Blind Signal Separation with Speech Enhancement

Chang-Hong Lin, Hsiao-Ping Lee, Jyun-Hong Li, Chih-Wei Su, Yu-Hao Chin, Jhing-Fa Wang and Jia-Ching Wang

Abstract A new speech enhancement architecture using convolutive blind signal separation (CBSS) and subspace-based speech enhancement is presented. The spatial and spectral information are integrated to enhance the target speech signal and suppress both interference noise and background noise. Real-world experiments were carried out in a noisy office room. Experimental results demonstrate the superiority of the proposed architecture.

Introduction

Many multiple- microphone speech enhancement methods have been proposed to exploit spatial information to extract the single source signal [1–4]. To significantly reduce the number of microphones and do not require a priori information about the sources, blind signal separation (BSS) methods are adopted to effectively separate interfering noise signals from the desired source signal. The second-order decorrelation based convolutive blind signal separation (CBSS) algorithm was recently developed [5]. Since estimating second-order statistics is numerically robust and the criteria leads to simple algorithms [6].

We proposed a novel architecture, in which critical-band filterbank is utilized as a preprocessor to provide improved performance and further savings on convergence time and computational cost. However, only critical-band CBSS does not work for removing background noise, which originates from a complex combination of a large number of spatially distributed sources. Therefore, a subspace-

C.-H. Lin · J.-H. Li · C.-W. Su · Y.-H. Chin · J.-C. Wang (✉)
Department of Computer Science and Information Engineering,
National Central University, Jhongli, Taiwan, Republic of China
e-mail: jcw@csie.ncu.edu.tw

H.-P. Lee · J.-F. Wang
Department of Electrical Engineering, National Cheng-Kung University,
Tainan, Taiwan, Republic of China

Y.-M. Huang et al. (eds.), *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*, Lecture Notes in Electrical Engineering 260,
DOI: 10.1007/978-94-007-7262-5_119, © Springer Science+Business Media Dordrecht 2014

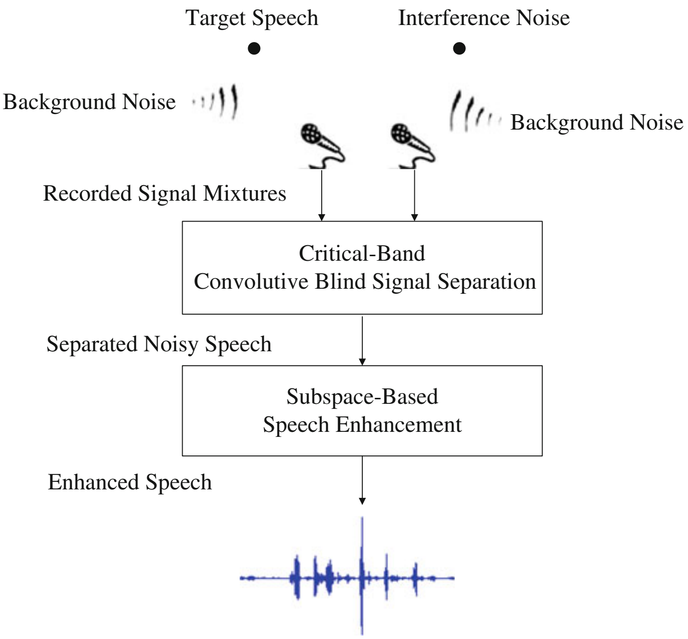


Fig. 1 Block diagram of proposed speech enhancement system

based speech enhancement method is utilized to reduce the background noise by exploiting additional spectral information [7].

Proposed Architecture

Figure 1 schematically depicts the block diagram of the proposed speech enhancement system. This architecture comprises a critical-band CBSS module and a subspace-based speech enhancement module. The input mixed signals are first processed by using the critical-band CBSS to separate the target speech from the interference noise. Next, the extracted target speech is fed into the subspace-based speech enhancement module to reduce the residual interferences and background noise. The proposed architecture adopts both spatial and spectral processing, and needs only two microphones.

Critical-Band Convolutive Blind Signal Separation

First, a critical-band filterbank based on the perceptual wavelet transform (PWT) is built from the psycho-acoustic model. The recorded signal mixtures are decimated into critical band time series by PWT. The CBSS is performed to separate the noisy speech and the interference noise in each critical band. A signal selection strategy based on high order statistics is then adopted to extract the target speeches. Finally, the inverse perceptual wavelet transform (IPWT) is applied to the critical-band extracted speeches to reconstruct the full-band separated noisy speech.

Perceptual auditory modeling is very popular for speech analysis and recognition. The wavelet packet decomposition is designed to adjust the partitioning of the frequency axis into critical bands which are widely used in perceptual auditory modeling. Within the 4 kHz bandwidth, this work uses 5-level wavelet tree structure to approximate the 17 critical bands derived based on the measurement [8, 9].

