

# Separation of Multiple Speech Signals by using Triangular Microphone Array

Nozomu Hamada<sup>1</sup>, Non-member

## ABSTRACT

Speech source separation has been an important topic to realize speech-based human-machine interfaces or high quality hand-free communication with machines. For source separation, Independent Component Analysis (ICA) and time-frequency masking are powerful methods as a tool of Blind Source Separation (BSS) of speech mixtures. The latter method is based on the assumption called “W-Disjoint Orthogonality” which implies the cell component sparsity of speech in the time-frequency domain. One of the topics treated in this article is to introduce the time-frequency masking scheme is applied to the equilateral triangular array where the three delay estimates from each microphone pairs are obtained. In addition, it is used to improve histogram-mapping algorithm by integrate and coordinate transformation of three delay estimates. Some experiments in real environment for separating multiple sources are performed to verify the effectiveness.

**Keywords:** Separation of speech signals, ICA, time-frequency masking, hands-free communication, human-machine interfaces

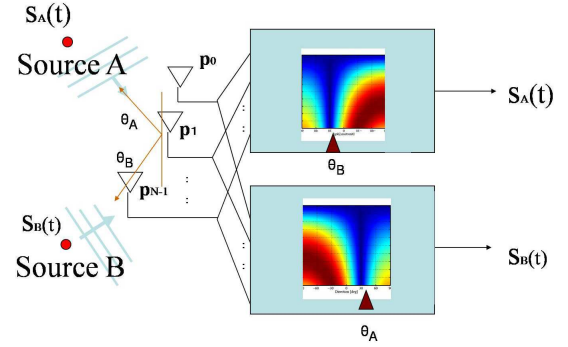
## 1. INTRODUCTION

### A. Natural Communication

Current sources are widely used in amplifiers, either single-stage or differential amplifiers. In these circuits, current source acts as a large resistor without consuming excessive voltage headroom. Some digital-to-analog converters (DAC) employed an array of current sources to produce an analog output proportional to digital input signal. Current sources, in conjunction with “current mirrors” can perform useful functions in analog signals. Some modifications to a current mirror, act as a low voltage cell can bring in lots of advantages to the analog design especially in the wireless communication field. With the trend towards fully integrated wireless transceivers which demand portable and low power consumption devices [1], a breakthrough of design techniques in transceivers is highly desirable.

Manuscript received on January 30, 2008 ; revised on January 20, 2008.

<sup>1</sup> The author is Department of Systems Engineering, Faculty of Science and engineering, Keio University, Yokohama, Hiyoshi, 223 8522, Japan, E-mail: hamada@hamada.sd.keio.ac.jp and hamadaab-sent@yahoo.co.jp



**Fig. 1:** Separation of signals by a pair of null beamformers

### B. Array Signal Processing

To implement basic and sophisticated sound capture systems, microphone array is indispensable. From the general signal processing viewpoints, sensor array system is considered as a spatial filter in order to enhance and suppress interference signal [4]. Depending on the required its role, beamformer (spatial band pass ) and/or null (spatial zero gain or notch) characteristics are realized. In particular, the sensor array system having null or zero gain characteristic to a specific direction is closely related to the separation of speech signals. Fig. 1 illustrates a source separation array system with a pair of null characteristic filters used to separate propagating signals according to its direction-of-angle (DOA). In the figure, upper processing system realizes zero gain to the direction angle  $\theta_B$  to suppress source B, then we obtain the signal A as its output. On the other hand, lower filter directs its null beam to the angle  $\theta_A$  to suppress signal A. Therefore, beamforming approach has to perform two main tasks. One is the propagating angle estimation ( $\theta_A$ ,  $\theta_B$ ), and the other is null beamformer realization to those directions.

### C. Cocktail Party Problem

The general issue of speech signal separation, called the “Cocktail party problem”, has been interested and investigated in the context beyond array signal processing [5]. The cocktail party problem is a challenging problem in human auditory perception, first proposed by Colin Cherry 1953. His definition is given by the following statement.

