Neurocomputing 73 (2009) 304–314

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

# A Bark-scale filter bank approach to independent component analysis for acoustic mixtures

Hyung-Min Park [a,*], Sang-Hoon Oh [b], Soo-Young Lee [c]

[a] Department of Electronic Engineering, Sogang University, Seoul 121-742, Republic of Korea
[b] Department of Information Communication Engineering, Mokwon University, Daejeon 302-729, Republic of Korea
[c] Department of Bio and Brain Engineering, and Brain Science Research Center, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Republic of Korea

ARTICLE INFO

*Article history:*
Received 11 July 2007
Received in revised form
14 July 2009
Accepted 24 August 2009
Communicated by X. Wu
Available online 30 September 2009

*Keywords:*
Independent component analysis
Filter banks
Blind source separation
The Bark scale

ABSTRACT

Uniform filter bank approach can be considered to perform independent component analysis (ICA) for convolved mixtures. It achieves better separation performance than the frequency domain approach and gives faster convergence speed with less computational complexity than the time domain approach. However, when the uniform filter bank approach is applied to natural audio signals, it provides slower convergence for low frequency subbands and gives inferior separation performance for high frequency subbands. Owing to spectral characteristics of natural signals, we present a filter bank approach that employs a Bark-scale filter bank. In the Bark-scale filter bank, low frequency region is minutely divided, whereas high frequency region has much wider subbands. The Bark-scale filter bank approach shows faster convergence speed than the uniform filter bank approach because it has more whitened inputs in the low frequency subbands. It also improves the separation performance as it has enough data to train adaptive parameters exactly in the high frequency subbands.


## 1. Introduction

Independent component analysis (ICA) is a signal processing method to express multivariate data as linear combinations of statistically independent random variables [1–3]. Resorting to higher order statistics, ICA has achieved impressive performance in many applications such as speech enhancement, telecommunications, medical signal processing, and feature extraction [4–7]. However, ICA for acoustic mixtures still remains as a challenging problem due to very complex reverberation involved with real-world acoustic mixing environments. To deal with convolutive mixtures of audio signals, some of the ICA approaches for instantaneous mixtures have been traditionally extended in the time domain [8] and the frequency domain [9–11]. Filter bank approaches have been proposed to overcome disadvantages of the time and frequency domain approaches [12–15]. A filter bank approach proposed by Park et al. [12,13] does not have performance limitation of the frequency domain approaches, since the ICA algorithm in each subband is basically the same as the time domain approach which is derived from the gradient of the output entropy. Since adaptive filters process subband signals at the decimated rate and the required adaptive filter length is shortened by a factor of the decimation, the number of multiplications in a subband is reduced by a factor of $1/M^2$ where $M$ is the decimation factor. If the number of subbands is $K$, computations are mainly saved by a factor of $K/M^2$ [16,17]. Furthermore, decimation improves convergence of the subband adaptive filters because subband signals are more whitened and the adaptive filter length is shortened [13,16].

However, the uniform filter bank approaches do not consider some properties of input signals. Fig. 1 shows the time-averaged power spectral densities of three natural sounds in the frequency domain which are used in the experiments. The energy of these signals is concentrated in low frequency region and generally decreases more steeply in low frequency region than in high frequency region as the frequency increases. These characteristics are commonly observed for most of the natural audio signals. When a uniform filter bank approach deals with such audio signals, it has more colored input signals in low frequency subbands than in high frequency subbands. This may result in relatively slower convergence for adaptation of ICA networks in the low frequency subbands which contain most of the signal energy. In addition, since audio signals have most energy in low frequency region, data in the high frequency subbands may not be enough to train adaptive filters exactly in the uniform filter bank approach resulting in inferior separation performance.

Several papers have proposed the use of nonuniform filter banks for adaptive filtering instead of uniform filter banks [18–21]. Schulz and Herfet described a mask-based approach, but it may provide inaccurate results because estimated masks

* Corresponding author. Tel.: +82 2 705 8916; fax: +82 2 706 4216.
*E-mail address:* hpark@sogang.ac.kr (H.-M. Park).

0925-2312/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2009.08.009

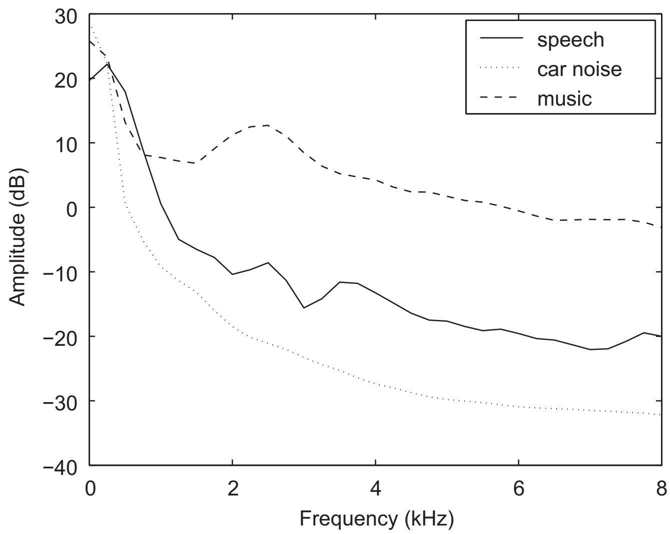



**Fig. 1.** Time-averaged power spectral densities for three natural sounds in the frequency domain.

specify which time–frequency components should be selected or not in a binary way [21]. On the other hand, Rutkowski et al. employed nonuniform filter banks with center frequencies based on estimation of a fundamental frequency [20]. However, the filter banks were constructed from the fundamental frequency of a speech signal with higher energy, so they may not be pertinent to separate other signals with lower energy. In addition, if the fundamental frequency is much changed, the filter banks also need to be changed. In this case, separation filters in subbands may have errors and should converge to another solution. The others were applied to system identification problems which use the least-mean-square (LMS) algorithm [18,19]. In those problems, rapidly changing spectral regions of the system response have narrow subbands, whereas smooth regions have wide subbands. Therefore, a filter bank approach using this nonuniform filter bank provides more uniform convergence speeds in all subbands than the uniform filter bank approach. However, it is not suitable for dealing with natural audio signals because we cannot know *a priori* information about their detailed spectral characteristics, in advance, which correspond to the time-invariant frequency response of an identified system in the system identification problems.

In this paper, a Bark-scale filter bank is considered to improve convergence and separation performance of the filter bank approach when applied to audio signals. It is known that the Bark-scale filter bank has narrow subbands in low frequency region and wide subbands in high frequency region. Also, its frequency response resembles the mammalian cochlea [22]. By employing this filter bank, we can attain faster convergence speed than the uniform filter bank approach because it has more whitened inputs in the low frequency subbands. It also gives better separation performance because it trains adaptive parameters more exactly by using enough data in the high frequency subbands. Although prewhitening of input signals may speed up convergence as shown in [23], the approach was based on time-averaged audio spectral characteristics. However, this could not remove the detailed correlation which temporarily exists only at an instance. Decimation of a filter bank is capable of removing both the averaged and detailed correlations.

The remainder of the paper is organized as follows: Section 2 briefly reviews a filter bank approach to ICA for convolved mixtures. In Section 3, our approach of utilizing nonuniform filter banks is presented. This method is compared with the corresponding uniform filter bank approach through several experiments in Section 4. Finally, some concluding remarks are presented in Section 5.

## 2. Review of a filter bank approach to ICA

Let us consider a set of unknown source signals, $\{s_j(n), j=1,\ldots,N\}$, such that the signals are zero-mean and mutually independent. If mixing involves convolution and time-delays, an observation is

$$x_i(n) = \sum_{j=1}^{N} \sum_{m=0}^{L_m-1} a_{ij}(m)s_j(n-m), \tag{1}$$

where $L_m$ and $a_{ij}(m)$ denote a mixing filter length and a coefficient, respectively [7].

To obtain the independent source signals from these observations, a filter bank approach can be considered as it shows better separation performance than the frequency domain approach and gives faster convergence with less computational complexity than the time domain approach [12,13]. Among the filter bank approaches, oversampled filter banks, where the decimation factor is smaller than the number of analysis filters, accomplish better performance than critically sampled filter banks. The oversampled filter banks can have negligible aliasing when each filter has a high stopband attenuation, so they make it possible to perform adaptive filtering without requiring cross adaptive filters between adjacent bands or distorting reconstructed signals [17,24,25].

Since ICA is performed in the oversampled filter bank, adaptive parameters in each subband can be adjusted without any information from other subbands [12,13]. Thus, the filter bank approach is appropriate for parallel processing. The inputs, which are mixtures of unknown independent signals, are decomposed into subband signals by analysis filters. Then, each subband signal is downsampled by a decimation factor. Since the downsampled signals are still convolved mixtures whose reverberation length has decreased by the decimation factor, a typical ICA algorithm for convolved mixtures can be used to obtain independent components from the downsampled signals at each subband. Here, the unmixing filter length is much shorter than that of the full-band time domain approach. The outputs from the ICA network are expanded, and the original independent signals can be reconstructed from the subband outputs through synthesis filters after fixing scaling and permutation.

