrm{s}_j$ is active, and $\mathrm{x}_{k\mathrm{s}_i} = \mathrm{h}_{ki} * \mathrm{s}_i$ ($*$ is a convolution operator, $k = i$ in our experiments). SIR is the ratio of a target-originated signal to jammer-originated signals.

### 14.4.2 Subband System

For subband analysis and synthesis, we used a polyphase filterbank [25] with single sideband (SSB) modulation/demodulation [28], which we mentioned in Section 14.3.1. Here, the number of subbands $N$ was 64 and the down-sampling rate $R$ was 16 ($R = \frac{N}{4}$). We decided this number of subbands $N$ so that the down-sampling rate of subband BSS corresponded to that of conventional frequency-domain BSS (see Section 14.4.3) of frame size $T = 32$ with the half frame-shift.

For the time-domain algorithm used in subband BSS, we estimated separation filters $w_{ij}^k(m)$ of 64 and 128-taps in each subband. The step-size for adaptation $\alpha$ was 0.02 and the number of blocks $B$ was fixed at 20 for three seconds of speech. We adopted $\theta_i$=±60° as the initial values of the separation filters (see Section 14.3.2).

### 14.4.3 Conventional Frequency-Domain BSS

The frequency-domain BSS iteration algorithm was a natural gradient based algorithm

$$\Delta\mathbf{W}_i(f) = \eta\left[\mathrm{diag}\left(\langle\Phi(\mathbf{y})\mathbf{y}^{\mathrm{H}}\rangle\right) - \langle\Phi(\mathbf{y})\mathbf{y}^{\mathrm{H}}\rangle\right]\mathbf{W}_i(f),$$

where $\mathbf{y}$=$\mathbf{y}(f,m)$, superscript H denotes a conjugate transpose and $\langle x(m)\rangle$ denotes the time average with respect to time $m$: $\frac{1}{L_m}\sum_{m=0}^{L_m-1} x(m)$. Subscript $i$ is used to express the value of the $i$-th step in the iterations, $\eta$ is a step-size parameter, and $\Phi(\cdot)$ is a nonlinear function. As the nonlinear function $\Phi(\cdot)$, we used $\Phi(\mathbf{y}) = \tanh(g \cdot \mathrm{abs}(\mathbf{y}))e^{j\arg(\mathbf{y})}$ [37], where $g$ is a parameter to control the nonlinearity and we utilized $g = 100$. As the initial value of the separation matrix, we utilized $\mathbf{W}(f) = \mathbf{H}^{-1}(f)$ with $\theta_i$=±60° (see Section 14.3.2).

We fixed the frame shift at half the STFT frame size $T$, so that the number of samples in the time-frequency domain were the same. To solve the scaling and permutation problems, we also used the beamforming approach [34]: first, from the directivity pattern obtained by $\mathbf{W}(f)$ we estimated the source directions and reordered the row of $\mathbf{W}(f)$ so that the directivity pattern formed a null toward the same direction in all frequencies, then we normalized the row of $\mathbf{W}(f)$ so that the gains of the target directions became 0 dB.

It should be noted that we used the time-average of $\mathbf{y}(f,m)$ of three seconds for adaptation, i.e., we used a *batch* algorithm. It should also be noted that if we fix the data length and frame shift at half the frame size, the number of samples $L_m$ of sequences $\mathbf{y}(f,m)$ in each frequency depends on the frame size $T$: roughly speaking, $L_m \propto$ (data length)$/T$.

Here we utilized the frequency-domain algorithm based on higher order statistics (HOS) despite the fact that we are using a time-domain algorithm relied on second order statistics (SOS). The performance of time-domain BSS based on SOS has already been compared with that based on HOS [38], and it was shown that the performance is not significantly different when we use an adaptive-beamformer-like initial value. It has also been shown [39] that the decorrelation-based algorithm and the fourth order moment-based algorithm perform identically for speech. Therefore, we consider that we will see the same tendency as that shown by our results if we compare time- subband- and frequency-domain BSS using HOS/SOS only.