“One of our most important faculties is our ability to listen to, and follow, one speaker in the presence

of others. This is such a common experience that we may take it for granted: we may call it the cocktail party problem. No machine has been constructed to do just this, to filter out one conversation from a number jumbled together. (C. Cherry, 1957 [6])

As remarked in above sentence, the cocktail party problem is essentially multi-disciplinary field. The most fruitful and established progress is known as “*Human auditory analysis*” by Bregman [7]. From engineering points of views, cocktail party problem is solved by realizing array machine capable of separation sound of interest in the real noisy environment as human hearing system do. The outstanding feature of our listening systems is to extract a particular sound selectively with no prior information on source signal, such as its direction. Blind source separation (BSS) algorithm tends to install these abilities in microphone array system. This approach is to extract the source signals using observed their mixtures with no information on mixing process, such as room acoustics. The BSS algorithm should be installed in speech-based human-machine communication interfaces.

#### D. Various Aspects of Separation Problem

Fig.2 illustrates the following aspects of speech separation problem as follows.

*Nature of sound sources*= { Fixed source location / Moving source }, *Environment Noise*= { Directional / Non-directional noise }, *Environment (Room) Acoustic*= { Echoic / Anechoic }, *Mixing Mechanism*= { Linear Convolution / Instantaneous }, *Sensor Characteristic*= { Omni-directional / Directional sensors }, *Separation System*= { Time Domain / Frequency Domain Filtering }, *Output Signal*= { Monaural / Stereo Sound }, *Number of Sources (N)*= { Known / Unknown }, *Number of Sensors (M)*= { M>N Over-determined Case / M<N Under-determined Case },

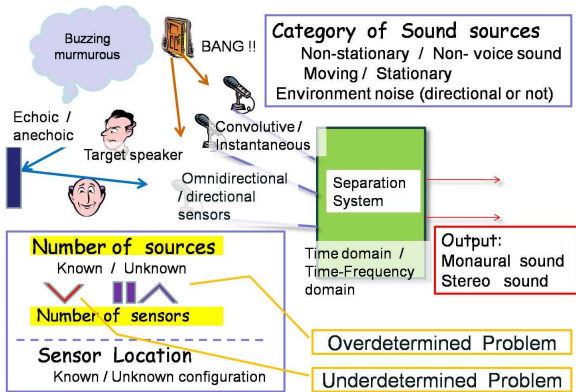


Fig.2: Various Aspects of Separation Problems

The most commonly studied and accepted problem in practice is the following conditions.

{Fixed source location, Convolutional mixing process, Omni-directional sensors: Other than these fac-

tors are not preconditioned.}

This problem setting gives the mixing formulation as follows. Consider N source signals  $s_i(n)$  ( $i = 1, \dots, N$ ), and their mixed observations by M sensors. The incoming microphone signals  $x_j(n)$  ( $j = 1, \dots, M$ ) are modeled by N-input/M-output (MIMO) dynamical system as the next convolution.

$$x_j(n) = \sum_{i=1}^N \sum_{l=1}^L h_{ji}(l) s_i(n-l+1) \quad (j = 1, \dots, M) \quad (1)$$

where  $h_{ji}(l)$  is the impulse response from a source  $i$  to a sensor  $j$ . The BSS is solved by estimating the separated signals  $\hat{s}_k(n)$  ( $k = 1, \dots, N$ ) by using only the observations  $x_j(n)$  ( $j = 1, \dots, M$ ).

Since the BSS is inherently an inverse system problem for MIMO linear dynamical systems, existing blind deconvolution schemes are applied. In particular, current major studies are based on either of the followings.