As an ICA network in each subband, one may use a feedback architecture [3,26] which is expressed as

$$\begin{aligned} u_i(k,n') = {} & \sum_{m'=0}^{L_a} w_{ii}(k,m')x_i(k,n'-m') \\ & + \sum_{j=1, j\neq i}^{N} \sum_{m'=1}^{L_a} w_{ij}(k,m')u_j(k,n'-m'), \end{aligned} \tag{2}$$

where $k$ and $L_a$ denote subband index and adaptive filter length, respectively. Usually, the length is shortened by a decimation factor, comparing with that of the corresponding adaptive filters in the full-band time domain approach. Here, adaptive filters $w_{ij}(k,m')$ force outputs $u_i(k,n')$ to reproduce the independent subband signals. Among different algorithms to find out the parameters, entropy maximization can provide a simple and biologically plausible adaptive learning algorithm [8,13]:

$$\begin{aligned} & w_{ii\ \mathrm{new}}(k,0) \\ & \quad = w_{ii\ \mathrm{old}}(k,0) + \mu(k,n')[1/w^*_{ii\ \mathrm{old}}(k,0) - \varphi(u_i(k,n'))x_i^*(k,n')], \end{aligned}$$

$$w_{ii\ \mathrm{new}}(k,m') = w_{ii\ \mathrm{old}}(k,m') - \mu(k,n')\varphi(u_i(k,n'))x_i^*(k,n'-m'), \quad m' \neq 0,$$

$$w_{ij\ \mathrm{new}}(k,m') = w_{ij\ \mathrm{old}}(k,m') - \mu(k,n')\varphi(u_i(k,n'))u_j^*(k,n'-m'), \quad i \neq j, \tag{3}$$

where $\mu(\cdot)$ denotes a step size and $\varphi(\cdot)$ is called a score function. Since subband signals are complex numbers, we use the polar-coordinate-based score function [27]:

$$\varphi(u_i(k,n')) = -\frac{\dfrac{\partial p(|u_i(k,n')|)}{\partial |u_i(k,n')|}}{p(|u_i(k,n')|)} \exp(j \cdot \angle u_i(k,n')), \tag{4}$$

where $p(|u_i(k,n')|)$ denotes the probability density function of $|u_i(k,n')|$.

## 3. ICA using Bark-scale oversampled filter banks

As mentioned in Section 1, the bandwidth of a filter bank should increase as the subband center frequency increases to avoid the undesired properties of the uniform filter bank approach to ICA of audio signals. As the bandwidth increases, the decimation factor should decrease to maintain negligible aliasing, so the adaptive filters should have long lengths to cover a certain time range. It is known that convergence of gradient-based algorithms depends on the condition number which is the ratio of the largest to the smallest eigenvalues of the correlation matrix computed from the input vector[1] [28]. By the Bordering theorem [29], the condition number is a monotonically nondecreasing function of filter length. Therefore, an increase of adaptive filter length can only decrease the convergence speed but never improve it [30].

Long adaptive filters should be avoided to prevent slow convergence. Therefore, while doing subband division, which is appropriate for ICA of audio signals, it is necessary to consider the trade-off between the undesired properties of the uniform filter bank approach and large adaptive filter length. By the trade-off, one may have a filter bank which provides similar convergence speeds for all subband adaptive filters.

Considering the trade-off, we use a Bark-scale filter bank [22] as a nonuniform filter bank which is suitable for audio signals. A wavelet filter bank [31] or an equivalent rectangular bandwidth filter bank [32] also can be considered. Although the former has been systematically and efficiently implemented in a form of a critically sampled filter bank, it has too wide high frequency subbands which may cause a slow convergence. In addition, the latter is easily implemented by a Gammatone filter. However, the filter has a wide transition band, which results in a relatively small decimation factor to maintain negligible aliasing. For a Bark-scale filter bank, the linear frequency $\omega$ can be warped into the Bark frequency $\Omega$ by

$$\Omega(\omega) = 6\log\left\{\frac{\omega}{1200\pi} + \left[\left(\frac{\omega}{1200\pi}\right)^2 + 1\right]^{0.5}\right\}, \tag{5}$$

where $\omega$ is the angular frequency in rad/s [33]. By this frequency warping, the filter bank approach which employs Bark-scale filter banks speeds up the convergence of the parameters as it has more whitened inputs in low frequency subbands. It also improves the separation performance because it has enough data to train the adaptive filters exactly in high frequency subbands.

[1] The correlation matrix is defined as $\mathbf{R} = E[\mathbf{x}_i(k,n')\mathbf{x}_i^H(k,n')]$, where the input vector $\mathbf{x}_i(k,n') = [x_i(k,n'), x_i(k,n'-1), \ldots, x_i(k,n'-L_a)]^T$.

In order to design a Bark-scale filter bank, we modified the design methodology used for uniform complex-valued filter bank [34], since it offers systematic formulation by a generalized discrete Fourier transform (GDFT) and the number of total subband samples is remarkably close to the critical decimation [35]. Especially, the methodology enables flexible design using an arbitrary cost function to build a Bark-scale filter bank with just a few modifications as explained in the next paragraph. In the method, the $k$ th subband analysis filter $h(k,n)$ is obtained from the corresponding real-valued low-pass prototype filter $q(k,n)$ by a GDFT,

$$h(k,n) = e^{j2\pi(F_k/F_s)(n-(L_{q_k}-1)/2)} \cdot q(k,n), \quad k = 1,2,\ldots,K, \quad n = 0,1,\ldots,L_{q_k}-1, \tag{6}$$

where $F_k, F_s$, and $L_{q_k}$ are the subband center frequency, sampling frequency, and length of $q(k,n)$, respectively. Complex-conjugate and time-reversed versions of the analysis filters are selected for synthesis filters,

$$f(k,n) = \tilde{h}(k,n) = h^*(k, L_{\max} - n - 1), \tag{7}$$

where $L_{\max}$ is the length of the prototype filter with maximum taps. Contrary to a uniform filter bank, the bandwidth of an analysis or synthesis filter in a subband is different from others. Thus, we have to design different prototype filters for every subbands.

The prototype filters can be designed by iterative least-square algorithm with a cost function which considers reconstructiveness and stopband attenuation. We may follow a detailed procedure presented in [34] except for the following modifications. If aliasing is sufficiently suppressed, the impulse response $t(n)$ of overall filter bank system can be written as a convolution of the analysis and synthesis filters:

$$t(n) = \sum_{k=1}^{K} \frac{1}{M_k} h(k,n) * f(k,n). \tag{8}$$

Here, $M_k$ denotes decimation factor at the $k$ th subband. A measure of the reconstruction error $\varepsilon_1$ can be evaluated by the Euclidean distance between the impulse response, $t(n)$, and a perfect delay, $\delta(n-(L_{\max}-1))$:

$$\varepsilon_1 = \sum_{n=0}^{2L_{\max}-2} [t(n) - \delta(n-(L_{\max}-1))]^2. \tag{9}$$

To measure the energy contained in the stopband of a linear phase prototype filter, $q(k,n)$, a dense grid of frequency points $\{\omega_0, \omega_1, \ldots, \omega_P\}$ covering the whole stopband is used for calculating

$$\varepsilon_{2_k} = \left\| \begin{bmatrix} 1 & \cdots & \cos(w_0 \cdot (L_{q_k}-1)) \\ 1 & \cdots & \cos(w_1 \cdot (L_{q_k}-1)) \\ \vdots & \ddots & \vdots \\ 1 & \cdots & \cos(w_P \cdot (L_{q_k}-1)) \end{bmatrix} \begin{bmatrix} q(k,0) \\ q(k,1) \\ \vdots \\ q(k,L_{q_k}-1) \end{bmatrix} \right\|^2. \tag{10}$$

Therefore, the cost function to be minimized is a combination of filter bank reconstruction error $\varepsilon_1$ and the stopband energies $\varepsilon_{2_k}$ of the prototype filters, which is

$$\varepsilon = \varepsilon_1 + \gamma \sum_{k=1}^{K} \varepsilon_{2_k}, \tag{11}$$

where $\gamma$ is a weighting factor. In order to compute the cost function, we need to determine the subband center frequencies, bandwidths, and decimation factors of the filters in advance. In a Bark-scale filter bank, the center frequencies are related to the bandwidths, and the decimation factors should be selected by taking the bandwidths into account so as to have negligible aliasing and reduced computational complexity. Thus, we first

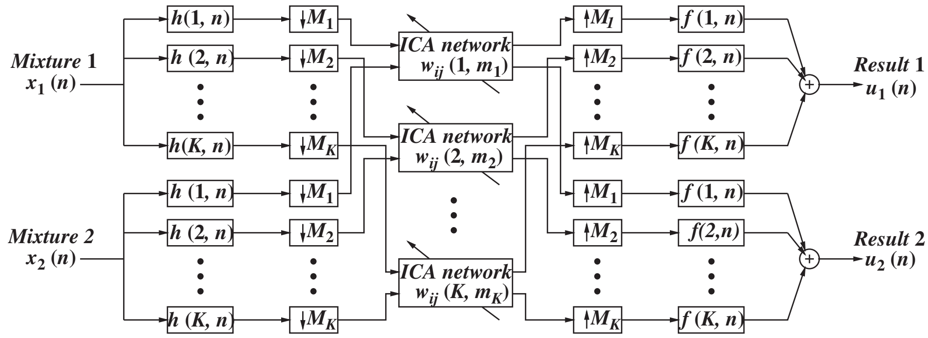



**Fig. 2.** A 2 × 2 network for the nonuniform oversampled filter bank approach to ICA.

choose the number of filters, and then set the center frequencies, bandwidths, and decimation factors for the filters at the same time taking their relationships in consideration.