### 14.4.4 Conventional Fullband Time-Domain BSS

We also examined fullband time-domain BSS. The algorithm was the same as that used in subband BSS, i.e., (14.5). In this case, the output signal vector $\mathbf{y}(n)$ consisted of the signals in the time domain $[\mathrm{y}_1(n), \cdots, \mathrm{y}_{N_{sm}}(n)]^T$. We used values of $\alpha = 0.002$ and $B = 20$. To obtain the initial condition of the separation filters, we also utilized $\mathbf{W}(f) = \mathbf{H}^{-1}(f)$ with $\theta_i = \pm 60^\circ$ and converted it into the time domain (see Section 14.3.2).

In fullband time-domain BSS, the output speech signals are distorted and whitened (see [40] and Section 14.4.6). We evaluated the SIR values after compensating for this whitening effect [32].

### 14.4.5 Results

**Subband System Evaluation.** To evaluate the subband analysis-synthesis system, we measured the signal-to-distortion ratio (SDR), which is defined as

$$\mathrm{SDR} = 10 \log \frac{\sum_n^{L_\delta} \mathrm{b}^2(n-D)}{\sum_n^{L_\delta} \{\mathrm{b}(n-D) - \mathrm{a}(n)\}^2} \ [\mathrm{dB}], \tag{14.8}$$

where the system input $\mathrm{b}(n) = \delta(n - \frac{L_\delta}{2})$, $L_\delta$ is the length of the delta function, $D$ is the delay caused by low-pass filters (LPF) in the analysis and synthesis stages, and $\mathrm{a}(n)$ is the output (impulse response) of the subband analysis-synthesis system. The SDR was 59.2 dB. This distortion caused by subband analysis and synthesis can be ignored because the separation performance SIR is at most 15 dB (see Fig. 14.7), and thus masks this distortion.

**Separation Performance of Subband BSS.** In order to confirm the superiority of subband BSS, we compared the separation performance of subband BSS with that of frequency-domain BSS and time-domain BSS.

Figure 14.7 shows the separation result SIR and the value of the average correlation coefficient between source signals $\mathrm{CC}(N) = \frac{1}{N} \sum_{k=1}^{N} |r_k|$, where

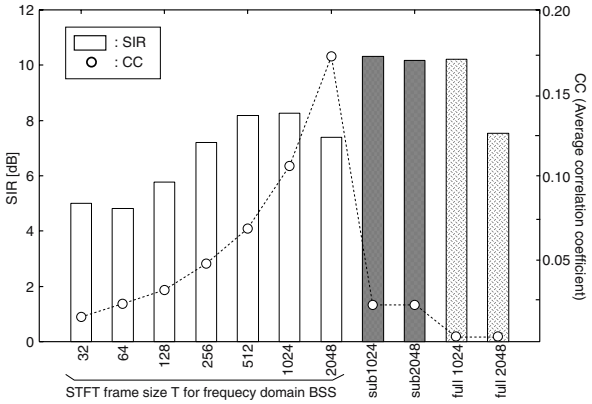


**Fig. 14.7.** Separation performance of frequency-domain BSS (white bars), subband BSS (black bars) and fullband time-domain BSS (gray bars). "CC": average correlation coefficient. Adaptation data length=3 s and separated data length=8 s. $T_{\mathrm{R}} = 300$ ms.

$N$ is the number of subbands for subband BSS or number of frequencies for frequency-domain BSS and $r_k$ is the correlation coefficient between source signals of a $k$-th frequency/subband.

For frequency-domain BSS, the parameter was the STFT frame size $T$. In Fig. 14.7, $T$ is shown by the horizontal axis. For subband BSS, we used separation filters $w_{ij}^k(m)$ of 64 and 128-taps in each subband; this corresponds to 1024 and 2048-taps in a fullband, respectively. In Fig. 14.7, they are shown as "sub1024" and "sub2048", respectively. Our $N = 64$ subbands with decimation $R = 16$ corresponds to $T = 32$ in frequency-domain BSS with regard to down-sampling rate. The number of learning data samples in the time-frequency domain was the same for subband and frequency-domain BSS.