- (a) Independent component analysis (ICA) [8-10]
- (b) Time-frequency masking algorithm [11-14]
- (c) Combination of ICA and Time-frequency masking [15,16]

The ICA uses statistical independency of source signals to obtain the separation system with no prior information. The methods are separated into two groups depending on the domain where ICA adaptive algorithm is performed, which are called the time-domain ICA and the frequency domain ICA (FDICA) respectively. In the FDICA [10], short time Fourier transform (STFT) is applied to convert the convolutive mixtures to the corresponding complex-valued instantaneous mixtures as follows.

$$X_j(k, l) = \sum_{i=1}^N H_{ji}(k) S_i(k, l) \quad (j = 1, \dots, M) \quad (2)$$

$H_{ji}$  is the transfer function from a  $i$ -th source to the  $j$ -th sensor,  $S_i(k, l)$  and  $X_j(k, l)$  denote short-time Fourier transformed sources and observed signals, respectively. where,  $k$  is the frequency index and  $l$  is the time frame of STFT. Therefore, the mixing mechanism is transformed to a tractable memoryless linear system at each frequency bin. The merit of it is its simple inverse algorithm due to the separability of the BSS.

## 2. TIME-FREQUENCY MASKING BY TRIANGULAR ARRAY

### A. W-Disjoint Orthogonality and Cell Clustering

Second BSS approach, known as time-frequency masking relies on the sparseness of the speech signals in the short time frequency domain. Instead of realizing the deconvolution filters as in FDICA, it directly separates the time-frequency components of mixed signal into each source and then synthesizes

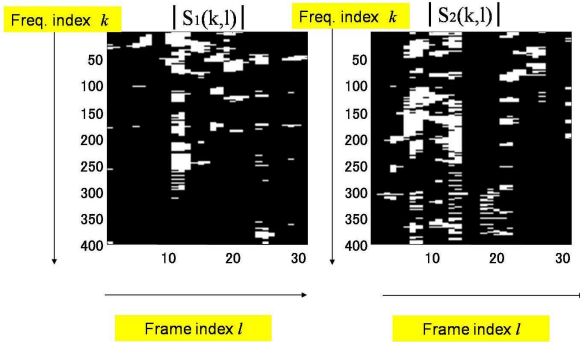
it. To this end, time-frequency masking is based on the assumption which is called W-disjoint orthogonality (WDO) between speech signals in time-frequency domain [11]. The WDO implies the following properties:

(a) *Speech signal has the sparsity against both time and frequency. Thus, (b) in STFT domain, though the observed speech signal is the mixture of several sources, most part of the time-frequency cells contain a component of at most one of the source signals.*

Let us consider two speech signals  $s_1(t)$ ,  $s_2(t)$  and their short time Fourier transform representations  $S_1(k, l)$ ,  $S_2(k, l)$ . The property (b) can be described mathematically as

$$S_1(k, l) \cdot S_2(k, l) \approx 0 \quad \forall(k, l) \quad (3)$$

Fig.3 shows the binary spectrogram images of amplitude  $|S_1(k, l)|$ , and  $|S_2(k, l)|$ . White cells in the graph show the components of each signal respectively. From this figure, we observe the sparsity of speech itself and also the WDO property between two speech signals.



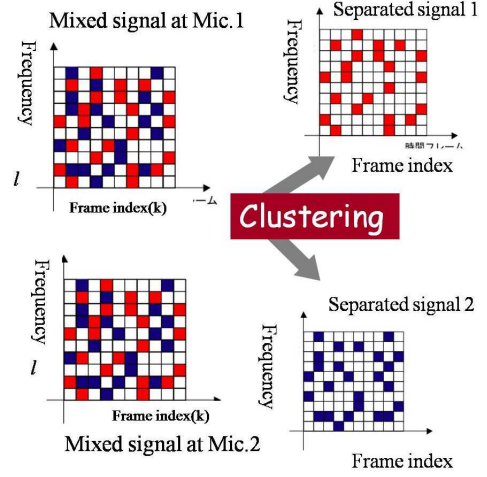
**Fig.3:** Sparseness and WDO properties of speech signals (Binary images of Fourier spectrum)

Most fundamental separation process in the masking method is to cluster time-frequency cells. Above WDO property ensures to develop the following separation or clustering mechanism. Figure 4 illustrates cell clustering mechanism. For a given time-frequency cell decomposition, if we could know to which source does each cell component belongs, we can separate the components into each corresponding source. Thus the parameterization of each cell utilized at classification phase is the task. Commonly used features of T-F cell are attenuation ratio and phase difference between a pair of microphones.