With a filter bank implemented by this method, a non-critical decimation of subband signals can provide negligible aliasing and make enough information available for successful adaptation in every subband independently [12,13,16,36]. Therefore, an appropriate structure for the nonuniform filter bank approach to ICA with two inputs and two outputs is shown in Fig. 2. The overall procedure is almost the same as that of the uniform filter bank approach. However, since nonuniform filter banks are used, bandwidth of a subband is different from others. Therefore, the decimation factor which makes aliasing error negligible in a subband is also different from others. Similar to the uniform filter bank approach, a feedback architecture can be considered for the ICA network in each subband. The outputs are expressed as

$$u_i(k,n_k) = \sum_{m_k=0}^{L_{a_k}} w_{ii}(k,m_k)x_i(k,n_k-m_k) + \sum_{j=1,j\neq i}^{N}\sum_{m_k=1}^{L_{a_k}} w_{ij}(k,m_k)u_j(k,n_k-m_k), \tag{12}$$

where $u_i(k,n_k), x_i(k,n_k), w_{ij}(k,m_k)$, and $L_{a_k}$ represent the estimated independent signals, input signals, adaptive filter coefficients, and adaptive filter length in the $k$ th subband, respectively. Entropy maximization can be used for the adaptation of filters in each subband, and the learning rules of the adaptive filter coefficients are given as

$$\begin{aligned} &w_{ii\ \mathrm{new}}(k,0) \\ &\quad = w_{ii\ \mathrm{old}}(k,0) + \mu(k,n_k)[1/w^*_{ii\ \mathrm{old}}(k,0) - \varphi(u_i(k,n_k))x_i^*(k,n_k)], \\ &w_{ii\ \mathrm{new}}(k,m_k) \\ &\quad = w_{ii\ \mathrm{old}}(k,m_k) - \mu(k,n_k)\varphi(u_i(k,n_k))x_i^*(k,n_k-m_k), \quad m_k \neq 0, \\ &w_{ij\ \mathrm{new}}(k,m_k) \\ &\quad = w_{ij\ \mathrm{old}}(k,m_k) - \mu(k,n_k)\varphi(u_i(k,n_k))u_j^*(k,n_k-m_k), \quad i \neq j. \end{aligned} \tag{13}$$

We may use Eq. (4) to evaluate the score function $\varphi(u_i(k,n_k))$.

Similar to the uniform filter bank approach, this method also has indeterminacy of estimated outputs up to arbitrary filtering for convolutive mixtures of temporally correlated audio signals. Although entropy maximization attempts to make outputs temporally whitened, it can be avoided by forcing direct filters $w_{ii}(k,m_k)$ to scaling factors while ICA is performed by a feedback architecture.

In addition, scaling and permutation of the estimated outputs in all subbands should be fixed because an ICA network in each subband is independently adapted. Since audio signals usually have time-varying statistical properties, we can employ a modification of the Murata's method which has been successfully used in the uniform filter bank approach [12,13].

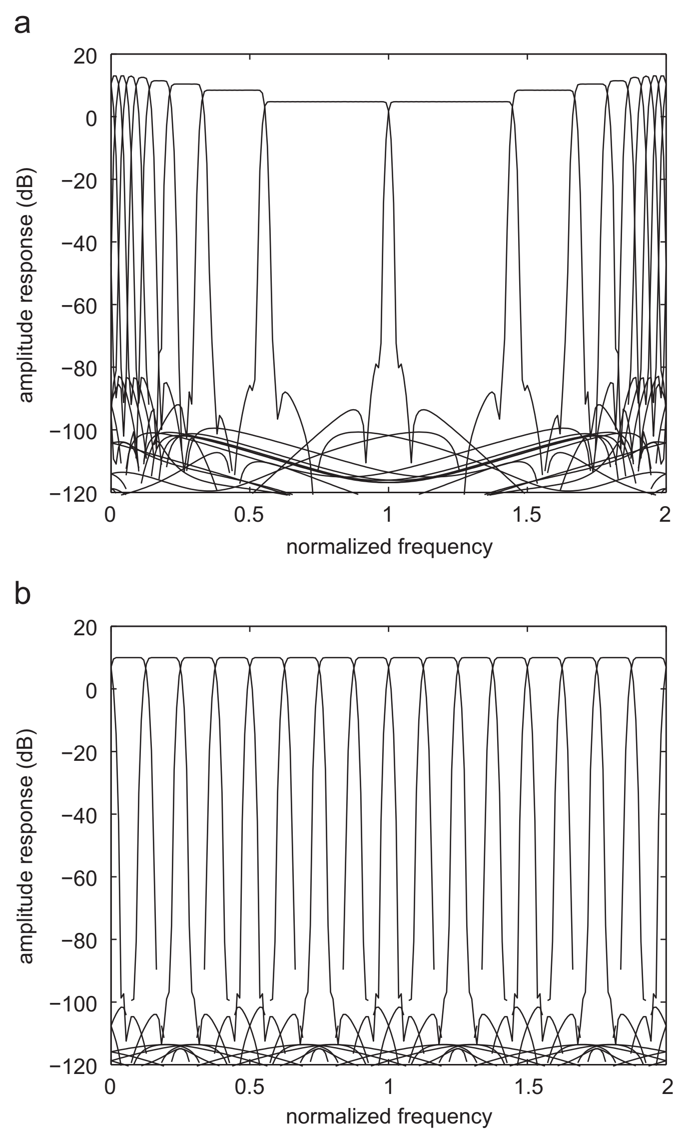



**Fig. 3.** Frequency responses of analysis filters of 16-channel oversampled filter banks: (a) a Bark-scale filter bank and (b) a uniform filter bank.

The main modification is as follows: The original Murata's method which was designed for a frequency domain approach [37] sorts frequency bins in order of weakness of correlation