With frequency-domain BSS, although we should use long frame to handle the reverberation, CC becomes large and the independent assumption seems to collapse as frame size $T$ becomes large. This is because the number of samples in each frequency becomes small. Therefore, the performance degraded when we used separation filters of 2048-taps (i.e., frame size $T = 2048$). Please note that the adaptation data length was three seconds and the half frame-shift was utilized.

With fullband time-domain BSS ("full1024" and "full2048" in Fig. 14.7), on the other hand, the CC was very small and we obtained a good result when the separation filter length was 1024. However, when we employed a separation filter length of 2048, it became difficult to estimate the separation filters and the performance degraded. The performance for various separation filter lengths with fullband time-domain BSS can be seen in [32].

By contrast, we achieved better separation performance with subband BSS even when we estimated separation filters of 2048-taps. Moreover, with subband BSS, we were able to confirm that the CC value was sufficiently small. From the CC values, we can say that the independence assumption held well in subband BSS. Another possible reason for the superior performance of subband BSS is that the permutation problem does not arise in the subbands. This point is discussed in the next subsection.

### 14.4.6 Discussion

Using subband BSS, we can maintain the number of samples in each subband and obtain better separation performance. Using one second of speech as adaptation data, we still obtained acceptable separation performance: SIR = 7.47 dB for $T_{\mathrm{R}} = 300$ ms. If the adaptive data length is sufficiently long, the same performance would be obtained by time-domain BSS, frequency-domain BSS, and subband BSS. Our experimental results showed that subband BSS works effectively when the adaptation data length is short.

Moreover, using subband BSS, we obtained separated signals with less whitening effect than when using fullband time-domain BSS. When we use the usual time-domain BSS algorithm, the output signal spectrum is flattened [40]. This is because we remove the time dependence of the speech signals. These whitened speech signals sound unnatural. In contrast, because this whitening effect is limited in each subband, it can be diminished by subband BSS. Figure 14.8 shows an example of separated speech obtained with time-domain BSS and subband BSS. The separated signal is whitened using time-domain BSS, while the shape of the spectrum holds well using subband BSS.

Furthermore, although we did not face the permutation problem due to the initialization with null beamformers, this problem occurs in frequency-domain BSS and subband BSS in general; the spectral components of sources are recovered in a different order at different frequencies/subbands. This makes the time domain reconstruction of separated signals difficult. However, this problem is less serious in subband BSS than in frequency-domain BSS. This is because the permutation problem does not occur in each subband as the separation procedure is executed in each subband. Therefore, we face a smaller number of permutation problems than with frequency-domain BSS. In particular, subband BSS encounters very few permutation problems in low frequency bands, where it is difficult to solve the problems with frequency-domain BSS [15]. Moreover, we can use a wider band signal than frequency-domain BSS to solve the permutation problem in between subbands. Therefore, we can use more information on separated signals and separation filters, and can solve the problem more easily than in frequency-domain BSS.

Finally, we discuss the computational cost. Because the calculation of convolution and correlation in the time domain (14.5) is expensive, we calculate them in the frequency domain. As discussed in [20], [21], we can reduce