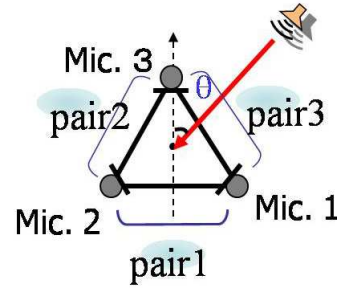
#### B. Triangular Array System

Our study [17] proposed a BSS system using three microphones located at the vertices of equilateral triangle as illustrated in Fig.5. In this section, the case of two speech signals is considered.

The triangular array configuration is the minimum number of sensor array configuration which copes



**Fig.4:** Clustering scheme of time-frequency cells



**Fig.5:** Equilateral triangular microphone array

with separation of arbitrary pair of speech sounds. The flow of separation process is given as follows.

(i) *STFT* Three observed signals  $x_1(t)$ ,  $x_2(t)$ ,  $x_3(t)$  by microphones are converted into time-frequency domain by STFT with appropriate window function. We set these in the vector form as,

$$\mathbf{X} = [X_1(k, l) \quad X_2(k, l) \quad X_3(k, l)]^T \quad (4)$$

(ii) *Delay Estimation* The arrival delays between each pair of microphones are estimated at each time-frequency cell. The estimated delays are integrated as a 3-D vector  $\delta(\mathbf{k}, l)$ .

$$\delta(\mathbf{k}, l) := [\delta_1(\mathbf{k}, l), \delta_2(\mathbf{k}, l), \delta_3(\mathbf{k}, l)]^T \quad (5)$$

Above delay  $\delta_i(k, l)$  is estimated by means of several ways. A convenient estimation method is the phase correlation method, which is written by the following equations.

$$\delta_i(k, l) = \frac{K}{2\pi k} \tan^{-1} \left\{ \frac{\text{Im}(\text{Corr}_i(k, l))}{\text{Re}(\text{Corr}_i(k, l))} \right\} \quad (6)$$

$$\text{Corr}_i(k, l) = \frac{X_i(k, l)X_{i+1}^*(k, l)}{|X_i(k, l)||X_{i+1}(k, l)|} \quad (X_4 = X_1) \quad (7)$$

where  $*$  denotes the complex conjugate. The amount of delay is the reflection of the source direction. Let us define theoretical delay values as a function of direction-of-arrival  $\theta$  by  $\Delta(\theta) := [\Delta_1(\theta), \Delta_2(\theta), \Delta_3(\theta)]^T$  between each pair of microphone. Then, equilateral triangular array configuration provides the following constraints on the delay elements.

$$\Delta_1^2(\theta) + \Delta_2^2(\theta) + \Delta_3^2(\theta) = \frac{3}{2} \left(\frac{d}{c}\right)^2 \quad (8)$$

$$\Delta_1(\theta) + \Delta_2(\theta) + \Delta_3(\theta) = 0 \quad (9)$$

Both conditions are satisfied on the circle, denoted by  $\mathbf{C}$ , in 3-D delay vector space. Since the relation (9) is valid for the estimated delay vector, all vectors  $\delta(\mathbf{k}, \mathbf{l})$  lie on the plane. We denote this plane by  $\mathbf{D}$ .

(iii) *Projection on 2-D Space* we dot a point at the place where delay vector of each T-F cell indicates. We investigate that the distribution of all delay vectors concentrate on plane  $\mathbf{D}$ . Actually estimated delay vectors (dots) would be spread around the circumference as shown in Fig.6., due to estimation error and noise effect.