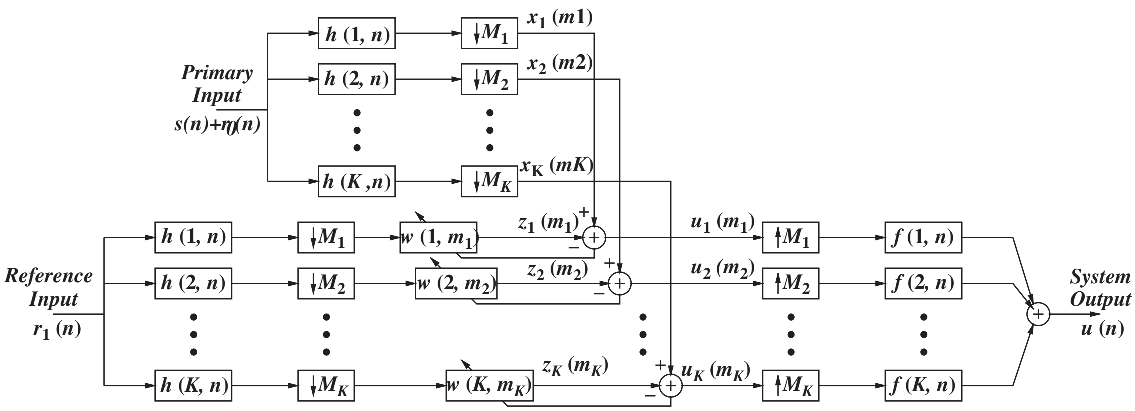



**Fig. 4.** A nonuniform oversampled filter bank approach to ANC.

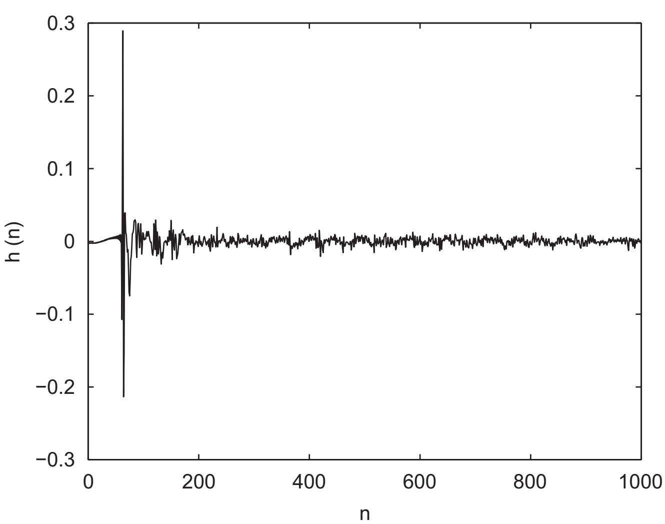



**Fig. 5.** A mixing filter measured in a normal office room.

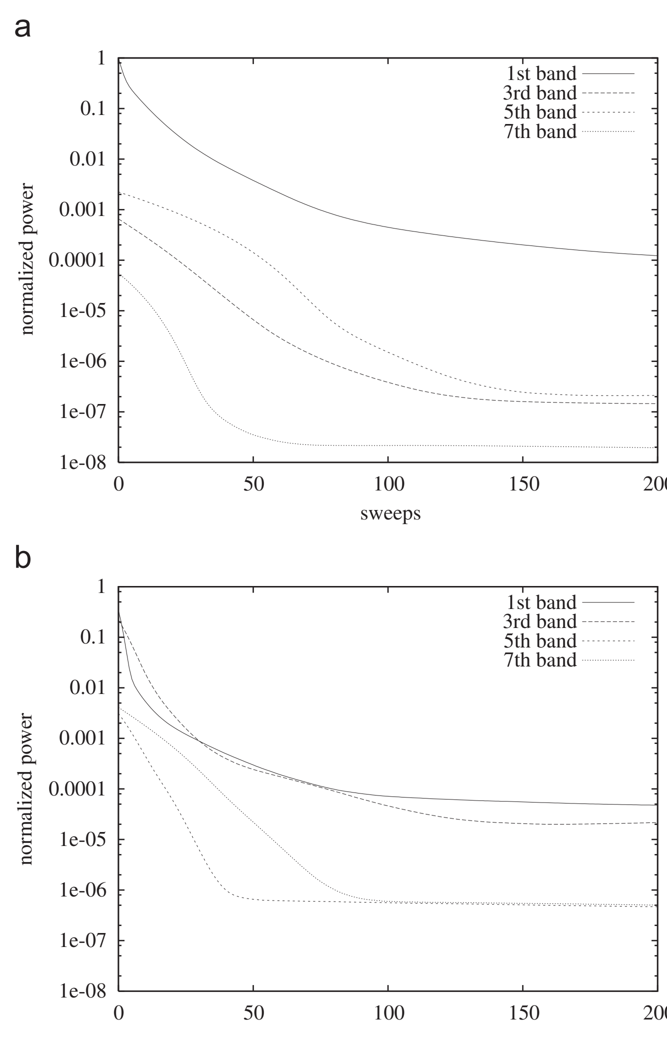



**Fig. 6.** Residual subband noise powers in the process of the filter bank approaches: (a) the uniform filter bank approach and (b) the Bark-scale filter bank approach.

among envelopes of estimated independent components. For the frequency bin which has the smallest correlation, its independent components are assigned to specific outputs. Then, for frequency bins sorted in the increasing order of the correlation, independent components are assigned to the outputs that have more correlation between envelopes of the frequency bin and averaged envelopes of the former frequency bins. However, if the next frequency bin is far from the former frequency bins, envelopes of the frequency spectra may be very different even though they are formed from the same source.

Therefore, we do not use this order to fix the permutation of subbands in the filter bank approach except a subband which has the smallest correlation. After assigning outputs for the subband which has the smallest correlation, we perform assignments for subbands adjacent to the previous subbands instead of the subband which has the next smallest correlation. In addition, when independent components are assigned to outputs in a subband, we use correlation between envelopes of the subband and weighted averages from envelopes of the previous subbands with a forgetting factor. This will emphasize the envelopes of close subbands. In this way, we may achieve more desirable results than those obtained by the original Murata's method [37], since the envelopes from close subbands will be more similar than those from distant subbands, and the indeterminacy of the filter bank approach can be successfully solved [13].

If a mixing environment is static, one may employ a method which exploits directivity patterns of sources to fix permutation [38]. Taking non-stationarity of signals or the static nature of the mixing environment into account, either permutation correction method can be used.

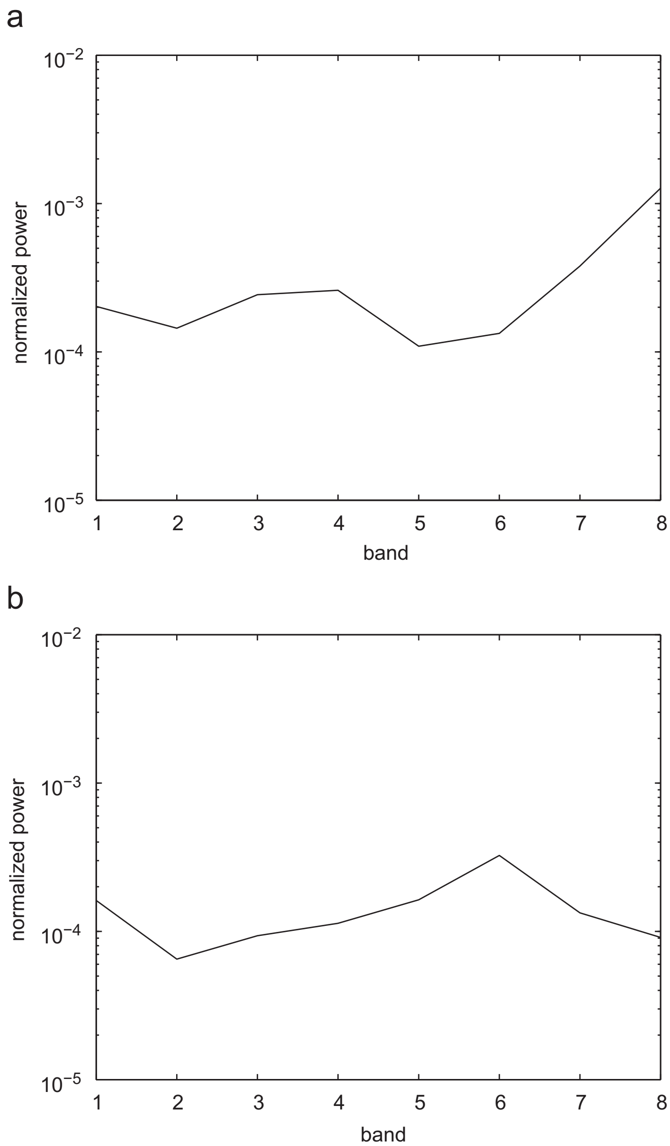



**Fig. 7.** Subband power ratios of residual noise components after 150 sweeps to those without cancellation in the filter bank approaches: (a) the uniform filter bank approach and (b) the Bark-scale filter bank approach.

