A SYSTEM FOR SEPARATING SOUND SOURCES PROPAGATED IN THE SAME DIRECTION

Kazuho ONO and Akio ANDO

Acoustics and Audio Signal Processing, NHK Science and Technical Research Laboratories
E-mail: {ono.k-gs, ando.a-io}@nhk.or.jp

ABSTRACT

This paper describes a microphone system that separates a target sound from other noise arriving in a single direction when the target cannot, therefore, be separated by directivity control. Microphones are arranged in a line toward the sources to form null sensitivity points at certain distances from the system. The null points exclude non-targeted sound sources on the basis of weighting coefficients for microphone outputs determined by blind source separation. The separation problem is thereby simplified into the instantaneous separation by adjusting the time-delays for microphone outputs. The system uses a direct (i.e. non-iterative) algorithm for blind separation based on second-order statistics, assuming that all sources are non-stationary signals. Simulations show that the 2-microphone system separates a target sound with separability of more than 40dB in the 2-source problem, and 25dB in the 3-source problem.

1. INTRODUCTION

A microphone system is investigated that separates a target sound from background noise [1]. A target sound arriving in a different direction from the noise can be separated effectively by a microphone system with sharp directivity. If, however, the sound arrives in the same direction as a noise source, such signal processing technology as blind source separation is required to separate it. This paper deals with the separation of unidirectional sounds. To solve the problem, the microphones are arranged in a line toward the sources, as shown in Fig.1. The time-delay in the output of microphone-1 in Fig.1 is adjusted to the time of the propagation of sound by inter-microphone distance, such that the signals of two outputs are in phase. This provides the important advantage that only the instantaneous mixing model needs to be considered.

In the blind separation of instantaneous signal mixtures, an N-dimensional measured signal $x(t)=[x_1(t), \cdots, x_N(t)]^T$ is assumed to be observed at each time point t, such that

$$x(t) = As(t) \ , \qquad (1)$$

where A is an unknown matrix of N^2 coefficient a_{ij}, and $s(t)=[s_1(t), \cdots, s_N(t)]^T$ is an N-dimensional unknown stochastic, independent source signal. The measured signals are processed by a linear operation where

$$y(t) = Cx(t) \ , \qquad (2)$$

in which C is a separation matrix of N^2 coefficient c_{ij}, and $y(t)=[y_1(t),\cdots,y_N(t)]^T$ is an N-dimensional output signal (an estimate of the source signal $s(t)$). The diagonal elements of C are set to 1 in this paper.

Many studies have considered the problem of blind separation of an instantaneous mixture of sources. Most of these use higher-order statistics to satisfy the statistical independence of the elements of output signal $y(t)$ [2]. Other approaches involve the use of second-order statistics on the assumption that the sources are non-stationary signals [3]. It is well known that the single-time decorrelation of elements of the output signal is insufficient to determine separation matrix C uniquely. Decorrelation at multiple points in time, however, adds other constraints that make it possible to determine C uniquely, except in the case of permutations [4]. These approaches have the advantage over the methods of

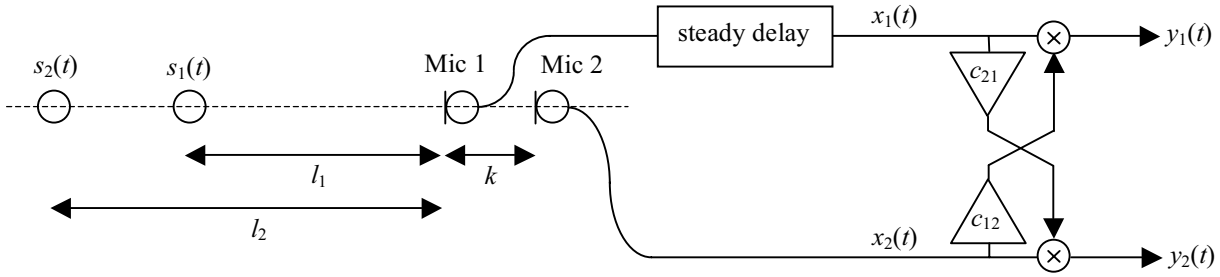

Fig.1 Block diagram of the 2-microphone system

higher-order statistics of providing stable statistical estimation and a smaller computational burden.