Convolutive Blind Signal Separation

This work assumes that two mixture signals $\bar{x}(t) = [x_1(t), x_2(t)]^{\mathrm{T}}$ composed of two point source signals $\bar{s}(t) = [s_1(t), s_2(t)]^{\mathrm{T}}$ and additive background noise $\bar{n}(t)$ are recorded at two different microphone locations:

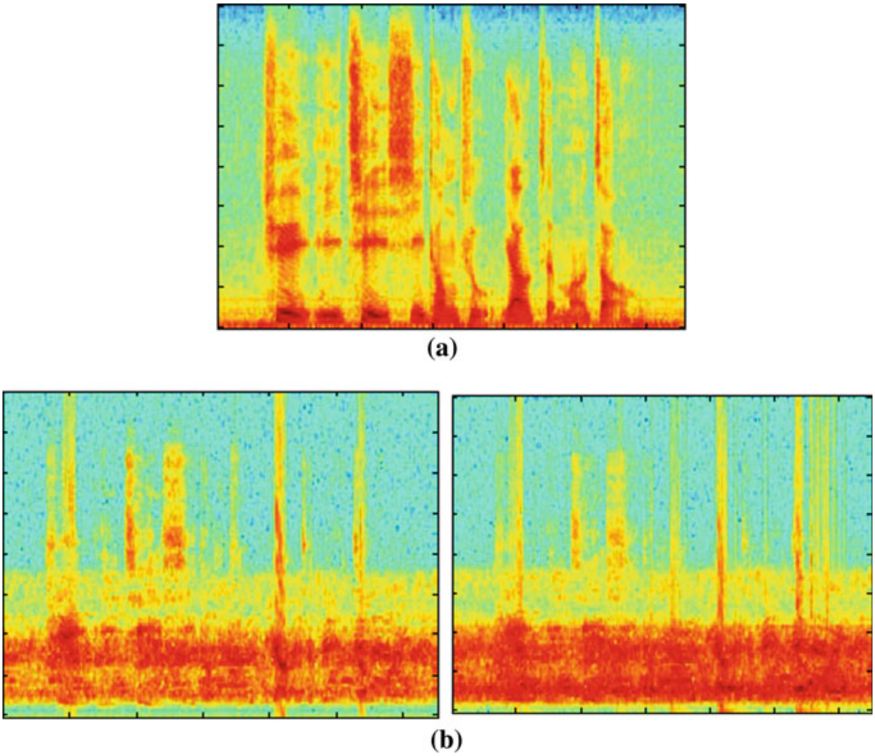
$$\bar{x}(t) = \sum_{\tau=0}^{P} \mathbf{A}(\tau)\bar{s}(t-\tau) + \bar{n}(\tau). \tag{1}$$

The mixing matrix $\mathbf{A}$ is a 2×2 matrix and P represents the convolution order. Passing through the critical-band filterbank, PWT separates mixture signals into 17 critical-band wavelet packet coefficients. In each critical band, using an M-point windowed discrete Fourier transformation (DFT), the time-domain equation (1) can be converted into frequency-domain. The convolutive BSS is then performed in each critical band.

Signal Selection

In each critical-band, the CBSS has separated the mixed signals as speech dominant and interference dominant signals. Next, we should identify the target speech from the two separated outputs. Nongaussianity can be considered as a measure for discriminating the target speech and the interference noise by using kurtosis [3].

The last stage simply synthesize the enhanced speech using the inverse perceptual wavelet transform (IPWT).


(a)

(b)

Fig. 2 **a** The original clean speech signals; **b** the 2-channel corrupted speeches under babble noise

Subspace-Based Speech Enhancement

The subspace-based speech enhancement is used to enhance the separated noisy speech by minimizing the background noise. The additive noise removal problem can be described as a clean signal $\bar{s}$ being corrupted by additive noise $\bar{n}$. The resulting noisy signal $\bar{u}$ can be expressed as

$$\bar{u} = \bar{s} + \bar{n}, \tag{2}$$

where $\bar{s}$= $[s(1), s(2), \ldots, s(L)]^{\mathrm{T}}$, $\bar{n} = [n(1), n(2), \ldots, n(L)]^{\mathrm{T}}$, and $\bar{u} = [u(1), u(2), \ldots, u(L)]^{\mathrm{T}}$. The observation period has been denoted as L. Henceforth, the vectors $\bar{s}$, $\bar{n}$, and $\bar{u}$ will be considered as part of real space R^L.

Ephraim and Van Trees proposed a subspace-based speech enhancement method [7]. The goal of this method is to find an optimal estimator that would minimize the speech distortion by adopting the constraint that the residual noise fell below a preset threshold.