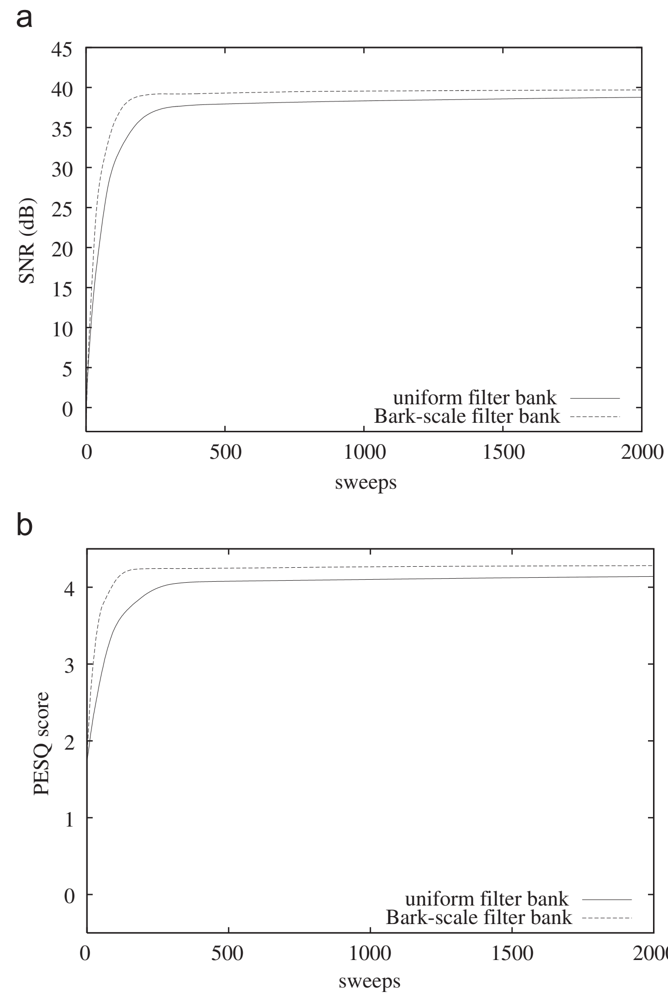



**Fig. 8.** Experimental results of the filter bank approaches to ANC for speech noise: (a) SNR and (b) PESQ score.

## 4. Experimental results

### 4.1. Design of filter banks

Fig. 3(a) shows the frequency response of analysis filters of a 16-channel Bark-scale GDFT oversampled filter bank. The filter bank was designed for subband center frequencies, [109 328 578 906 1391 2156 3516 6203]Hz, at the 16-kHz sampling rate with the corresponding decimation factors of [22 22 20 18 14 11 7 3] in the lower half subbands. To reduce computational loads, the decimation factors were the largest integers which maintained negligible aliasing. The number of filters was selected to offer reasonable decimation for whitening the inputs in low frequency subbands and fairly sufficient data for training adaptive filters. However, it was only a reasonable candidate and was not tuned to the experimental data to show robustness on it. For comparison, a uniform filter bank was also constructed as shown in Fig. 3(b). In all subbands, the common decimation factor was 10. All prototype filters to build both the filter banks had 220 taps. The oversampling ratios, the ratios of the number of total samples in all subbands to the corresponding original samples, were 167% and 160%.

### 4.2. Preliminary experiments on adaptive noise canceling (ANC)

To verify several statements which motivate the nonuniform filter bank approach, experiments on ANC were performed because they avoid the ICA indeterminacy. ANC is an approach to reduce noise based on reference signal [39]. In a typical ANC system, the primary input signal is a combination of a signal $s(n)$ and a noise $r_0(n)$ in which the signal $s(n)$ is transmitted over a channel from a signal source to a sensor and the noise $r_0(n)$ is added in the sensor from a noise source. Another sensor receives a noise signal $r_1(n)$ through another channel from the same noise source, and this signal acts as the reference. The goal is to get a system output $u(n)$ in which noise components are removed as much as possible. The ANC can be regarded as a special case of blind source separation (BSS) where some source signals can be obtained without being interfered by others. Adaptive filter coefficients should be estimated to achieve this goal, where the output $u(n)$ is

$$u(n) = s(n) + r_0(n) - \sum_{m=1}^{L_a} w(m) r_1(n-m). \tag{14}$$

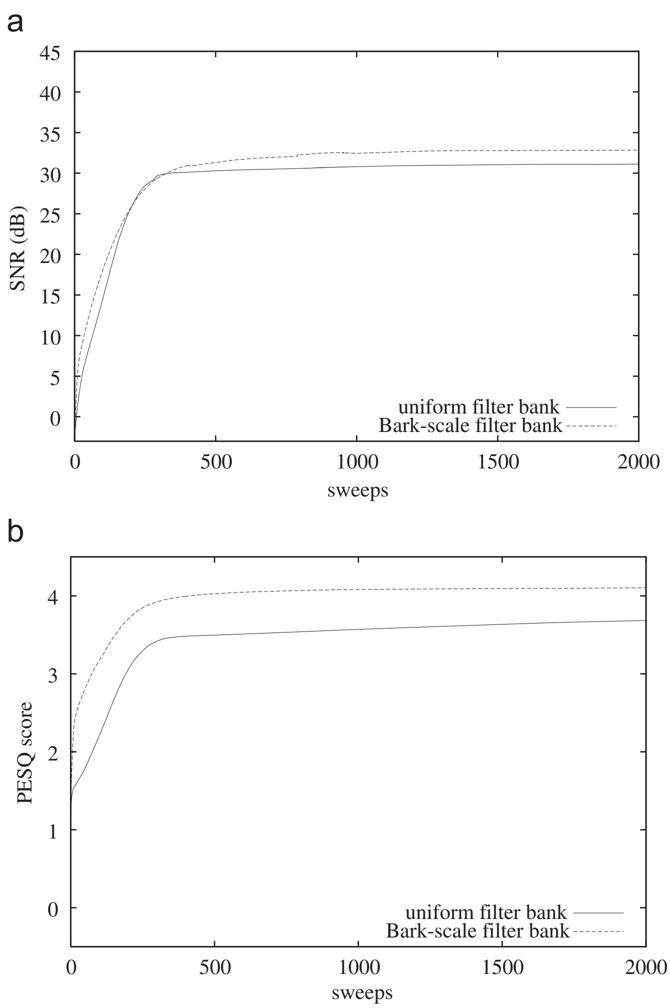



**Fig. 9.** Experimental results of the filter bank approaches to ANC for car noise: (a) SNR and (b) PESQ score.

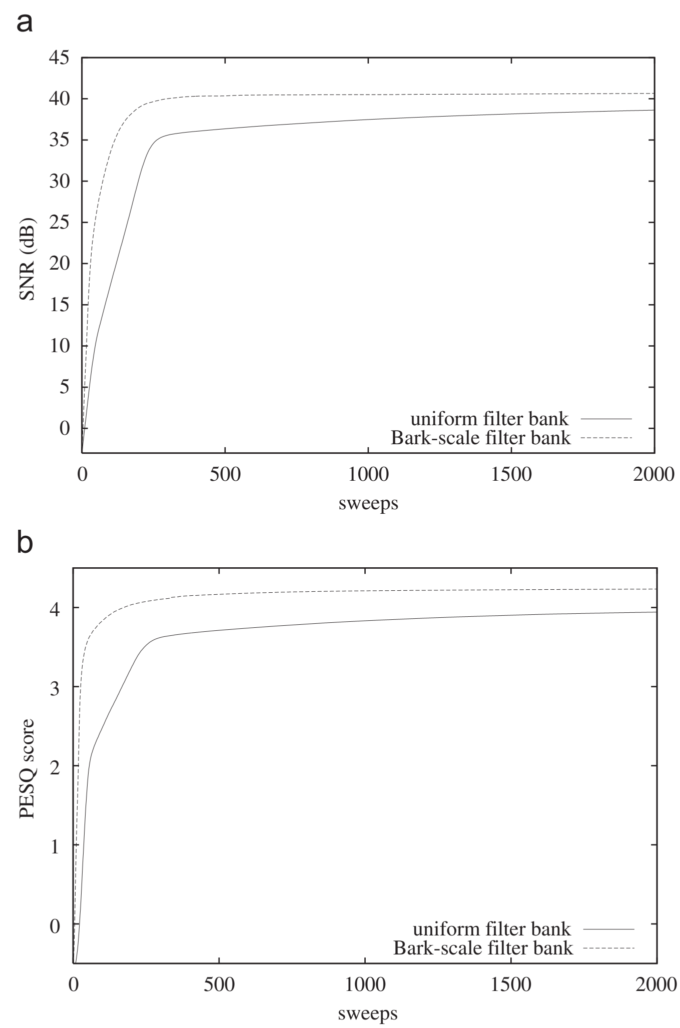



**Fig. 10.** Experimental results of the filter bank approaches to ANC for music noise: (a) SNR and (b) PESQ score.

Although the most popular algorithm for noise cancellation is the LMS algorithm, performance of ANC systems can be improved by ICA [40]. Furthermore, a filter bank approach can be applied to the ANC system as a more efficient method [12]. The overall structure for a filter bank approach to ANC is shown in Fig. 4. Using entropy maximization technique, learning rules of adaptive filter coefficients $w(k, m_k)$ can be derived as [12]

$$w_{\text{new}}(k, m_k) = w_{\text{old}}(k, m_k) + \mu(k, n_k)\varphi(u(k, n_k))r_1^*(k, n_k - m_k). \tag{15}$$

In this approach, noise components in the primary input are cancelled by the filtered reference signal in each subband, and the adaptive filters do not introduce any distortions to the desired signal components at the system output. Therefore, the filter bank approach to ANC is not affected by the ICA indeterminacy.

To perform experiments on the ANC system with filter bank approaches, two streams of speech were used as the signal and noise sources. Each signal had 10-s length at the 16-kHz sampling rate. It is known that speech signal approximately follows Laplacian distribution. Thus, $\text{sgn}(|u|)\exp(j \cdot \angle u)$ was used as the score function $\varphi(u)$, where $\text{sgn}(\cdot)$ denotes the signum function. The mixing filters from the signal source to the primary input and from the noise source to the reference input were simple linear scales. The scale values were chosen to obtain desired initial SNRs. For the mixing filter from the noise source to the primary input, we have used a measured filter in a normal office room as shown in Fig. 5.