This paper describes the principles of the new microphone system and proposes a new algorithm for blind source separation using second-order statistics [5]. In contrast to almost all existing algorithms of blind source separation, which use an iterative algorithm to determine C, the new algorithm uses a direct (i.e. non-iterative) procedure. There is, therefore, no need to consider the stability and convergence of the algorithm.

2. SEPARATION MECHANISM OF THE SYSTEM

We assume that each source produces a wave that spreads spherically outward. The wave does not change shape as it spreads out, but its amplitude diminishes in proportion to the propagation distance. Since all sources and microphones lie on the same line, as shown in Fig.1, those positions can be represented by a 1-dimensional coordinate. We then denote the position of a source $s(t)$ as ξ, the positions of two microphones as ξ_1 and ξ_2 ($\xi < \xi_1 < \xi_2$), and the observed signals of $s(t)$ at ξ_1 and ξ_2 as $f_1(\xi, t)$ and $f_2(\xi, t)$, respectively. From these assumptions, $f_1(\xi, t)$ is given by

$$f_1(\xi, t) = \frac{1}{\xi_1 - \xi} s(t - \tau) \quad . \tag{3}$$

In (3), τ is the time-delay of sound propagation between the source and microphone at ξ_2, because the steady time-delay in the system is properly adjusted as described in the preceding section. $f_2(\xi, t)$ is also given by

$$f_2(\xi, t) = \frac{1}{\xi_2 - \xi} s(t - \tau) \quad . \tag{4}$$

Then a process, which outputs a weighted sum

$$f_1(\xi, t) + c_{12} f_2(\xi, t) = \left(\frac{1}{\xi_1 - \xi} + \frac{c_{12}}{\xi_2 - \xi} \right) s(t - \tau)$$
$$= \frac{c_{12}\xi_1 + \xi_2 - (1 + c_{12})\xi}{(\xi_1 - \xi)(\xi_2 - \xi)} s(t - \tau) \quad ,$$

can generate a null point at

$$\xi = \frac{c_{12}\xi_1 + \xi_2}{1 + c_{12}} \quad , \tag{5}$$

and suppress the source signal at that point.

Another process, which outputs a weighted sum $c_{21} f_1(\xi, t) + f_2(\xi, t)$, can also suppress the source at the point

$$\xi = \frac{\xi_1 + c_{21}\xi_2}{1 + c_{21}} \quad . \tag{6}$$

The system controls coefficient c_{12} or c_{21} to produce a null point in one source position and picks up the other source signal when the number of sources is 2. The next section gives a definite algorithm for solving coefficients c_{12} and c_{21} automatically.

3. NEW SEPARATION ALGORITHM

A direct decorrelation method exists for the 2-source separation problem [6], which can solve a pair of non-linear simultaneous equations to realize the decorrelation at two time points. While this method avoids the stability problem of iterative learning, it is sensitive to the estimation instabilities of the correlation functions of the measured signal because it is based only on the estimated values of correlation functions at two time points. Another problem with this method is that the solution at a given time often becomes a permutated solution of the previous time. The new algorithm can decorrelate the elements of the output signal at more than two time points. It is a two-stage algorithm. In the first stage, linear multiple regression coefficients among correlation functions of the measured signals are estimated at multiple points in time. The elements of the separation matrix are calculated to form the coefficients in the second stage.