Next, the estimated delay vectors are projected on 2-D plane  $\mathbf{D}$  as follows. Two orthonormal basis  $\mathbf{e}_1 = [0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]^T$ ,  $\mathbf{e}_2 = [\frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}]^T$  on  $\mathbf{D}$  are introduced. Then, the projection of  $\delta(\mathbf{k}, \mathbf{l})$  onto  $\mathbf{D}$  gives 2-D vector

$$\mathbf{f}(\mathbf{k}, \mathbf{l}) := \begin{bmatrix} f_1(k, l) \\ f_2(k, l) \end{bmatrix} = \begin{bmatrix} \delta(\mathbf{k}, \mathbf{l})^T \mathbf{e}_1 \\ \delta(\mathbf{k}, \mathbf{l})^T \mathbf{e}_2 \end{bmatrix} \quad (10)$$

We obtain the distribution of  $\mathbf{f}(\mathbf{k}, \mathbf{l})$  on  $\mathbf{D}$  as shown in Fig.7

(iv) *Histogram Peaks and Clustering, Masking* Using the vectors which locate on and around the circumference, we make a histogram and detect the peaks of distribution. These peaks are used to cluster the time-frequency cells. (Fig. 8)

According to this clustering criterion, the binary masks taking out the cell components are made.

Finally, clustered cell components are converted into time domain using the inverse STFT, such as overlap-add (OLA) method for smoothing the separated signals.

In the proposed system, clustering is based on the location of delay vector associated with each cell. The separation strategy here uses linear discriminate. In fact, each histogram peaks of the clusters are determined and the Euclidian distance between delay vector (dot) of cell at  $(k, l)$  and these peaks is used. For a cell component, distances between its delay vector and two centers are calculated. Then, the cell is identified into the component of a cluster whose peak location is closer to it. For determining center of cluster in  $\mathbf{D}$ , histogram of  $\mathbf{f}(\mathbf{k}, \mathbf{l})$  around the circle  $\mathbf{C}'$  (projection of  $\mathbf{C}$  on  $\mathbf{D}$ ) is utilized. Namely, we use

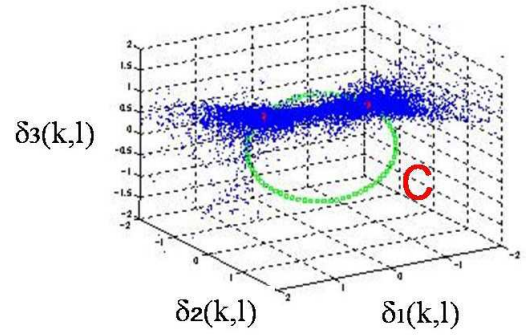


Fig.6: 3D space distribution of delay vectors

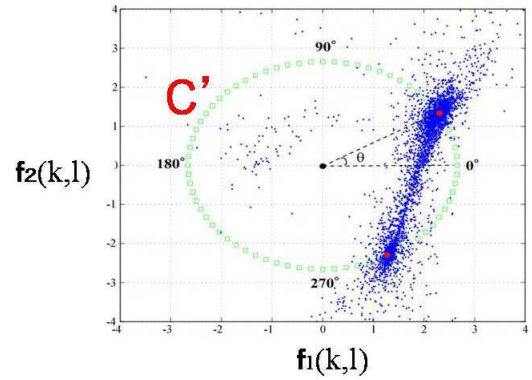


Fig.7: Projection on 2D space  $\mathbf{D}$  and dot distribution

the delay vectors satisfying the following conditions.

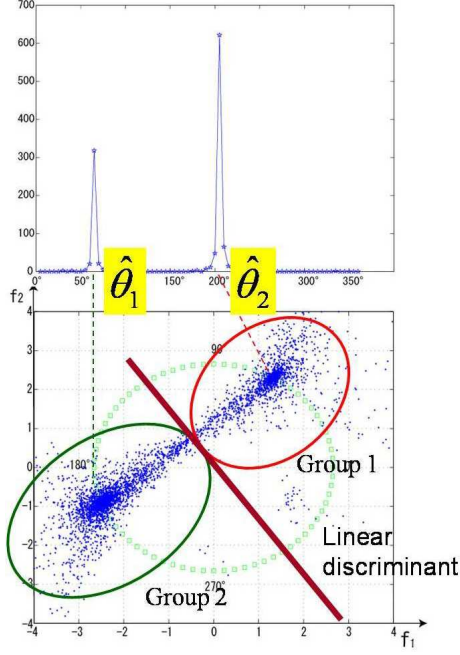
$$\frac{3}{2} \left(\frac{d}{c}\right)^2 - \alpha \leq \|\mathbf{f}(\mathbf{k}, \mathbf{l})\| \leq \frac{3}{2} \left(\frac{d}{c}\right)^2 + \alpha \quad (11)$$

