頻域獨立成分分析法對於人工電子耳的應用分析
Analysis of Independent Component Analysis in Frequency-Domain on Cochlear Implant

桂禮璿[1] 李宜軒[*1] 蔡德明[2]
Li-Hsuan Kuei[1], Yi-Hsuan Lee[*1], Charles T. M. Choi[2]

摘要

人工電子耳使用者對語音的辨識能力遠低於聽覺正常者，特別是在有噪音的環境下。因此如何保留語音的完整性，對於欲將噪音抑制方法應用於人工電子耳而言至關重要。獨立成分分析（Independent Component Analysis，ICA）最初是用來處理雞尾酒派對（cocktail-party）問題，而獨立成分分析相較於其他噪音抑制方法，可以保留語音的完整性，不會產生音樂性噪音（musical noise）。本研究將時域獨立成分分析方法與頻域獨立成分分析方法應用於人工電子耳，試圖提高人工電子耳使用者辨識語音的能力，同時進行聽覺實驗評估其效能。由實驗結果得知使用獨立成分分析方法確實可提高語音辨識度。
關鍵詞：人工電子耳，獨立成分分析，噪音抑制，聽覺實驗

Abstract

The hearing ability in recognizing speech of cochlear implant (CI) users has been inferior than those of normal people, especially in noisy environments. Therefore, retaining the integral information of speech is of utmost importance to the noise reduction methods implemented on CI issue. Independent component analysis (ICA) was initially developed to deal with the cocktail-party problem. Compared to other noise reduction methods, ICA can preserve overall completeness of speech signals without generating music noise. In this study, Time-Domain ICA (TD-ICA) and Frequency-Domain ICA (FD-ICA) methods to increase the speech recognition rate to cochlear implant users is applied. Meanwhile, normal hearing experiments to evaluate these methods are also conducted. Experimental results show that ICA method does improve speech intelligibility.
Keywords: Cochlear Implant (CI), Independent Component Analysis (ICA), noise reduction, normal hearing experiment

I. INTRODUCTION

Sounds in the real world are always mixed with a diversity of noise ranging from different originators to echoes and reverberations [1]. *Cochlear implant* (*CI*) users have a higher speech intelligibility in quiet, but their hearing performance rapidly deteriorates in the presence of competing speakers and noises. So, unless noise reductions and separations are conducted, CI users will have difficulty in identifying the desired speech from background noises [2].

Consider a hearing impaired person who wears a cochlear implant. If the only speech processor receives distorted speech data, it will give wrong information to CI user and then reduce their speech intelligibility. Instead of directly eliminating noises based on band subtraction. *Independent Component Analysis* (*ICA*) extracts desired speech patterns from noisy signals. Thus the integrity of speech data can be retained, and the music noise will not be generated [3-4].

The primitive ICA assumes source signals are linearly mixed. However, composite signals in real world are actually convolutively mixed due to noisy and reverberant environment and cannot be well separated [3-4]. *Frequency-Domain* implementations involving complex valued signals have advantages over *Time-Domain* implementations in ICA [5]. Complex-valued ICA for instantaneous sound mixtures is applied in each frequency bin. The merit of this approach is that the ICA algorithm becomes simple and can be implemented separately at each

[1]國立臺中教育大學資訊工程學系 [2]國立交通大學電機工程學系
[*]Corresponding author. E-mail: yslee@mail.ntcu.edu.tw
[1] Department of Computer Science, National Taichung University of Education, Taichung, Taiwan, R.O.C.
[2] Department of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
Manuscript received 19 November 2012; accepted 4 January 2013

frequency [5], and the resulting mixtures separated by FD-ICA method is better than TD-ICA method. This study applies TD-ICA method and FD-ICA method to extract desired speech signal. We also conduct normal hearing experiments to evaluate whether our methods are more effective. Experimental results show that both TD-ICA and FD-ICA does improve the speech intelligibility and the recognition % for FD-ICA method is higher than TD-ICA method.