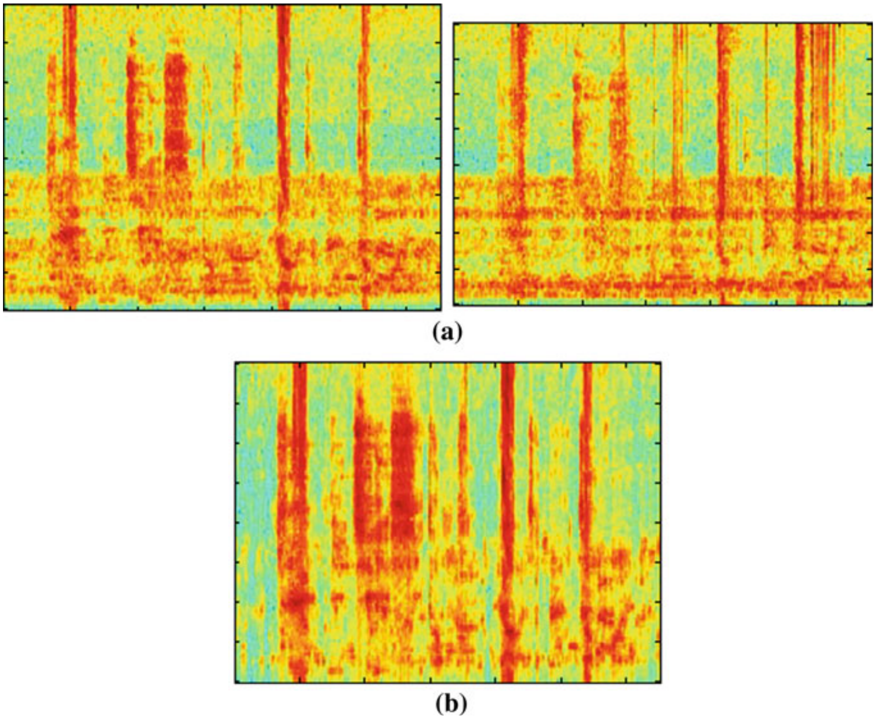


Fig. 3 **a** The two-channel critical-band CBSS outputs; **b** the selected enhanced speech

Experiment Results

The experiment was performed with a speech source and a babble interference noise at an angle of 150° and a distance of 40 cm from the center of the microphone array. Twenty different spoken sentences were played, each with about 50,000 samples and babble noise in AURORA database was employed as interference noise.

For objective evaluation, the SNR measure was adopted to evaluate these speech enhancement algorithms. Additionally, the modified Bark spectral distortion (MBSD) was also applied to assess speech quality. Since MBSD measure, presented by Yang et al. [10], is a perceptually motivated objective measure for mimicking human performance in speech quality rating. In both measures, the proposed architecture significantly outperforms conventional subspace enhancement method.

Figure 2 shows the spectrograms of original clean speech and two speeches corrupted by babble noise. Figure 3 illustrates the spectrograms of the critical-based CBSS outputs and the enhanced result. Figure 3a clearly reveals that one output is target speech dominant, while the other is interference dominant.

Conclusion

This work develops a spatio-spectral architecture for speech enhancement. The architecture consists of a critical-band CBSS module and a subspace-based speech enhancement module. The spatial and spectral information are exploited to enhance the target speech, and to suppress strong interference noise and background noise using two microphones. Kurtosis analysis is then adopted to select the target CBSS output. The enhancement performance is improved significantly.

References

1. VanVeen BD, Buckley KM (1988) Beamforming: a versatile approach to spatial filtering. IEEE Acoust, Speech Sig Process Mag 5:4–24
2. Kellermann W (1991) A self-steering digital microphone array. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing, vol 5, April pp 3581–3584
3. Low SY, Nordholm S, Togneri R (2004) Convolutive blind signal separation with post-processing. IEEE Trans Speech Audio Process 12(5):539–548
4. Visser E, Otsuka M, Lee TW (2003) A spatio-temporal speech enhancement scheme for robust speech recognition in noisy environments. Speech Commun 41(2):393–407
5. Parra L, Spence C (2000) Convolutive blind source separation of nonstationary sources. IEEE Trans Speech Audio Process 8(3):320–327
6. Parra L, Fancourt C (2002) An adaptive beamforming perspective on convolutive blind source separation. Davis G (ed) In: noise reduction in speech applications, CRC Press LLC
7. Ephraim Y, Van Trees HL (1995) A signal subspace approach for speech enhancement. IEEE Trans Speech Audio Process 3(4):251–266
8. Zwicker E, Terhardt E (1980) Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. J Acoust Soc Am 68:1523–1525
9. Chen SH, Wang JF (2004) Speech enhancement using perceptual wavelet packet decomposition and Teager energy operator. J VLSI Sig Process Syst 36(2–3):125–139
10. Yang W, Benbouchta M, Yantorno R (1998) Performance of the modified bark spectral distortion as an objective speech quality measure. In: Proceedings of IEEE international conference on acoust, speech, signal process pp 541–544