We assume N=2 in this paper. The output signals are then given by

$$\begin{aligned} y_1(t) &= x_1(t) + c_{12} x_2(t) \\ y_2(t) &= c_{21} x_1(t) + x_2(t) \end{aligned} \quad . \tag{7}$$

The cross-correlation between $y_1(t)$ and $y_2(t)$ is written by

$$r_y^{(12)}(t) \equiv E\left[y_1(t) y_2(t) \right]$$
$$= c_{21} r_x^{(1)}(t) + c_{12} r_x^{(2)}(t) + (1 + c_{12} c_{21}) r_x^{(12)}(t) \quad , \tag{8}$$

where

$$r_x^{(kl)}(t) \equiv E\left[x_k(t) x_l(t) \right] \quad , \tag{9}$$

and $r_x^{(11)}(t)$ and $r_x^{(22)}(t)$ are simply denoted by $r_x^{(1)}(t)$ and $r_x^{(2)}(t)$, respectively. The decorrelation of the output signal is realized by c_{12} and c_{21}, satisfying

$$c_{21} r_x^{(1)}(t) + c_{12} r_x^{(2)}(t) + (1 + c_{12} c_{21}) r_x^{(12)}(t) = 0 \quad . \tag{10}$$

Eq. (10) is a non-linear equation in the variables c_{12} and c_{21}, but a linear equation in the variables $r_x^{(1)}(t)$, $r_x^{(2)}(t)$ and $r_x^{(12)}(t)$. From (10), we obtain the following linear relationship between the correlation functions:

$$r_x^{(12)}(t) = -\frac{c_{21}}{1 + c_{12} c_{21}} r_x^{(1)}(t) - \frac{c_{12}}{1 + c_{12} c_{21}} r_x^{(2)}(t) \quad . \tag{11}$$

The first stage of the proposed algorithm applies the following linear multiple regression model to (11) :

$$r_x^{(12)}(t) = -w_1 r_x^{(1)}(t) - w_2 r_x^{(2)}(t) \quad , \tag{12}$$

and solves the regression coefficient w_1 and w_2 at multiple points in time $t_1, t_2, \ldots, t_m$ (e.g. $t_i = t - (i-1)T$, T:time interval):

$$\begin{pmatrix} r_x^{(1)}(t_1) & r_x^{(2)}(t_1) \\ r_x^{(1)}(t_2) & r_x^{(2)}(t_2) \\ \cdot & \cdot \\ \cdot & \cdot \\ r_x^{(1)}(t_m) & r_x^{(2)}(t_m) \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = - \begin{pmatrix} r_x^{(12)}(t_1) \\ r_x^{(12)}(t_2) \\ \cdot \\ \cdot \\ r_x^{(12)}(t_m) \end{pmatrix} \quad . \tag{13}$$

The coefficients can be solved by least square estimation. Then, from the non-linear equations

$$\frac{c_{21}}{1+c_{12}c_{21}} = w_1 , \quad \frac{c_{12}}{1+c_{12}c_{21}} = w_2 \quad ,$$

the second stage gives the two solutions of c_{12} and c_{21} as

$$c_{12} = \frac{1+\sqrt{1-4w_1w_2}}{2w_1}, \quad c_{21} = \frac{1+\sqrt{1-4w_1w_2}}{2w_2} \qquad (14)$$

and

$$c_{12} = \frac{1-\sqrt{1-4w_1w_2}}{2w_1}, \quad c_{21} = \frac{1-\sqrt{1-4w_1w_2}}{2w_2} \quad . \qquad (15)$$

It should be noted that Eqs. (14) and (15) are the alternative permutations.

4. PERMUTATION CONTROL

Permutation is one of the drawbacks of blind source separation. There are some methods to solve the problem by controlling the directivity pattern of the microphone array [7]. This paper proposes a different method to control permutation based on the distance between the null point and the microphones.