II. METHOD

Blind source separation (*BSS*) is an approach taken to estimate original source signals from their mixtures. Independent component analysis (ICA) is a dominated method to deal with this problem [6].

Fig. 1 illustrates the Frequency-Domain ICA formulation. The observed signals in which multiple source signals are mixed are given by $X(f) = A(f)S(f)$, where $X(f) = [X_1(f),\ldots,X_k(f)]^T$ is the observed signal vector, and $S(f) = [S_1(f),.\ .\ ,S_L(f)]^T$ is the source signal vector. $A(f)$ is the mixing matrix which is assumed to be complex-valued [7]. In the Frequency-Domain ICA, usually applies windowed Fourier transform to transfer mixed signals to frequency-domain, and does separation in each frequency bin individually. It may be more time efficient, but the derived permutation and scaling problems must be well resolved.

In this study, we follow steps to conduct FD-ICA. We use a FFT filter to divide the signal into several sub-bands. Each sub-band is still Time-Domain, and a complex-valued ICA is applied to separate. Then, speech-like signals from all sub-bands are directly combined to from the final Time-Domain signal.

III. RESULTS

A. Subjects

10 normal-hearing subjects (4 females and 6 males) between the ages of 22 and 26 years old (mean 23.4 years) participated in this study. All subjects were native speakers of Taiwanese Mandarin.

B. Stimuli

We use the material proposed in [8] which contains 300 Taiwanese Mandarin sentences for normal hearing experiment. Each sentence contains 7~10 words and designed to recognize the last word. These sentences are evenly classified into two categories: high predictability (HP) and low predictability (LP). A sentence with HP means the last word is easier to predict from the sentence. In contrast, a LP sentence doesn't contains any cues can predict its last word. We recruit a female talker to record this sentence material. Each sentences were recorded at a sampling rate of 44.1 kHz and a bit depth of 16 bits. After recording, each sentence would verify by five other subjects to ensure that all sentences could be fully recognized in quite condition. The masker was a multi-talker babble available from CD (AudiTec Ltd, St Louis).

C. Procedure

All programming performed the signal processing via software routines in MATLAB. The directional hearing was simulated by using head-related transfer functions (HRTFs) [9]. The input signal which combined speech and noise was processed by filtering it with the corresponding HRTFs for each angle of incidence. The CI simulation was created using a vocoder [10].

Fig. 2 shows how we use HRTFs to produce different source angle of incidence. Speech source is always fixedly originated in the front at 0 azimuth, and noise source is placed at an angle of incidence of 270° (S0N270) or 315° (S0N315) azimuth. The distance between source and KEMAR was 1.4 meters. Each sentences are mixed by multi-talker babble in 5/0/-5 dB *signal-to-noise ratio* (*SNR*). For the listening mode, stimuli were presented only vocoder in left ear to form unilateral CI-alone, and the headphone we used is AKG K181DJ.

D. Experimental Results

We selected TD-ICA and FD-ICA to increase the speech intelligibility of CI users. Fig. 3 shows the experimental result conducted by normal hearing subjects. In Fig. 3, the "Original" represents the signal before noise reduction. The recognition % of FD-ICA and TD-ICA are 22%~65% higher than the "Original" case. A *one way analysis of variance* (*ANOVA*) analyzes the recognition % of SNR, method, and source angle of incidences. From S0N270