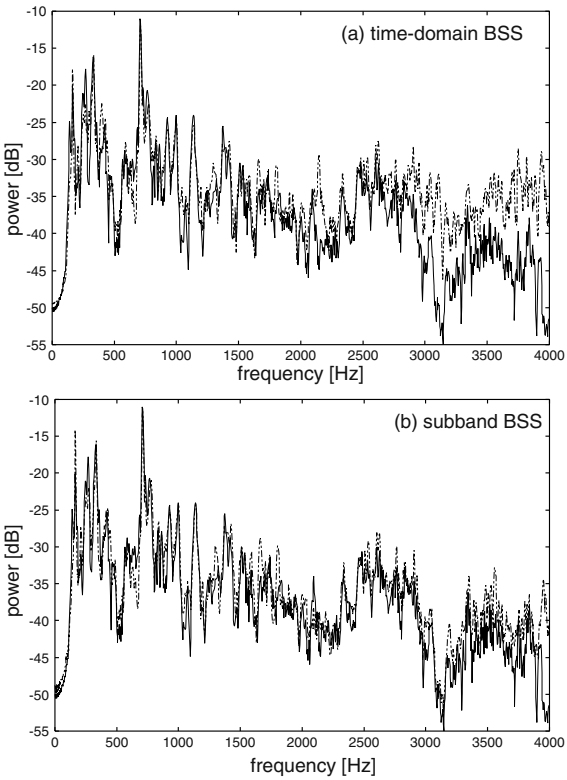


**Fig. 14.8.** Example spectra of a separated signal with (a) time-domain BSS and (b) subband BSS (broken lines). The solid lines show the spectrum of the original speech.

the computational cost by using subband processing. When we consider the decimation $R$, the computational cost for $N$ subbands per time is reduced to about $(N/2+1)/(R \times R)$ times that of fullband time-domain BSS. As $R = N/4$ in our case, we can reduce the computational cost by about $2/R$.

## 14.5 Frequency-Appropriate Processing for Further Improvement

Subband BSS allows us to use different separation methods to estimate the separation filter for different subbands. By exploiting this advantage, in this section, we concentrate on low frequency bands for speech separation.

With speech separation, the SIR is generally worse in low frequency bands as shown in Fig. 14.9, which plots the SIR values of separated signals for each subband. One reason for the poor performance at low frequencies is that the impulse response is usually longer (see Fig. 14.10) and therefore it is harder

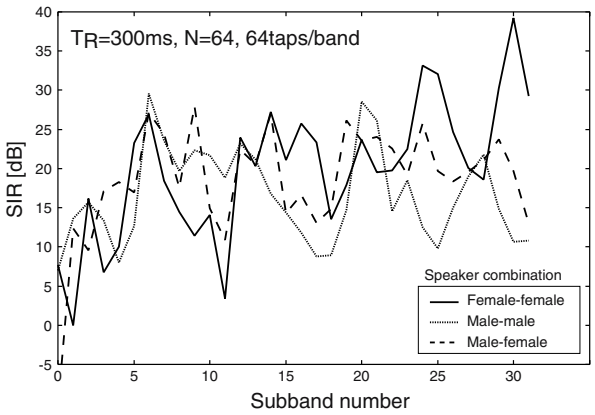


**Fig. 14.9.** SIR of separated signals in each subband. We can see that the SIR is poor in low frequency bands for every speaker combination.

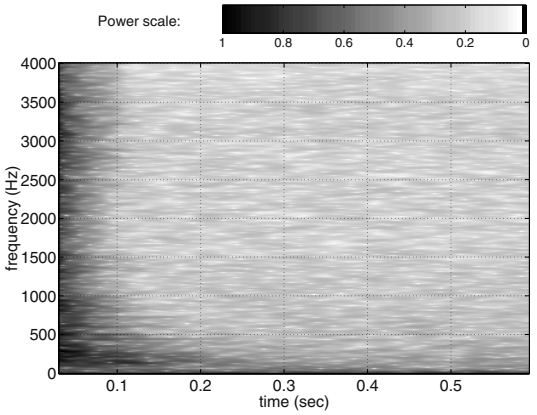


**Fig. 14.10.** Spectrogram example of a room acoustic impulse response. Black indicates high power and white indicates low power. We can see that the reverberation is longer at low frequencies than at high frequencies.

to separate signals in low frequency bands than in high frequency bands. Moreover, since speech signals have high power in low frequency bands, the performance in these bands dominates the overall speech signal separation performance. Therefore, it is important for speech separation to improve the separation performance in low frequency bands to obtain better overall separation performance.