By substituting (1) for (2), we obtain

$$\begin{aligned} y_1(t) &= b_{11}s_1(t) + b_{12}s_2(t) \\ y_2(t) &= b_{21}s_1(t) + b_{22}s_2(t) \end{aligned} \qquad (16)$$

where

$$\begin{aligned} b_{11} &= a_{11} + c_{12}a_{21}, \quad b_{12} = a_{12} + c_{12}a_{22} \\ b_{21} &= c_{21}a_{11} + a_{21}, \quad b_{22} = c_{21}a_{21} + a_{22} \end{aligned} \quad . \qquad (17)$$

Thus, according to Fig.1, we obtain

$$\begin{aligned} y_1(t) &= \left(\frac{1}{l_1} + \frac{c_{12}}{l_1+k}\right)s_1(t) + \left(\frac{1}{l_2} + \frac{c_{12}}{l_2+k}\right)s_2(t) \\ y_2(t) &= \left(\frac{c_{21}}{l_1} + \frac{1}{l_1+k}\right)s_1(t) + \left(\frac{c_{21}}{l_2} + \frac{1}{l_2+k}\right)s_2(t) \end{aligned} . \qquad (18)$$

From (18), one solution is obtained when

$$l_1 = -\frac{c_{21}}{1+c_{21}}k, \quad l_2 = -\frac{1}{1+c_{12}}k \quad . \qquad (19)$$

The other solution is obtained when

$$l_1 = -\frac{1}{1+c_{12}}k, \quad l_2 = -\frac{c_{21}}{1+c_{21}}k \quad . \qquad (20)$$

Since $0<l_1<l_2$ and $k>0$, if

$$-\frac{1}{1+c_{12}} > -\frac{c_{21}}{1+c_{21}} \quad , \qquad (21)$$

it follows that $s_1(t)$ is separated into $y_1(t)$ and $s_2(t)$ into $y_2(t)$. The permutation problem can, therefore, be solved by always picking up the solution that satisfies (21),.

5. SEPARATION EXPERIMENTS

The 2-microphone, 2-output system was evaluated using experiments that involved the 2-source and 3-source problems.

5.1. Separability measure

From (16), if $s_1(t)$ is mainly separated into $y_1(t)$ and $s_2(t)$ into $y_2(t)$, the separated signals are $b_{11}s_1(t)$ and $b_{22}s_2(t)$. In this case, $b_{12}s_2(t)$ and $b_{21}s_1(t)$ can be considered as noise for $b_{11}s_1(t)$ and $b_{22}s_2(t)$, respectively. Then, S/N ratios (dB) in $y_1(t)$ and $y_2(t)$ are given by

$$10\log_{10}\left|\frac{E\left[\left(b_{11}s_1(t)\right)^2\right]}{E\left[\left(b_{12}s_2(t)\right)^2\right]}\right| = 20\log_{10}\left|\frac{b_{11}}{b_{12}}\right| + 10\log_{10}\left|\frac{E\left[\left(s_1(t)\right)^2\right]}{E\left[\left(s_2(t)\right)^2\right]}\right|$$

$$10\log_{10}\left|\frac{E\left[\left(b_{22}s_2(t)\right)^2\right]}{E\left[\left(b_{21}s_1(t)\right)^2\right]}\right| = 20\log_{10}\left|\frac{b_{22}}{b_{21}}\right| + 10\log_{10}\left|\frac{E\left[\left(s_2(t)\right)^2\right]}{E\left[\left(s_1(t)\right)^2\right]}\right| .$$

We thus define separability S_{ep} for the 2-source problem as

$$S_{ep} \equiv 10 \times \left(\log_{10}\left|\frac{b_{11}}{b_{12}}\right| + \log_{10}\left|\frac{b_{22}}{b_{21}}\right|\right) \quad . \qquad (22)$$

If $s_1(t)$ is mainly separated into $y_2(t)$ and $s_2(t)$ into $y_1(t)$, S_{ep} is defined as the negative of the right-hand side of (22).

