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

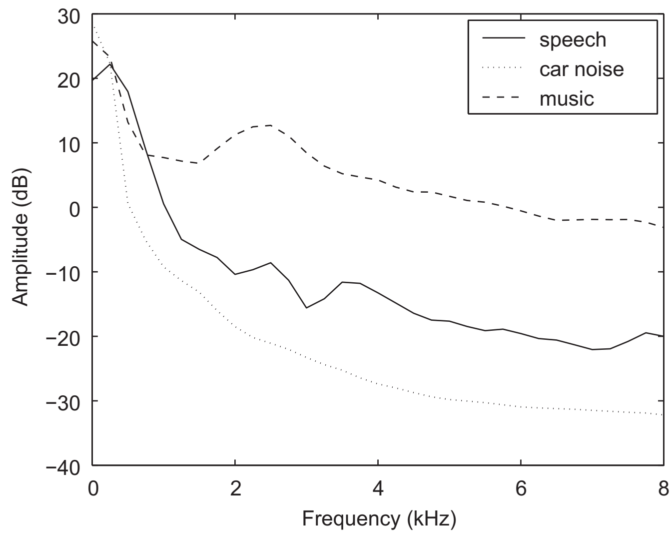

Fig. 1. Time-averaged power spectral densities for three natural sounds in the frequency domain.

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

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

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]:

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