### 14.5.1 Longer Separation Filters in Low Frequency Bands

One possible way to improve the SIR in low frequency bands is to estimate longer separation filters in these bands in order to cover the long reverber-

**Table 14.1** Separation performance of subband BSS. (A)-(F) the overlap-blockshift was executed only for low frequency bands 0-5, and (G) and (H) the overlap-blockshift was executed for *all* subbands. $N = 64$.

| | # of taps | | SIR [dB] | | |
|---|---|---|---|---|---|
| | band 0-5 | band 6-32 | no-overlap | overlap (x2) | overlap (x4) |
| (A) | 32 | 32 | 6.0 | | |
| (B) | 64 | 32 | 9.9 | 9.8 | |
| (C) | 128 | 32 | 9.5 | 10.1 | 10.4 |
| (D) | 64 | 64 | 10.3 | 10.8 | 10.7 |
| (E) | 128 | 64 | 10.5 | 11.4 | 12.2 |
| (F) | 128 | 128 | 10.1 | 11.0 | 11.7 |
| (G) | 64 | 64 | 10.3 | 10.7 | 10.7 |
| (H) | 128 | 128 | 10.1 | 11.2 | 12.2 |

ation. If the length of the separation filters is insufficient, we cannot reduce reverberant components of interferences that are longer than the filters and we obtain poor SIR [24].

We therefore employ longer separation filters for low frequency bands (bands 0-5). Figure 14.10 shows that the reverberation is long below about 600 Hz. Therefore, we used long filters for these frequency bands. The column labelled "no-overlap" in Table 14.1 shows the separation performance for each separation filter length condition.

In Table 14.1 (A)-(C), we used a 32-tap separation filter for high frequency bands, and we changed the filter length for low frequency bands (bands 0-5). We can see that a 32-tap long separation filter cannot achieve good performance [see Table 14.1 (A)]. This is conceivable that it cannot cover the reverberation in low frequency bands. When we used long separation filters only in low frequency bands [Table 14.1 (B)], the separation performance was greatly improved. However, when we used 128-taps in low frequency bands, the separation performance degraded [see Table 14.1 (C)]. Figure 14.11 shows the SIR for cases (A) - (C). We can see that the performance of (C) is worse than (B). This is attributed to the fact that the number of samples in each subband is too small to allow us to estimate a 128-tap separation filter precisely. The proposal in the next section (Section 14.5.2) will overcome this problem.

### 14.5.2 Overlap-Blockshift in Low Frequency Bands

Another possible way to improve the SIR in low frequency bands is to utilize a fine overlap-blockshift in the time-domain BSS stage. Using the fine overlap-blockshift, we can increase outwardly the number of samples in each

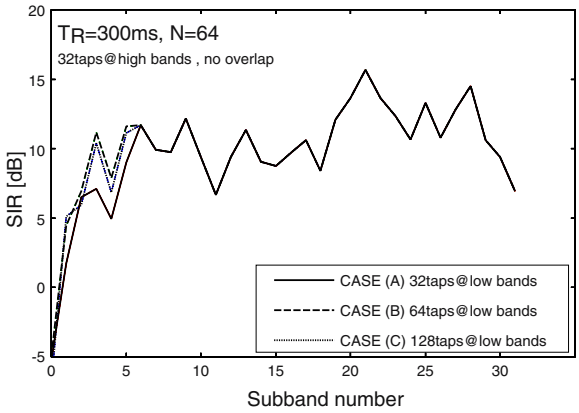


**Fig. 14.11.** Effect of filter length for low frequency bands.

subband, and can estimate the separation filters more precisely. Since our time-domain BSS algorithm (14.5) divides signals into $B$ blocks to utilize the non-stationarity of signals, we can divide signals into blocks with an overlap, as long as the non-stationarity is expressed among blocks. It should be noted that this overlap-blockshift is executed in the separation stage, i.e., after the decimation for subband analysis.

In Table 14.1 [(B)-(F)], the columns show the SIR obtained by the overlap-blockshift only for low frequency bands (bands 0-5). "Overlap ($\times 2$)" and "overlap ($\times 4$)" means that the blockshift rate $S = 2$ and 4 in (14.5), respectively. Table 14.1 [(B)-(F)] show that when we used the overlap-blockshift only for low frequency bands, we obtained better separation performance. With a fourfold overlap-blockshift for (E), we were able to estimate the separation filters of 128-taps in low frequency bands, and we obtained the best separation performance (underlined in Table 14.1). Figure 14.12 shows the effect of the fine overlap-blockshift in low frequency bands.