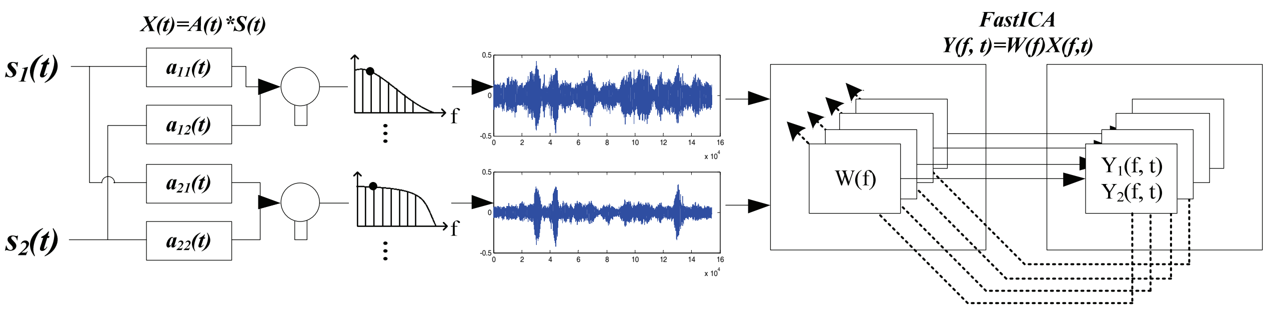

Fig. 1 Illustration of frequency-domain ICA formulation

and SON315 we can conclude that all cases of TD-ICA and FD-ICA are both significantly different against the “Original” case, except TD-ICA method with S0N315 under SNR5 condition. Between S0N270 and S0N315 the significant difference only presents under SNR0 condition. In addition, all SNR conditions are significantly different from each other and the recognition rates of FD-ICA are a little higher than TD-ICA.

Although from Fig. 3 the recognition % of FD-ICA is slightly higher than TD-ICA. Signals after de-noising using FD-ICA and TD-ICA are very different. Signals separated by TD-ICA tend to be unstable, which indicates separated signals may be all speech-like or noise-like. The reason may be the TD-ICA is not well suit to separate acoustic signals. In contrast, signals separated by FD-ICA can be more easily determined to speech-like or noise-like. These differences are not displayed in Fig. 3.

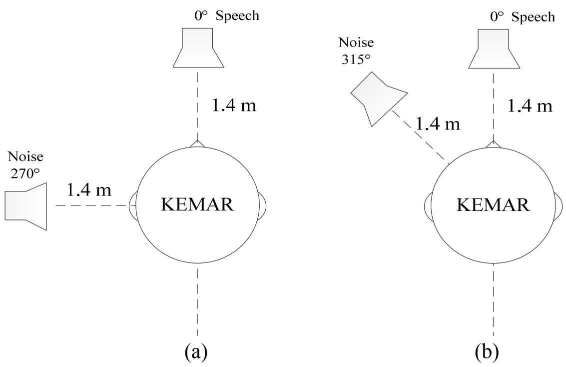

Fig. 2 The source angle of incidence in this experiment. Speech source is originated in the front at 0 azimuth. Noise source is placed at an angle of incidence of (a) 270˚, (b) 315˚

IV. CONCLUSIONS

In this study, we use noise reduction methods to increase the ability of CI user to recognize desired speech. TD-ICA method and FD-ICA method are selected to extract desired speech signal in order to retain the integrity of speech data. We also conduct normal hearing experiments to evaluate our methods. In our experimental results show that FD-ICA and TD-ICA can effectively improve speech intelligibility of CI users in almost condition. The recognition % of FD-ICA is even higher than TD-ICA in most case.

ACKNOWLEDGEMENTS

This research was supported by the National Science Council of the Republic of China under contract numbers NSC 100-2221-E-142-001. We would like to thank the assistance from the following people. Prof. C. T. M. Choi from NCTU provided the vocoder. Prof. H. C. Chen from NKNU provided Taiwanese Mandarin materials.