Where  $\alpha$  is a parameter determining the width of the region for histogram. The angle of the position on  $\mathbf{C}'$  is quantized into each 5 degree levels in the range  $[0, 360]$ . The quantization element of this level is shown as the small square cells. Fig.8 shows an example of the resulted histogram, and its two prominent peaks located at the angles  $\hat{\theta}_1$  and  $\hat{\theta}_2$ . We denote the vectors corresponding these positions on  $\mathbf{D}$  by  $\mathbf{f}_{\theta_1}, \mathbf{f}_{\theta_2}$ . Then, the linear discrimination criterion yields the separation masks.

$$M_1(k, l) = \begin{cases} 1, & \|\mathbf{f}(\mathbf{k}, \mathbf{l}) - \mathbf{f}_{\theta_1}\| < \|\mathbf{f}(\mathbf{k}, \mathbf{l}) - \mathbf{f}_{\theta_2}\| \\ 0, & \text{elsewhere} \end{cases} \quad (12)$$

$$M_2(k, l) = \begin{cases} 1, & \|\mathbf{f}(\mathbf{k}, \mathbf{l}) - \mathbf{f}_{\theta_1}\| > \|\mathbf{f}(\mathbf{k}, \mathbf{l}) - \mathbf{f}_{\theta_2}\| \\ 0, & \text{elsewhere} \end{cases} \quad (13)$$

All the time-frequency cells  $(k, l)$  are separated into two groups by multiplying the binary masks  $M_1(k, l), M_2(k, l)$  to the observed time-frequency cell



**Fig.8:** Delay Histogram and clustering by linear discrimination

domain of  $X_1(k, l)$ . Namely

$$\hat{S}_i(k, l) = M_i(k, l)X_1(k, l) \quad (i = 1, 2) \quad (14)$$

The  $\hat{S}_1(k, l), \hat{S}_2(k, l)$  are inversely transformed to the time domain using OLA method in order to generate original signals.

### 3. EXPERIMENTS AND EVALUATION

We show the result of experiments in a conference room using Japanese voiced sounds, known as "ASJ Continuous Speech Corpus for Research". Several parameter's values are listed below.

Array aperture(d): 40mm, Height of speakers & microphone array: 1.20m, Room size: 15m×18m,

Sampling rate: 8 kHz, SIFT flame length: 1024 sample, Window: Hamming, Flame overlap: 521 sample.

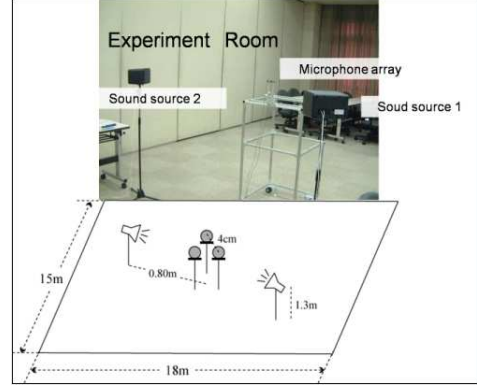
As the value of performance evaluation, we use the measure of W-disjoint Orthogonality (WDOM). It is computed from two other criteria PSR (the preserved-signal ratio) and SIR (the signal-to-interference ratio) defined as follows: [11],