In the 3-source problem, the output signal is described as

$$\begin{aligned} y_1(t) &= b_{11}s_1(t) + b_{12}s_2(t) + b_{13}s_3(t) \\ y_2(t) &= b_{21}s_1(t) + b_{22}s_2(t) + b_{23}s_3(t) \end{aligned} \quad . \qquad (23)$$

If $s_1(t)$ is mainly separated into $y_1(t)$, the S/N ratio is given by

$$10\log_{10}\left|\frac{E\left[\left(b_{11}s_1(t)\right)^2\right]}{E\left[\left(b_{12}s_2(t)+b_{13}s_3(t)\right)^2\right]}\right| \quad .$$

Then, we define Sep_{ij}, the separability of $s_i(t)$ mainly separated into $y_j(t)$ as

$$Sep_{ij} = 20\log_{10}\left|\frac{b_{ij}}{\sum_{(k<j)\cup(k>j)} b_{ik}}\right| \quad . \qquad (24)$$

5.2. Outline of the experiments

The 2-source experiments were performed using the system shown in Fig.1, in which the measured signal was generated according to (1). We selected l_1=2.0, l_2=3.0 and k=0.5. Three algorithms were evaluated by average separability in 30 experiments, in which source signals were all combinations of two out of six speech signals, consisting of one set each of English male and female speech and 2 each of Japanese male and female speech. The length of all speech signals was 20 seconds. The sampling frequency and quantization accuracy were 44.1kHz and 16 bits, respectively.

The proposed algorithm was compared with the conventional direct algorithm [6] and iterative algorithm [1], which minimize the following objective function by the Simplex method:

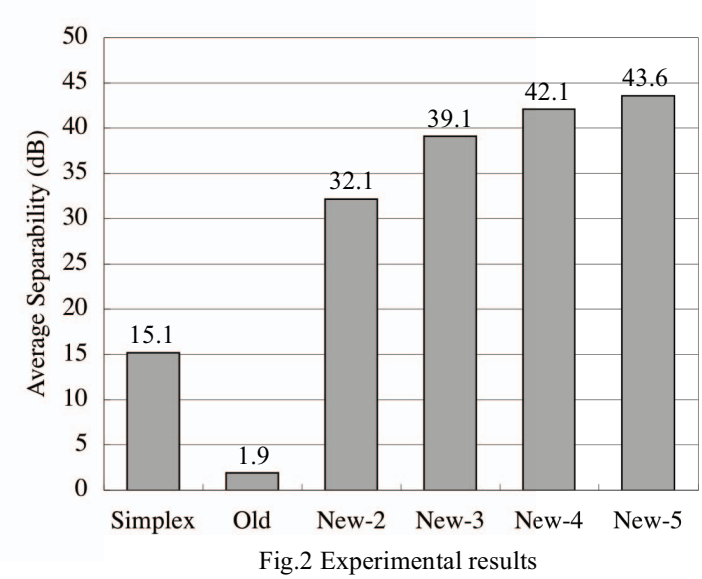

Fig.2 Experimental results
(2-source separation with 2-microphone system)