### 14.5.3 Discussion

Even when we used 128-taps for all the frequency bands [(F) in Table 14.1], the performance was no better than when we used 128-taps only for the low frequency bands [(E) in Table 14.1]. Figure 14.13 shows the SIR in each subband for (E) and (F). We can see that the use of the long separation filters is not so effective in the high frequency bands. Sometimes, short filters achieve better separation performance than long filters in the high frequency bands. We can say that the employment of long separation filters only in low frequency bands is enough for the separation.

Furthermore, when the overlap-blockshift was used in all subbands [see (G) and (H) in Table 14.1], the increase in SIR was very small compared with the SIR for (D) and (F) in Table 14.1. Figure 14.14 shows the improvement

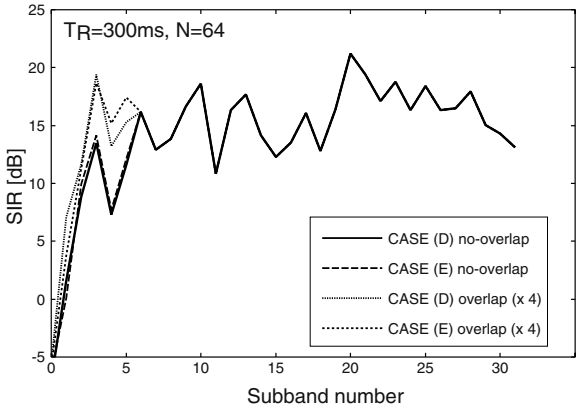


**Fig. 14.12.** Effect of overlap-blockshift only in low frequency bands.

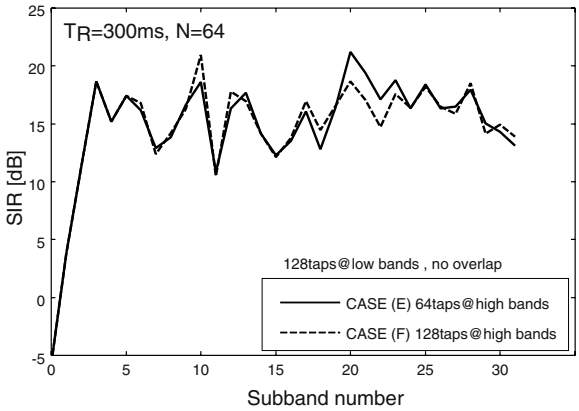


**Fig. 14.13.** Example of SIR in each subband when we use a long filter in all frequency bands.

in separation performance provided by the overlap-blockshift. The overlap-blockshift is also effective in high frequency bands. However, the contribution of the improvement to SIR in the high frequency bands is not significant for the whole performance [see (F) and (H) in Table 14.1]. This is because the original power of the high frequency components of the speech signal is smaller than that of the low frequency components. Therefore, we can conclude that the use of a fine overlap-blockshift only in low frequencies is sufficient to obtain improved performance.

By using long separation filters and the fine overlap-blockshift technique only in low frequency bands, we can efficiently separate convolutive mixtures of speech. Such frequency-dependent processing is impossible with time-domain BSS and intricate with frequency-domain BSS. Moreover, we can save the computation cost without degrading the separation performance by lim-

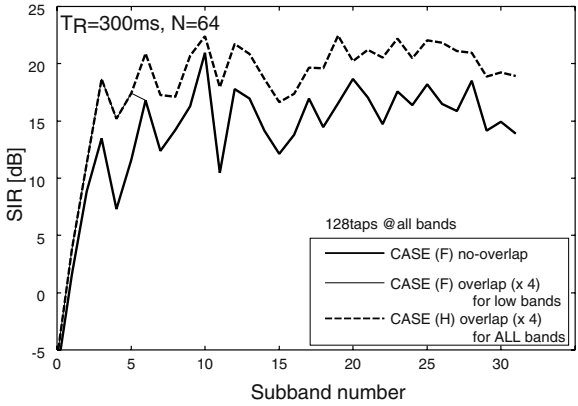


**Fig. 14.14.** Example of SIR in each subband obtained with the overlap-blockshift in all subbands.

iting the use of long separation filters and the fine overlap-blockshift only to low frequency bands.