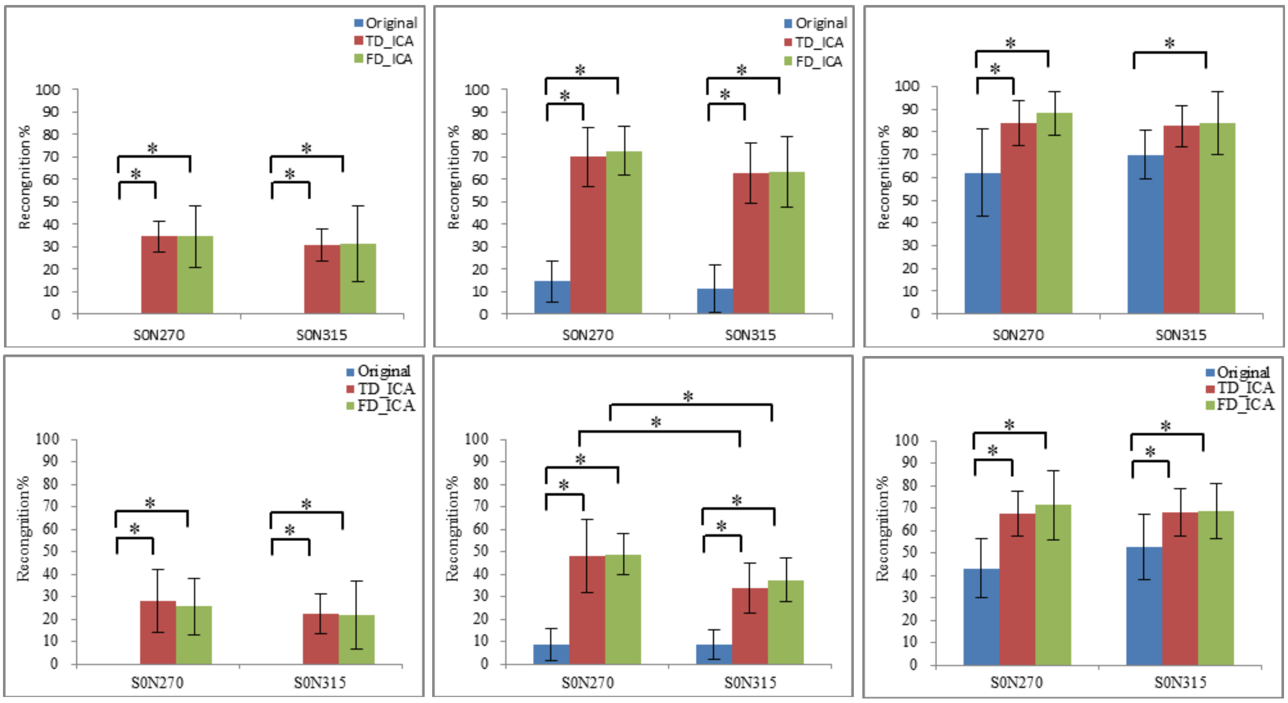

Fig. 3 The performance of speech recognition as a function of the source angle of incidence with comparing method (control group (Original), Time domain ICA and Frequency domain ICA). Recognition rates are plotted as the mean percent correct across subjects with the error bars indicating the standard deviation in SNR (a) -5, (b) 0, and (c) 5 dB for HP (upper panel) and LP (lower panel) sentence

REFERENCES

[1] H. Hakin and B. V. Veen, *Signals and Systems*, 2nd ed., John Wiley & Sons. Inc., 2003.

[2] B. S. Wilson and M. F. Dorman, "Cochlear implants: current designs and future possibilities," *Journal of Rehabilitation Research and Development*, vol. 45, no. 5, pp. 695-730, Feb. 2008.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley InterScience, 2001.

[4] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.

[5] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1-8, Feb. 2000.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Speech Audi. P.*, vol. 12, no. 5, pp. 530-538, Sep. 2004.

[7] H. Saruwatari, T. Kawamura, K. Sawai, A. Kaminuma, and M. Sakata, "Blind source separation based on fast-convergence algorithm using ICA and beamforming for real convolutive mixture," *in Proc. ICASSP*, pp. 921-924, 2002.

[8] 蔡志浩和陳小娟，「噪音背景辨識語音測驗編製研究」，*特殊教育研究學刊*，第 23 卷，第 121-140 頁，2002。

[9] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," Media Lab Perceptual Computing, Univ. MIT, Tech. Rep. #280, May 1994.

[10] C. T. M. Choi, C. H. Hsu, W. Y. Tsai, and Y. H. Lee, "A vocoder for a novel cochlear implant stimulating strategy based on virtual channel technology," *in Proc. ICBME*, pp. 310-313, 2009.