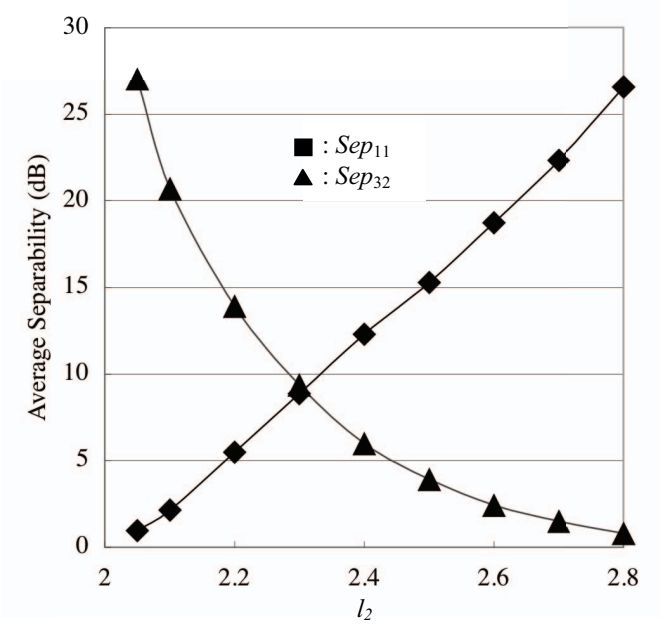

Fig.3 Experimental results
(3-source separation with 2-microphone system)

$$f(c_{12},c_{21}) = \left(c_{21}r_x^{(11)}(t) + c_{12}r_x^{(22)}(t) + (1+c_{12}c_{21})r_x^{(12)}(t)\right)^2 . \quad (25)$$

To estimate correlation functions, each speech signal was divided into time blocks. In this kind of case, the length of the block may influence the performance of the algorithm. We selected a block size of 8,192 samples (approximately 186ms) because this produced the best results in preliminary experiments using the iterative algorithm,

Fig.2 shows the experimental results where "*m*" of "New-*m*" corresponds to "*m*" in (13). The conventional algorithm ("Old" in Fig.2) did not produce good results mainly due to the permutation problem. Permutation control improved the performance of the new algorithm (for example, the 32.1dB of "New-2" rises to 32.9 dB). It was remarkably effective with the conventional method, where the separability was raised to the level of "New-2".

We also performed 3-source experiments with the new algorithm. The 120 experiments in total all consisted of combinations of three out of the above-mentioned 6 speech signals. In addition to the notations in 2-source experiments, we denote the distance between source $s_3(t)$ and the microphone "Mic-1" as l_3. The separation performance in this case strongly depended on the arrangement of sources. We therefore calculated the average values of Sep_{11} and Sep_{32} of (24), changing l_2 from 2.25 to 2.8, with l_1=2.0, l_3=3.0 and k=0.5. The result is shown in Fig.3, where either $s_1(t)$ or $s_3(t)$ was separated with separability of more than 9dB. Fig.3 shows that if two of three sources are adjacent, the other source can be separated with separability of more than 25dB when the null point generated by the method suppresses two sources simultaneously.

6. CONCLUSION

A new method for separating sound sources propagated in the same direction and a new direct algorithm for blind source separation were evaluated in 2-source and 3-source separation experiments. The algorithm can also be applied to frequency domain separation, which would enable application of the method to sound separation in the reverberant field. Such application and the solution to the problems of more than 3 microphone systems remain to be tackled.

7. REFERENCES

[1] M.Iwaki and A.Ando, "Selective Microphone System using Blind Separation of Block Decorrelation of Output Signal," *Proc. ICA2003*, P5A-09, pp. 1023-1028, (2003).

[2] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley, (2001)

[3] K. Matsuoka, M. Ohya, and M. Kawamoto, "A Neural Net for Blind Separation of Nonstationary Signals", *Neural Network*, Vol.8, no.3, pp.411-419, (1995)

[4] L. Parra, and C. Spence, "Convolutive Blind Separation of Non-Stationary Sources," *IEEE Trans. Sp. and Sig. Proc.*, Vol.8, no.3, pp.320-327, (2000)

[5] A. Ando, and K. Ono, "A Blind Separation Algorithm for Separation of Nonstationary Sources," *Tech. Rep. of IEICE*, EA2004-22, pp.31-36, (2004)

[6] Y. Takahashi, M. Toyama, and M. Iwaki, "Sound Source Separation by Decorrelation of 2-point Microphone Signal", *Proc. 17th Int. Cong. on Acoustics,* 3D.04.02, (2001)

[7] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of Blind Signal Separation Method Using Directivity Pattern under Reverberant Conditions", *Proc. ICASSP2000*, Vol.5, pp.3140-3143 (2000)