Fig. 6 displays every other subband powers of residual noise components in the process of cancellation with the filter bank approaches using designed two filter banks. All of the powers were normalized by total powers without cancellation. Since speech has more energy in low frequency region than in high frequency region, differences among the subband powers for the uniform filter bank were much larger than those for the Bark-scale filter bank. In addition, the residual powers for the Bark-scale filter bank approach decreased at much more similar speeds in all subbands than those for the uniform filter bank approach. On the other hand, the first subband showed very slow decrease in the residual power when the uniform filter bank was used for the filter bank approach. It is consistent with the fact that the uniform filter bank approach has more colored input signals in low frequency subbands than in high frequency subbands. Unfortunately, the first subband usually dominates convergence speed of the uniform filter bank approach, since the subband has most of the power.

To estimate the extents of decrease in residual noise components with the filter bank approaches, a ratio of a subband power of residual noise components after 150 sweeps to the corresponding subband power of these components without cancellation was computed in each subband and is shown in Fig. 7. The uniform filter bank approach had difficulty in canceling many noise components in high frequency subbands, whereas the Bark-scale filter bank approach removed these noise components in rather similar ratios across subbands. It strengthens our argument that data in high frequency subbands may not be sufficient to learn adaptive filters exactly in the uniform filter bank approach. Furthermore, it may cause performance degradation which is magnified from the viewpoint of human perception.

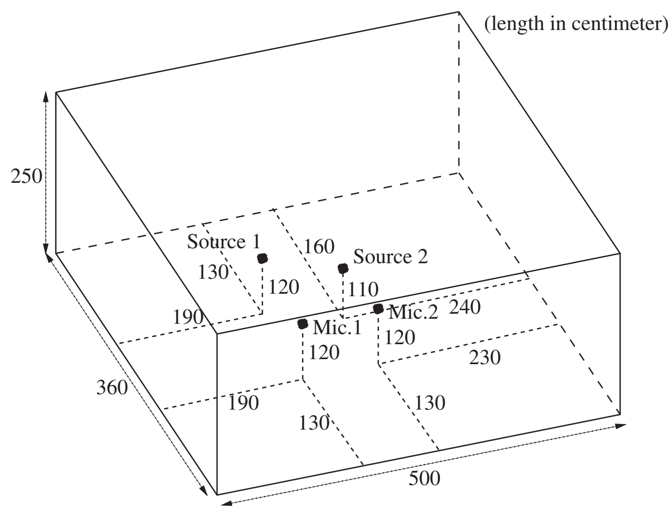



**Fig. 11.** Virtual room to simulate impulse responses from two speaker points to two microphone points.

The presented Bark-scale filter bank approach was compared with the uniform filter bank approach in terms of signal-to-noise ratio (SNR), which we define as a ratio of the signal power to the noise power at the system output $u(n)$ in the ANC system,

$$\mathrm{SNR(dB)} = 10\log\frac{\langle (s(n))^2 \rangle}{\langle (r_0(n) - \sum_{m=1}^{L_a} w(m) r_1(n-m))^2 \rangle}, \tag{16}$$

where $\langle \cdot \rangle$ denotes an operator to return a time-averaged value over a whole signal. In addition, perceptual evaluation of speech quality (PESQ) score [41,42] was used to assess speech quality of the output $u(n)$. Fig. 8 shows the experimental results. The number of taps of adaptive filter coefficients was $\lceil \frac{1024}{M_k} \rceil$ in the $k$ th subband. For both the SNR and the PESQ score, the Bark-scale filter bank approach provided faster convergence speed and higher SNRs or PESQ scores than those of the uniform filter bank approach. This is due to characteristics of natural audio signals. That is, the uniform filter bank approach has more colored input signals than the Bark-scale filter bank approach in low frequency subbands, and the former provided inferior cancellation performances to the latter in high frequency subbands. Figs. 9 and 10 also display experimental results when a car noise or a music signal was used as the noise source. These showed the same tendency.

### 4.3. Experiments on BSS

The two different filter bank approaches were compared through experiments on BSS. For independent source signals,

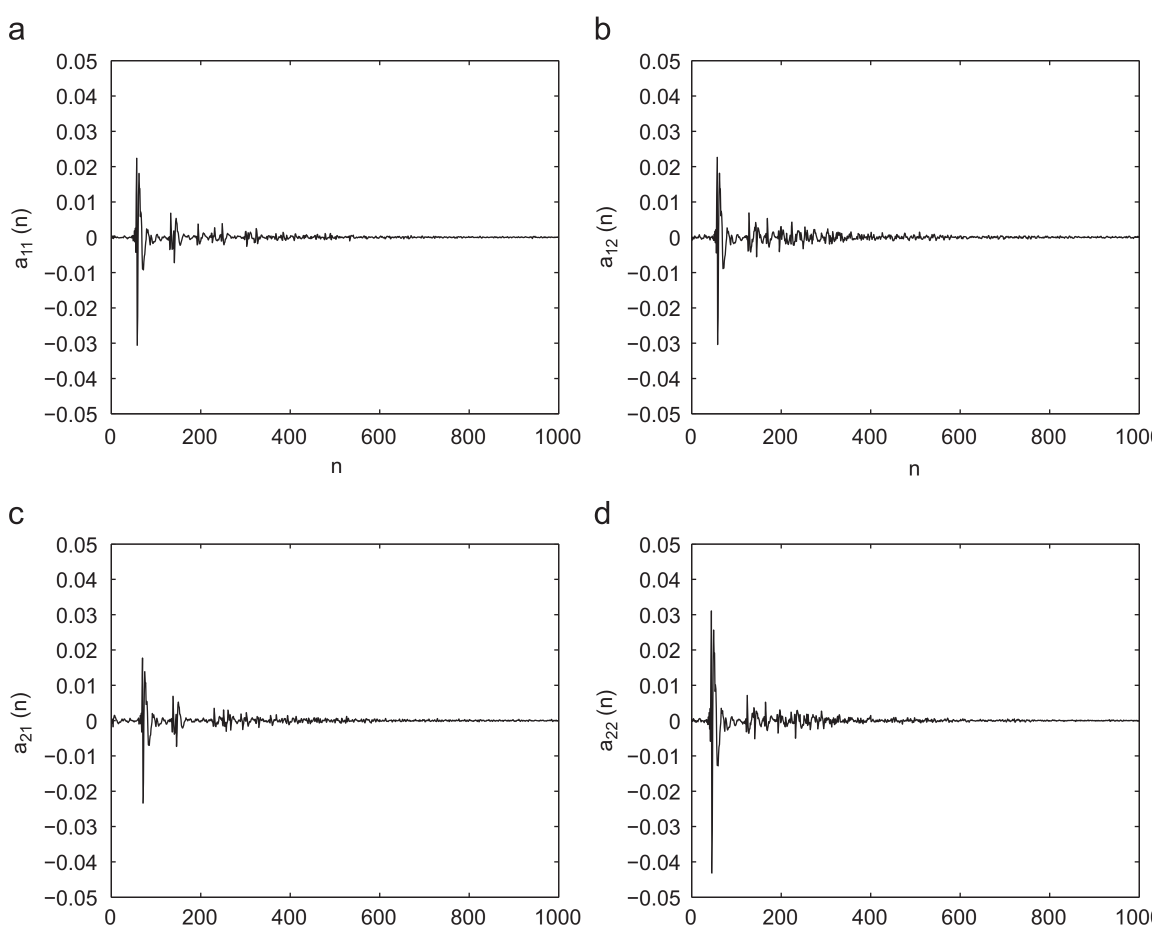



**Fig. 12.** Impulse responses of the mixing system for experiments on BSS: (a) $a_{11}$, (b) $a_{12}$, (c) $a_{21}$ and (d) $a_{22}$.