$$WDOM := PSR_M - \frac{PSR_M}{SIR_M} \quad (15)$$

$$PSR_M := \frac{\|M_i(k, l)S_i(k, l)\|^2}{\|S_i(k, l)\|^2} \quad (16)$$

$$SIR_M := \frac{\|M_i(k, l)S_i(k, l)\|^2}{\|M_i(k, l)S_j(k, l)\|^2}$$

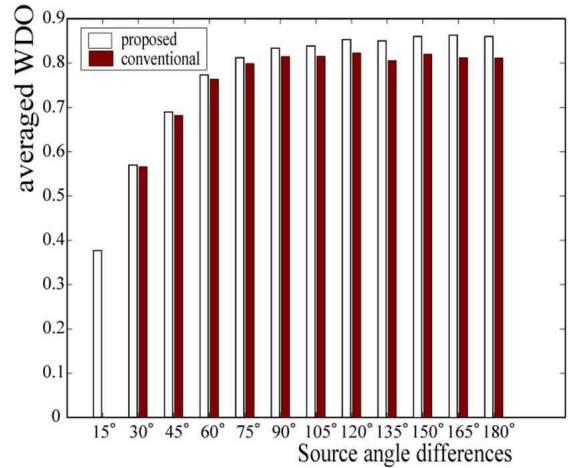
$M_i(k, l)$  is the corresponding binary mask.  $S_i(k, l)$ .



**Fig.9:** Experiment room and devices

When  $WDOM = 1$ , separation is perfectly attained, because it implies that  $PSR_M = 1$  and  $SIR_M = \infty$ . In order to have  $WDOM \simeq 1$ , i.e., well demixing performance, it must simultaneously preserve the energy of interest while suppressing the energy of the interference.

Fig.10 shows the averaged WDOM values with respect to source angle difference. The conventional method in this case is the masking method using a pair of microphone which is selected from three pairs as the most suitable pair to separate the delay histogram. This result proves that the proposed triangular array system exceeds the conventional method for every source angle differences. In the case of 15°, the conventional method fails to separate the signals, because it could not estimate the location of two peaks correctly.



**Fig.10:** Variance of WDO with respect to angle difference

### 4. CONCLUSIONS REMARKS

In the first part of this paper, we took a brief look at the blind source separation problems using microphones. Array signal processing, especially null

or zero gain formation is the very important key characteristic to separation. Then, a blind speech separation method using equilateral triangular array is introduced. Triangular microphone array gives three delays between the pair of microphones. The constraints on these delays, brought by the unilateral triangularity, are utilized in the process of time-frequency cell clustering. In particular, histogram peak search would become more reliable by the use of delay vector on the constrained circle. Experiments show that the method achieves a high performance separation even for mixtures of closely located two speakers. From other experimental results, we can confirm that the proposed method improves the separation performance in a real acoustic environment.

As we mentioned in the introduction, speech separation is even one field of the cocktail party problem (CCP). For future studies on the BSS, the cocktail party problem approach would give significant suggestions. One of which is given by S. Haykin as a criticism from auditory scene analysis [5]. (i) Most ICA/BSS algorithms require that the number of sensors is assumed to be larger than and equal to the number of sources. But the human auditory system requires merely two outer ears to solve CPP. The ICA framework usually requires the separation of all source signals. But our auditory systems focus on extracting a single speaker of interest. The ICA learning algorithms assume constant number and direction signal sources, namely constant auditory scene. In future, BSS system corresponding to the case that more than three sources are exist, and to the strong reverberant case should be considered. To these challenging problems, moving sound source separation have been started [18]. At last, we would like to introduce one research which stands at the definitely opposite side from human auditory system. That is the largest microphone array in the world which makes beamformer by use of 1020 microphones [19].

## 5. ACKNOWLEDGEMENTS

The author wishes to acknowledge Mr. Yosuke Takenouchi for his pioneering work on triangular microphone array in Keio University. Also wishes to Mr. Masahiro Yashita for his help on this manuscript preparing, also to Mr. Masahiro Yashita for his help on this manuscript preparing.