There could be other ways to improve the separation performance. For instance, we may be able to use different microphone pairs with appropriate spacing for each subband. From a beamforming point of view, the resolution of a spatial cancellation is related to the frequency. If the microphone spacing is greater than half the wavelength, spatial aliasing occurs. This tends to happen at high frequencies. On the other hand, if the spacing is too small, the phase and amplitude difference between observations at low frequency becomes too small and therefore, it becomes difficult to achieve good performance. That is, the small phase difference between the observations at the microphones is also a reason for the poor performance in low frequency bands. A low frequency generally prefers a long spacing and a high frequency likes a short spacing [41]. In this chapter, we considered the case of $N_m$ microphones whose number and spacing are fixed and ignored the multiple spacing microphone case. However, if we could configure the microphone spacing according to frequency, we would obtain better performance.

## 14.6 Conclusions

In this chapter, subband processing was applied to BSS for convolutive mixtures of speech. The subband-based BSS approach offers a compromise between the time-domain technique, which is usually difficult and slow with many separation filter coefficients to estimate, and a frequency domain technique, which has difficulty estimating statistics when the adaptation data length is insufficient. Our proposed subband BSS can maintain a sufficient number of samples to estimate the statistics in each subband and estimate

a separation filter long enough to cover the reverberation. We confirmed the effectiveness of subband BSS experimentally.

Furthermore, making good use of subband processing, i.e., employing an appropriate separation method for each frequency band, we showed that we can improve the separation performance with long separation filters and the overlap-blockshift technique only in low frequency bands. Subband BSS is a powerful separation tool when the source signals $s_i$ or the impulse response of the system $h_{ji}$ have different characteristics in different frequency bands.

## Acknowledgments

We wish to thank Dr. H. Saruwatari for his valuable discussions and for providing the impulse responses used in the experiments. We also thank Dr. Y. Haneda, Mr. A. Nakagawa, and Mr. S. Sakauchi for their help with the SSB filterbank, and Mr. R. Aichner and Mr. T. Nishikawa for detailed discussions on time-domain blind source separation.

## References

1. S. Haykin, *Unsupervised Adaptive Filtering*. John Wiley & Sons, 2000.
2. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
3. T. W. Lee, *Independent Component Analysis – Theory and Applications*. Kluwer Academic Publishers, 1998.
4. S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, 1997, pp. 101–104.
5. K. Torkkola, "Blind separation of delayed and convolved sources," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed., vol. 1, pp. 321–375, John Wiley & Sons, 2000.
6. M. Kawamoto, K. Matsuoka, and N. Ohnishi, "A method of blind separation for convolved non-stationary signals," *Neurocomputing*, vol. 22, pp. 157–171, Nov. 1998.
7. H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: a unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds., pp. 255–293, Kluwer Academic Publishers, 2004.
8. S. C. Douglas, "Blind separation of acoustic signals," in *Microphone Arrays: Techniques and Applications*, M. Brandstein and D. B. Ward, Eds., pp. 355–380, Springer-Verlag, 2001.
9. P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, Nov. 1998.
10. S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. ICA*, 1999, pp. 365–370.