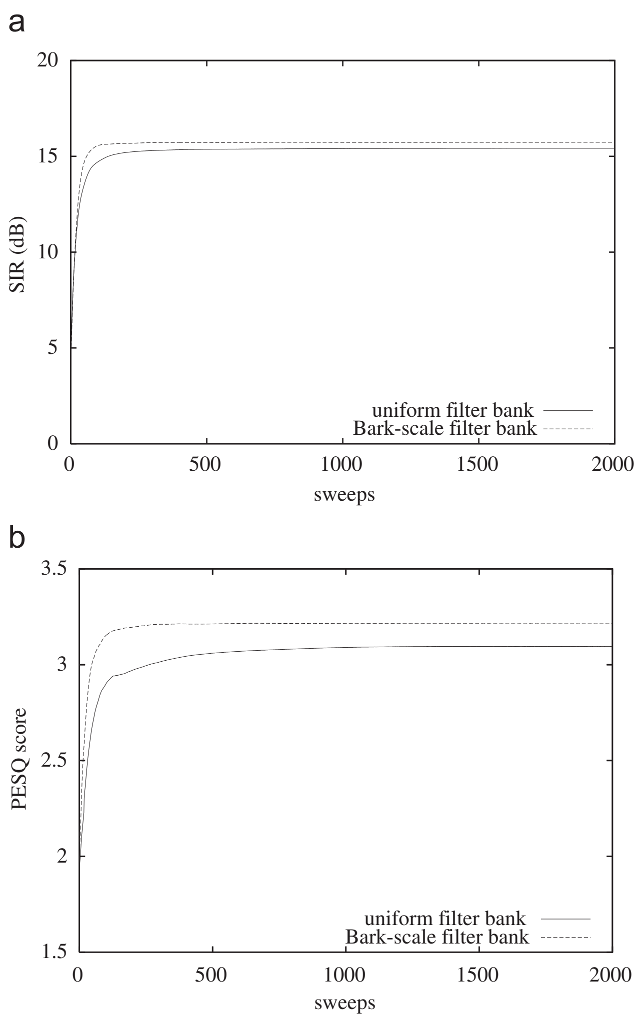



**Fig. 13.** Experimental results of the filter bank approaches to BSS for speech sources: (a) SNR and (b) PESQ score.

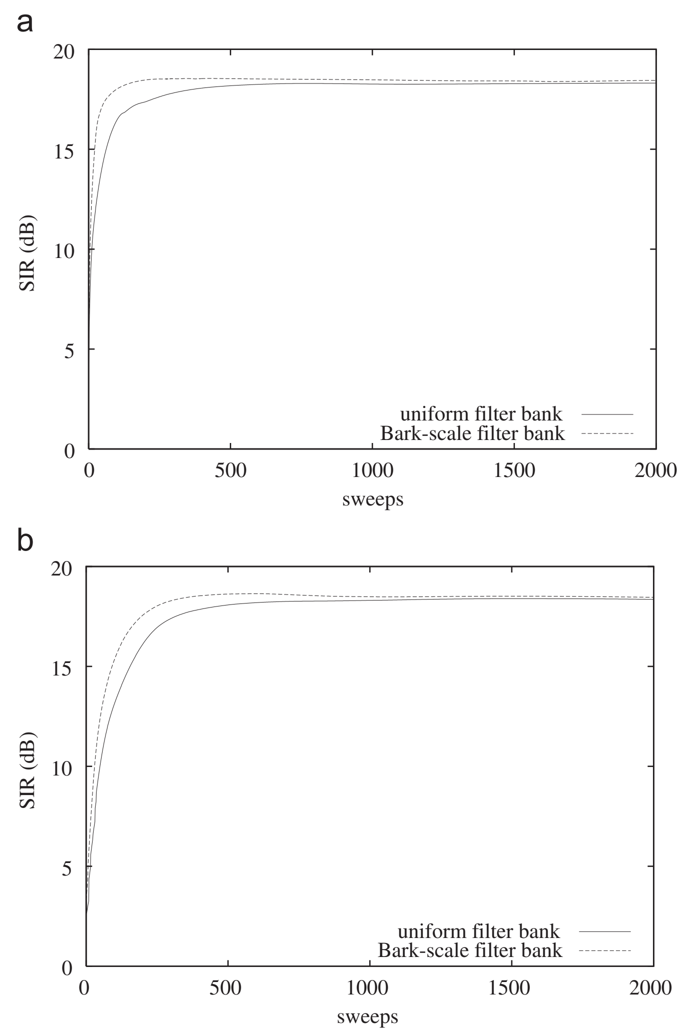



**Fig. 14.** Experimental results of the filter bank approaches to BSS for other sources: (a) speech and car noise and (b) speech and music.

we have used the same signals as in the previous experiment except that each signal had 5-s length. To construct a $2 \times 2$ mixing system, impulse responses were generated by the image method, which simulates acoustics between two points in a rectangular room [43]. Figs. 11 and 12 show a virtual room to simulate the impulse responses from two speaker points to two microphone points and the resulting impulse responses, respectively. All reflection coefficients were 0.6, and the reverberation time $RT_{60}$ was 230 ms.

Fig. 13 shows experimental results when the two streams of speech were used as sources. Using the speech sources, the score function was $\mathrm{sgn}(|u_i|)\exp(j \cdot \angle u_i)$. A feedback architecture was used for the separation network in each subband, and each filter of the feedback network in the $k$ th subband had $\left\lceil \frac{2048}{M_k} \right\rceil$ taps. Experimental results were compared in terms of signal-to-interference ratio (SIR). For a $2 \times 2$ mixing/unmixing system, the SIR is defined as a ratio of the desired signal power to the interference power at outputs [44],

$$\mathrm{SIR(dB)} = \frac{1}{2} \cdot \left| 10\log\left( \frac{\langle (u_{1,s_1}(n))^2 \rangle}{\langle (u_{1,s_2}(n))^2 \rangle} \cdot \frac{\langle (u_{2,s_2}(n))^2 \rangle}{\langle (u_{2,s_1}(n))^2 \rangle} \right) \right|. \tag{17}$$

Here, $u_{i,s_j}(n)$ denotes the $i$ th output of the cascaded mixing/unmixing system when only $s_j(n)$ is active. In addition, PESQ scores averaged from the two outputs were also displayed in Fig. 13 to assess quality of recovered speech. As in the experiments on ANC, the Bark-scale filter bank approach showed faster convergence and gave better SIRs or PESQ scores than those of the uniform filter bank approach.

Fig. 14 displays experimental results when a car noise or a music signal replaced one of the source signals. Assuming that the probability density functions of these signals can be approximated by Gaussian distribution in each subband, a linear function was used as the score function. Since a linear function may return much larger values than $\mathrm{sgn}(\cdot)$ and the values may be dominant for adaptation of an ICA network, we have used $\varphi(u) = u/5$ as the score function for signals normalized by estimated standard deviations. The same tendency was observed in these curves as in the previous results.

## 5. Concluding remarks

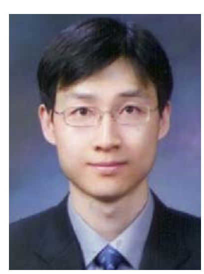
In this paper, we have described a Bark-scale filter bank approach to ICA. By measuring subband noise powers in ANC, we demonstrated that the Bark-scale filter bank approach provided much more similar convergences and cancellation performances in all subbands than the uniform filter bank approach. Especially, experiments on BSS of audio signals demonstrated that the Bark-scale filter bank approach achieved faster convergence speed and better separation performances than the uniform filter bank approach. The reason for performance improvements of the Bark-scale filter bank is due to more whitened input signals in low frequency subbands and enough data in high frequency subbands.

## Acknowledgments

This work was supported by the Brain Neuroinformatics Research Program sponsored by Korean Ministry of Science and Technology and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0070250).

