BLIND SPEECH SEPARATION ALGORITHM FOR DYNAMICALLY MIXING SYSTEMS

Chi-Tat Leung and Wan-Chi Siu
Center for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hung Hom, KLN., Hong Kong

ABSTRACT

This paper presents a blind speech separation algorithm that is capable of extracting a speech signal from background noise or music based on a microphone-array. A variable rearrangement is derived to convert convolution operations into a simple matrix multiplication in dynamically mixing systems. The fast fixed-point algorithm is then extended to separate a speech signal from background noise in a realistic room with acoustic reverberation.

INTRODUCTION

Separating speech from background noise or music is an extremely difficult issue but has many potential applications such as mobile phones and speech recognition [1]. Blind speech separation (BSS) addresses this difficult issue and becomes one of the emerging research topics of speech processing. In this paper, a BSS algorithm is developed to process the signals picked up from a microphone-array. The algorithm is able to directly estimate the speech signals from background noise or music.

THE ALGORITHM

An extension of Hyvärinen's fixed-point algorithm [2] is derived for extracting a speech signal from background noise or music based on a dynamically mixing mechanism. A thorough experimental comparison in ref. [3] concluded that the fixed-point algorithm provided the best trade-off in terms of accuracy and computational requirements for blind source separation. However, the fixed-point algorithm cannot tackle the realistic dynamically mixing mechanism or the convolutive mixing models that characterizes acoustic reverberation effects in real-room environments. In this paper, a generalization of Hyvärinen's algorithm is developed to handle the dynamically mixing situation.

As the original fixed-point algorithm was designed for the instantaneous mixing model, this paper presents a rearrangement of variables such that the convolution operations are transformed into a simple matrix multiplication for dynamically mixing systems. The fixed-point algorithm can then be applied to extract one of the speech signals from their convolutive mixtures.

Let us assume the underlying mixing mechanism to be a linear time-invariant multiple-input-multiple-output (MIMO) system. The p non-observable source signals, $\mathbf{s}(k)$, pass through the mixing system or channel with an unknown transfer function, $\mathbf{A}(z)$. At the outputs of the mixing channel, only q mixed signals, $\mathbf{x}(k)$, are accessible. The MIMO system is widely modeled by a convolutive mixing model, which is given by

$$\begin{aligned}\mathbf{x}(k) &= \mathbf{A}(z)\mathbf{s}(k)+\mathbf{n}(k)\\ &= \sum_{p=0}^{K}\mathbf{A}(p)\mathbf{s}(k-p)+\mathbf{n}(k),\end{aligned} \tag{1}$$

where $\mathbf{n}(k)$ is the additive noise perturbation at time step k. In order to extract the speech signal, another demixing FIR polynomial matrix, $\mathbf{c}(z)$, is applied to the mixtures, $\mathbf{x}(k)$, viz.,

$$\mathbf{y}(k)=\mathbf{c}(z)\mathbf{x}(k)=\sum_{p=0}^{L}\mathbf{c}(p)\mathbf{x}(k-p). \tag{2}$$

Since the original fixed-point algorithm is designed for the instantaneous mixing model, this paper presents a rearrangement of variables such that the convolution operation in (2) is transformed into a simple matrix multiplication. Let us redefine a new observation vector as $\mathbf{X}(k)=(x_1(k), \ldots, x_1(k-L), x_2(k), \ldots, x_2(k-L), \ldots, x_q(k), \ldots, x_q(k-L))^T$, and a demixing weight vector as $\mathbf{c}=(c_1(0), \ldots, c_1(L), c_2(0), \ldots, c_2(L), \ldots, c_q(0), \ldots, c_q(L))^T$. Hence, the convolution in (2) is then converted into a simple matrix multiplication as shown below:

$$y(k)=\mathbf{c}^T\mathbf{X}(k). \tag{3}$$

The original fixed-point algorithm can then be applied to extract a speech signal from its convolutive mixtures.

0-7803-6622-0/01 $10.00 © 2001 IEEE

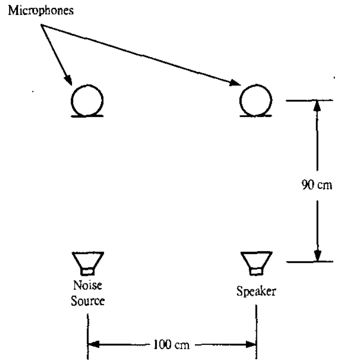

Fig. 1 Experimental Setup

The proposed method has been validated by real world recordings. The experimental setup is shown in Fig. 1. Two microphones were used to pick up the audio signals. An engine noise and a speaker formed the two audio sources. The experiments were carried out in a realistic office environment of 21 square meters. A female speaker and a male speaker were employed in this experiment. Fig. 2 and Fig. 3 illustrate the recordings of the two speakers' utterances and the extracted speech signal. The experimental simulation results show that the extended fixed-point algorithm can separate the speeches from the engine noise based on the recordings. They also indicate that the noise components in the extracted speeches were significantly suppressed. These simulation results indicate that the extended fixed-point algorithm is capable of extracting the speech signals from their mixtures in a real-world office environment.

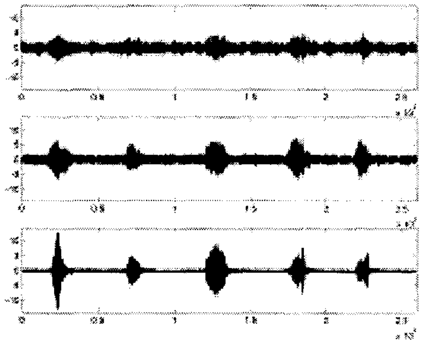

Fig. 2 Experimental Recorded Signals (The Upper Two Signals) and the extracted Female Speech (The Lowest One Signal)

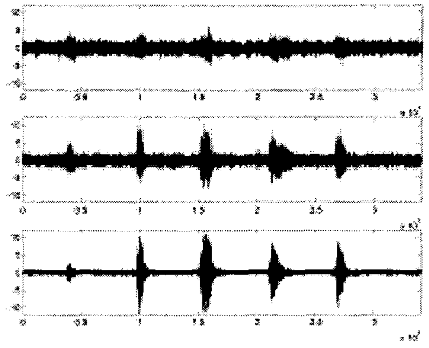

Fig. 3 Experimental Recorded Signals (The Upper Two Signals) and the extracted Male Speech (The Lowest One Signal)

CONCLUSION

In this paper, an extension of the fixed-point algorithm for dynamically mixing was developed. The proposed algorithm is capable of extracting a speech signal from the convolutive mixtures with background noise in a multi-channel dynamically mixing environment. Real-world recordings were used to verify the extended fixed-point algorithm. In real-world recordings using two channels, the algorithm is capable of significantly reducing the noise components of the speech signals.

Acknowledgement

This work is supported by the Center for Multimedia Signal Processing of the Hong Kong Polytechnic University.

REFERENCES

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

[2] A. Hyvräinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.

[3] X. Giannakopoulos, J. Karhunen, and E. Oja, "An experimental comparison of neural algorithms for independent component analysis and blind separation," *Int. J. Neural Systems*, vol. 9, no. 2, pp. 99-114, April 1999.