## References

- [1] I. Marsic, A. Medel, and J. Flanagan, "Natural Communication with Information Systems," *Proceedings of the IEEE*, Vol.88, No.8, pp.1354–1365, 2000.
- [2] Special Issue on Integrated technologies of Robotic Hearing, (Japanese), *Journal of the Society of the Instrument and Control Engineers Japan*, Vol.46, No.6, June 2007
- [3] [http://www.aist.go.jp/aist\\_e/latest\\_research/2005/20050620/20050620.html](http://www.aist.go.jp/aist_e/latest_research/2005/20050620/20050620.html)
- [4] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*, Prentice-Hall, 1983
- [5] S. Haykin, *The Cocktail Party Problem*, Neural Computation Vo.17, pp.1875–1902, 2005
- [6] E.C.Cherry, "Some Experiments on the recognition of speech, with one and two years," *Journal of the Acoustical Society of America*, Vol.25, pp.975–979, 1953
- [7] A.S.Bregman, "Auditory scene analysis: The perceptual organization of sound." Cambridge, MA; MIT Press
- [8] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & sons, 2001.
- [9] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, pp.101–104, April 1997.
- [10] S. Makino, "Blind source separation of convolutive mixtures", *Proceedings SPIE*, 624
- [11] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," *IEEE Trans. on signal processing*, Vol.52, No.7, pp.1830–1847, 2004.
- [12] S. Makino, H. Sawada, R. Mukai, and S. Araki, "Blind source separation of convolutive mixtures of speech in frequency domain," *Invited in IEICE Trans. Fundamentals*, vol. E88-A, no. 7, pp. 1640–1655, July 2005.
- [13] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "A combined approach of array processing and independent component analysis for blind separation of acoustic signals," *Proc. ICASSP 2001*, pp.2729–2732, May 2001.
- [14] J. Huang, N. Ohnishi, N. Sugie, "A Biomimetic System for Localization and Separation of Multiple Sound Sources," *IEEE Trans. on Instrumentation and Measurement*, Vol.44, No.3, pp.733–738, June. 1995.
- [15] M.S.Pedersen, et al, "Overcomplete blind source separation by combining ICA and binary time-frequency masking", *IEEE International workshop on Machine Learning for Signal Processing*, pp.15–20, 2005
- [16] F. Yang, N. Hamada, "Solution of Underdetermined Speech Separation Problems by Combining ICA and Time-Frequency Masking Methods," *IEICE Technical Report*, vol.107, no.239, SP2007–57, pp.1–6, Sept. 2007.
- [17] Y. Takenouchi and N.Hamada, "Time-frequency masking for BSS problem using equilateral triangular microphone array", *Proceedings of 2005 International Symposium on Intelligent*

*Signal Processing and Communication Systems*, pp.185-188, Dec. 13-16, 2005

- [18] A. Fujita and N. Hamada, "Separation of Moving Sound Sources by Time-Frequency Masking Method", *Journal of Signal Processing*, Vol.10, No.4, pp.271-274, July 2006.
- [19] E. Weinstein, K. Steele, A. Agarwal, J. Glass, "LOUD: A 1020-Node Microphone Array and Acoustic Beamformer," *Proc. ICSV*, pp.571-578, Cairns, July. 2007.



**Nozomu Hamada** received the B.S., the M.S. and Ph.D. degrees in electrical engineering from Keio University. In 1974, he became an Instructor in electrical engineering at Keio University. He has been a Professor there in the Department of System Design Engineering since 1991. He was the visiting researcher in Australian National University in 1982. Currently, he is the adjunct professor of Xi'an Jiaotong University and Xi'an Jiaotong University City College.

His research interests include circuit theory, stability theory of dynamical system, design of one or multi-dimensional digital filters, and image processing. One of his recent research fields is a realization of human interface system using microphone array. The main topics in this study are acquisition of audio signal from spatially distributed sound sources and the separation of multiple speech signals by ICA. He is the author of "Linear Circuits, Systems and Signal Processing (Chapter 5) (Marcell Dekker Inc. 1990), "Two-Dimensional Signal and Image Processing" (SICE 1996), "Introduction of Modern Control Systems" (Corona Pub. Inc. 1997). He was the chair of IEEE signal Processing Society in Japan Chapter (2004) and editorial board in *Journal of Signal Processing*. He is the guest editor of special issue relevant to multi-dimensional signal processing: its application and realization technique in the IEICE (2000), *Signal Processing* (2002, 2003, 2006, 2007), etc.