## References

[1] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (1994) 287–314.
[2] S. Haykin (Ed.), Unsupervised Adaptive Filtering, Wiley, New York, NY, 2000.
[3] T.-W. Lee, Independent Component Analysis, Kluwer Academic Publishers, Boston, MA, 1998.
[4] S. Amari, A. Cichocki, H.H. Yang, A new learning algorithm for blind source separation, Advances in Neural Information Processing Systems, vol. 8, MIT Press, Cambridge, MA, 1996, pp. 757–763.
[5] A.J. Bell, T.J. Sejnowski, Edges are the independent components of natural scenes, Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, MA, 1997, pp. 831–837.
[6] J.-H. Lee, T.-W. Lee, H.-Y. Jung, S.-Y. Lee, On the efficient speech feature extraction based on independent component analysis, Neural Processing Letters 15 (3) (2002) 235–245.
[7] T.-W. Lee, A.J. Bell, R.H. Lambert, Blind separation of delayed and convolved sources, Advances in Neural Information Processing Systems, vol. 9, MIT Press, Cambridge, MA, 1997, pp. 758–764.
[8] K. Torkkola, Blind separation of convolved sources based on information maximization, in: Proceedings of the IEEE International Workshop on NNSP, 1996, pp. 423–432.
[9] H.-M. Park, H.-Y. Jung, T.-W. Lee, S.-Y. Lee, Subband-based blind signal separation for noisy speech recognition, Electronics Letters 35 (23) (1999) 2011–2012.
[10] P. Smaragdis, Information theoretic approaches to source separation, Master's Thesis, MIT Media Arts and Science Department, MIT, Boston, June 1997.
[11] P. Smaragdis, Blind separation of convolved mixtures in the frequency domain, Neurocomputing 22 (1998) 21–34.
[12] H.-M. Park, S.-H. Oh, S.-Y. Lee, A uniform oversampled filter bank approach to independent component analysis, Proceedings of IEEE ICASSP 5 (2003) 249–252.
[13] H.-M. Park, C.S. Dhir, S.-H. Oh, S.-Y. Lee, A filter bank approach to independent component analysis for convolved mixtures, Neurocomputing 69 (2006) 2065–2077.
[14] S. Araki, S. Makino, R. Aichner, T. Nishikawa, H. Saruwatari, Subband based blind source separation with appropriate processing for each frequency band, in: Proceedings of the International Conference on ICA and BSS, 2003, pp. 499–504.
[15] J. Huang, K.-C. Yen, Y. Zhao, Subband-based adaptive decorrelation filtering for co-channel speech separation, IEEE Transactions on Speech and Audio Processing 8 (4) (2000) 402–406.
[16] S. Weiß, S.R. Dooley, R.W. Stewart, A.K. Nandi, Adaptive equalization in oversampled subbands, Electronics Letters 34 (15) (1998) 1452–1453.
[17] S. Weiß, A. Stenger, R.W. Stewart, R. Rabenstein, Steady-state performance limitations of subband adaptive filters, IEEE Transactions on Signal Processing 49 (9) (2001) 1982–1991.
[18] J.D. Griesbach, T. Bose, D.M. Etter, Non-uniform filterbank bandwidth allocation for system modeling subband adaptive filters, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999, pp. 1473–1476.
[19] J.D. Griesbach, M. Lightner, D.M. Etter, Spectral domain MMSE estimators for non-uniform system modeling subband adaptive filters, in: Asilomar Conference on Signals, Systems and Computers, 2000, pp. 1489–1493.
[20] T. Rutkowski, A. Cichocki, A. K. Barros, Speech enhancement using adaptive filters and independent component analysis approach, in: International Conference on Artificial Intelligence in Science and Technology, AISAT'2000, 2000, pp. 191–196.
[21] S. Schulz, T. Herfet, Binaural source separation in non-ideal reverberant environments, in: International Conference on Digital Audio Effects, DAFx-07, 2007, pp. 133–140.
[22] L. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[23] Y. Zhao, R. Hu, X. Li, Speedup convergence and reduce noise for enhanced speech separation and recognition, IEEE Transactions on Audio, Speech, and Language Processing 14 (4) (2006) 1235–1244.
[24] A. Gilloire, M. Vetterli, Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation, IEEE Transactions on Signal Processing 40 (8) (1992) 1862–1875.
[25] Y. Yamada, H. Ochi, H. Kiya, A subband adaptive filter allowing maximally decimation, IEEE Journal on Selected Areas in Communications 12 (9) (1994) 1548–1552.
[26] K. Torkkola, Blind separation for audio signals—are we there yet?, in: Proceedings of the International Conference on ICA and BSS, 1999, pp. 239–244.
[27] H. Sawada, R. Mukai, S. Araki, S. Makino, A polar-coordinate based activation function for frequency domain blind source separation, in: Proceedings of the International Conference on ICA and BSS, 2001, pp. 663–668.
[28] S. Haykin, Adaptive Filter Theory, fourth ed., Prentice-Hall, Upper Saddle River, NJ, 2002.
[29] M.H. Hayes, Statistical Digital Signal Processing and Modeling, Wiley, New York, NY, 1996.
[30] M.-H. Lu, P.M. Clarkson, The performance of adaptive noise cancellation systems in reverberant rooms, Journal of the Acoustical Society of America 93 (2) (1993) 1122–1135.
[31] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, NY, 1998.
[32] M. Slaney, An efficient implementation of the Patterson–Holdsworth auditory filterbank, Technical Report 35, Apple Computer, 1993.
[33] H. Hermansky, Perceptual linear predictive (PLP) analysis of speech, Journal of the Acoustical Society of America 87 (4) (1990) 1738–1752.
[34] M. Harteneck, S. Weiß, R.W. Stewart, Design of near perfect reconstruction oversampled filter banks for subband adaptive filters, IEEE Transactions on Circuits Systems II 46 (1999) 1081–1085.
[35] S. Weiß, L. Lampe, R. W. Stewart, On implementation and design of filter banks for subband adaptive systems, in: Proceedings of the IEEE Workshop on Signal Processing Systems, 1998, pp. 172–181.
[36] S. Weiß, On adaptive filtering on oversampled subbands, Ph.D. Thesis, Signal Processing Division, University of Strathclyde, Glasgow, May 1998.
[37] N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, Neurocomputing 41 (2001) 1–24.
[38] C. S. Dhir, H.-M. Park, S.-Y. Lee, Permutation correction of filter bank ICA using static channel characteristics, in: International Conference on Neural Information Processing, ICONIP'2004, 2004, pp. 1076–1081.
[39] B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong Jr., R.C. Goodlin, Adaptive noise cancelling: principles and applications, Proceedings of the IEEE 63 (12) (1975) 1692–1716.
[40] H.-M. Park, S.-H. Oh, S.-Y. Lee, Adaptive noise cancelling based on independent component analysis, Electronics Letters 38 (15) (2002) 832–833.
[41] A.W. Rix, J.G. Beerends, M.P. Hollier, A.P. Hekstra, Perceptual evaluation of speech quality (PESQ)—the new ITU standard for end-to-end speech quality assessment. Part I—time-delay compensation, Journal of the Audio Engineering Society 50 (10) (2002) 755–764.
[42] J.G. Beerends, A.P. Hekstra, A.W. Rix, M.P. Hollier, Perceptual evaluation of speech quality (PESQ)—the new ITU standard for end-to-end speech quality assessment. Part II—psychoacoustic model, Journal of the Audio Engineering Society 50 (10) (2002) 765–778.
[43] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, Journal of the Acoustical Society of America 65 (4) (1979) 943–950.
[44] D. Schobben, K. Torkkola, P. Smaragdis, Evaluation of blind signal separation methods, in: Proceedings of the International Conference on ICA and BSS, 1999, pp. 261–266.



**Hyung-Min Park** received the B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1997, 1999, and 2003, respectively. From 2003 to early 2005, he was a Post-Doc. at the Department of Biosystems, Korea Advanced Institute of Science and Technology. From 2005 to 2006, he was with the Language Technologies Institute, Carnegie Mellon University. Since early 2007, he has been an Assistant Professor at the Department of Electronic Engineering, Sogang University, Seoul, Korea. His current research interests include the theory and applications of independent component analysis, blind source separation, adaptive noise canceling, and binaural or multi-microphone processing for noise-robust speech recognition.

**Sang-Hoon Oh** received his B.S. and M.S. degrees in Electronics Engineering from Pusan National University in 1986 and 1988, respectively. He received his Ph.D. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology in 1999. From 1988 to 1989, he worked for the LG semiconductor, Ltd., Korea. From 1990 to 1998, he was a senior researcher in Electronics and Telecommunications Research Institute (ETRI), Korea. From 1999 to 2000, he was with Brain Science Research Center, KAIST. In 2000, he was with Brain Science Institute, RIKEN in Japan as a research scientist. In 2001, he was an R&D manager of Extell Technology

Corporation, Korea. Since 2002, he has been with the Department of Information Communication Engineering, Mokwon University, Daejeon, Korea, and is now an Associate Professor. Also, he was with the Division of Computational Science and Engineering, College of Computing, Georgia Institute of Technology, Atlanta, USA, as a visiting scholar from August 2008 to August 2009. His research interests are supervised/unsupervised learning for intelligent information processing, speech processing, pattern recognition, and bioinformatics.

**Soo-Young Lee** received B.S., M.S., and Ph.D. degrees from Seoul National University in 1975, Korea Advanced Institute of Science in 1977, and Polytechnic Institute of New York in 1984, respectively. From 1977 to 1980 he worked for the Taihan Engineering Co., Seoul, Korea. From 1982 to 1985 he also worked for General Physics Corporation at Columbia, MD, USA. In early 1986 he joined the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, as an Assistant Professor and now is a Full Professor at the Department of BioSystems and also Department of Electrical Engineering and Computer Science. In 1997 he established *Brain Science Research Center*, which is the main research organization for the *Korean Brain Neuroinformatics Research Program*. The research program is one of the Korean Brain Research Promotion Initiatives sponsored by Korean Ministry of Science and Technology from 1998 to 2008, and currently about 35 Ph.D. researchers have joined the research program from many Korean universities. He is a Past-President of *Asia-Pacific Neural Network Assembly*, and has contributed to *International Conference on Neural Information Processing* as Conference Chair (2000), Conference Vice Co-Chair (2003), and Program Co-Chair (1994, 2002). Dr. Lee is the Editor-in-Chief of the newly-established online/offline journal with double-blind review process, *Neural Information processing—Letters and Reviews*, and is on Editorial Board for two international journals, i.e., Neural Processing Letters and Neurocomputing. He received Leadership Award and Presidential Award from *International Neural Network Society* in 1994 and 2001, respectively, and APPNA Service Award in 2004. His research interests have resided in *artificial brain, the human-like intelligent systems based on biological information processing mechanism in our brain*. He has worked on the auditory models from the cochlea to the auditory cortex for noisy speech processing, information-theoretic binaural processing models for sound localization and speech enhancement, the unsupervised pro-active developmental models of human knowledge with multi-modal man-machine interactions, and the top-down selective attention models for superimposed pattern recognitions. His research scope covers the mathematical models, neuromorphic chips, and real-world applications. Also, he had recently extended his research into brain–computer interfaces with simultaneous fMRI and EEG measurements.