11. M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," in *Proc. IEEE ICASSP*, 2000, pp. 1041–1044.
12. J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation," in *Proc. ICA*, 2000, pp. 215–220.
13. S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 109–116, Mar. 2003.
14. N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, pp. 1–24, Oct. 2001.
15. H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust approach to the permutation problem of frequency-domain blind source separation," in *Proc. IEEE ICASSP*, 2003, pp. 381–384.
16. K. Rahbar and J. P. Reilly, "A new fast-converging method for BSS of speech signals in acoustic environments," in *Proc. IEEE WASPAA*, 2003, pp. 21–24.
17. K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA*, 2001, pp. 722–727.
18. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Blind source separation for convolutive mixtures of speech using subband processing," in *Proc. SMMSP (International Workshop on Spectral Methods and Multirate Signal Processing)*, 2002, pp. 195–202.
19. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband based blind source separation for convolutive mixtures of speech," in *Proc. IEEE ICASSP*, 2003, pp. 509–512.
20. J. Huang, K.-C. Yen, and Y. Zhao, "Subband-based adaptive decorrelation filtering for co-channel speech separation," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 402–406, July 2000.
21. F. Duplessis-Beaulieu and B. Champagne, "Fast convolutive blind speech separation via subband adaptation," in *Proc. IEEE ICASSP*, 2003, pp. 513–516.
22. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband based blind source separation with appropriate processing for each frequency band," in *Proc. ICA*, 2003, pp. 499–504.
23. S. Araki, S. Makino, R. Aichner, T. Nishikawa, and H. Saruwatari, "Subband-based blind separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Processing*, submitted.
24. R. Mukai, S. Araki, H. Sawada, and S. Makino, "Evaluation of separation and dereverberation performance in frequency domain blind source separation," *Acoustical Science and Technology*, vol. 25, pp. 119–126, Mar. 2004.
25. M. R. Portnoff, "Implementation of the digital phase vocoder using the fast Fourier transform," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 24, pp. 243–248, June 1976.
26. N. Grbic, X.-J. Tao, S. E. Nordholm, and I. Claesson, "Blind signal separation using overcomplete subband representation," *IEEE Trans. Speech Audio Processing*, vol. 9, pp. 524–533, July 2001.
27. S. L. Gay and R. J. Mammone, "Fast converging subband acoustic echo cancellation using RAP on the WE DSP16A," in *Proc. IEEE ICASSP*, 1990, pp. 1141–1144.

28. R. Crochiere and L. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
29. P. L. Chu, "Weaver SSB subband acoustic echo canceller," in *Proc. IWAENC*, 1993, pp. 173–176.
30. S. Makino, J. Noebauer, Y. Haneda, and A. Nakagawa, "SSB subband echo canceller using low-order projection algorithm," in *Proc. IEEE ICASSP*, 1996, pp. 945–948.
31. T. Nishikawa, H. Saruwatari, and K. Shikano, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals*, vol. E86-A, pp. 846–858, Apr. 2003.
32. R. Aichner, S. Araki, S. Makino, T. Nishikawa, and H. Saruwatari, "Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," in *IEEE International Workshop on Neural Networks for Signal Processing*, 2002, pp. 445–454.
33. S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1157–1166, 2003.
34. 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. IEEE ICASSP*, 2000, pp. 3140–3143.
35. H. Saruwatari, S. Kurita, and K. Takeda, "Blind source separation combining frequency-domain ICA and beamforming," in *Proc. IEEE ICASSP*, 2001, pp. 2733–2736.
36. H. Sawada, R. Mukai, and S. Makino, "Direction of arrival estimation for multiple source signals using independent component analysis," in *Seventh International Symposium on Signal Processing and its Applications*, 2003, vol. 2, pp. 411–414.
37. H. Sawada, R. Mukai, S. Araki, and S. Makino, "Polar coordinate based nonlinear function for frequency domain blind source separation," in *Proc. IEEE ICASSP*, 2002, pp. 1001–1004.
38. T. Nishikawa, H. Saruwatari, K. Shikano, S. Araki, and S. Makino, "Multistage ICA for blind source separation of real acoustic convolutive mixture," in *Proc. ICA*, 2003, pp. 523–528.
39. S. Van Gerven, D. Van Compernolle, L. Nguyen Thi, and C. Jutten, "Blind separation of sources: A comparative study of a 2nd and a 4th order solution," in *Signal Processing VII, Theories and Applications*, M. J. J. Holt, C. F. N. Cowan, P. M. Grant, and W. A. Sandham, Eds., Elsevier, pp. 1153–1156, 1994.
40. X. Sun and S. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," in *Proc. ICA*, 2001, pp. 59–64.
41. H. Sawada, S. Araki, R. Mukai, and S. Makino, "Blind source separation with different sensor spacing and filter length for each frequency range," in *IEEE International Workshop on Neural Networks for Signal Processing*, 2002, pp. 465–474.