Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Blind source separation and sound source localization on time-frequency domain considering time lag information

Shogo Ueda, Fumio Sasaki, Osamu Tanaka and Masahito Yasuoka

Department of Architecture, Tokyo University of Science, 1-3 Kagurazaka, Shinjuku-ku,
162-8601 Tokyo, Japan
shogo_0604@yahoo.co.jp

The blind source separation and sound source localization based on independent component analysis on time-frequency domain considering time lag information between source signals and observation signals are conducted. The formulation based on the independency of time-frequency domain and the linearity of source signal is presented. The method which can be conducted not only the separation of source signals but also the specification of location of source signals is proposed through the consideration of time lags. Using this method, it can be analyzed even if observation signals include an intermittent noise, under the assumption of some independency of source signals. First of all, the number of source signals is specified through the quotient of complex valued time-frequency information of arbitrary two observation signals. Next, the locations of source signals are specified using the relationship of relative time lags between source signals and observation signals. Then, the source signals are obtained by use of the Fourier Transform. The numerical test is conducted to confirm our method, and then the locations of source signals and source signals are obtained by high accuracy.

1 Introduction

The cocktail party effect is known as auditory ability to distinguish particular sound and voice among other sounds and background noises. The cause of the cocktail party effect is tried solving from various fields and various points of view. The blind source separation problem corresponds to a way to enable computers to solve the cocktail party effect.

The methods for the specification of the number of source signals and the separation of the source signals are proposed using time-frequency information of source signals as a technique of blind source separation [1], [2]. In the paper [3], [4], the specification of the number of source signals and the separation of source signals are conducted by using wavelet analysis assuming some kind of independency for time-frequency information of source signals. However, in these papers, the time lags between source signals and observation signals are not considered.

In this paper, the method which can be conducted not only the separation of source signals but also the specification of location of source signals is proposed through the consideration of time lags. Moreover, the numerical test is conducted to confirm our method.

2 Formulation

2.1 Assumption of Independency of Time-Frequency Information

Let $\mathbf{s}(t)$ be a N dimension real valued vector function of source signal data $s_j(t)$ ($1 \le j \le N$)

$$\mathbf{s}(t) = (s_1(t), \cdots, s_j(t), \cdots, s_N(t))^T . \tag{1}$$

Let $\mathbf{x}(t)$ be a M dimension real valued vector function of observation data $x_k(t)$ ($1 \le k \le M$)

$$\mathbf{x}(t) = (x_1(t), \cdots, x_k(t), \cdots, x_M(t))^T . \tag{2}$$

Where we assume $M \ge N$.

Let $\mathbf{A}=(a_{kj})$ ($1 \le k \le M$, $1 \le j \le N$) be a damping matrix, and moreover c_{kj} represents time lag between k component (x_k) and j component (s_j). Where a_{kj} and c_{kj} are real values.

In this paper, the reflections of source signals aren't considered. The linearity is assumed between $x_k(t)$ and $s_j(t)$, on the free sound field such that

$$x_k(t) = \sum_{j=1}^{N} a_{kj} s_j(t - c_{kj}) \quad (1 \le k \le M) . \tag{3}$$

Here, $\mathbf{x}(t)$ is only known data, and $\mathbf{A}$, c_{kj}, $\mathbf{s}(t)$ are all unknown data.

Let $\mathbf{S}(t,\omega)$ be a time-frequency information vector of source signal vector $\mathbf{s}(t)$ and be a complex valued vector function. In this paper, continuous wavelet transform of which integral kernel consists complex wavelet is adopted to obtain $\mathbf{S}(t,\omega)$. Therefore, time-frequency information of $s_j(t - c_{kj})$ can be represented by the form $S_j(t - c_{kj}, \omega)$ which is only the time shift on time-frequency domain.

The independency of time-frequency information of $S_j(t - c_{kj}, \omega)$ is assumed as follows.

Let E_{kj} be the set of time-frequency domain as

$$E_{kj} = \{(t,\omega) \mid S_j(t - c_{kj}, \omega) \ne 0, \text{and } \forall i (\ne j) \text{ s.t. } S_j(t - c_{ki}, \omega) = 0\} . \tag{4}$$

Assumption 1 $E_{kj} \ne \{\phi\}$ and measurable

Measurable means that E_{kj} has some area on time-frequency domain. The assumption 1 is roughly explained as follows. The concept of the assumption 1 is illustrated in Fig.1. Where, we set $N = 4$ in this figure. This shows that the time-frequency information $X_k(t,\omega)$ of the observation signal $x_k(t)$ is represented a mixture of four source signals $a_{kj} S_j(t - c_{kj}, \omega)$ ($1 \le j \le 4$). Assumption 1 means that independent domains E_{kj} of $a_{kj} S_j(t - c_{kj}, \omega)$ which doesn't overlap each other exist on every k and j.

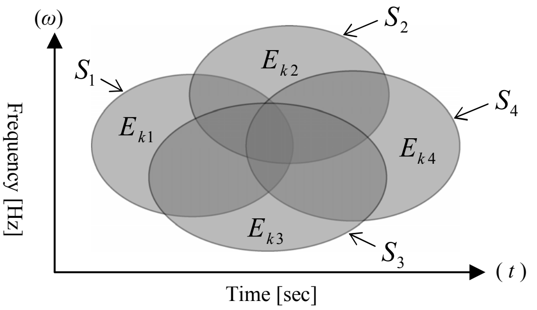

Fig.1 Illustration of independent domains E_{kj} ($1 \le j \le 4$) on $X_k(t,\omega)$.

Let $\widetilde{E}_{kj}$ be a time shift of E_{kj} to $-c_{kj}$. That is, $\widetilde{E}_{kj}$ is shifted E_{kj} by time direction $-c_{kj}$. And, Let G_j be

$$G_j = \bigcap_{k=1}^{M} \widetilde{E}_{kj} . \tag{5}$$

Assumption 2 $G_j \ne \{\phi\}$ and measurable

These assumptions make sense on general sound field.

2.2 Specification of the Number of Source Signals

In this section, the specification method of the number of source signals (*N*) is shown. For $\forall k,l \ (1 \le k,l \le N)$, time-frequency information $X_k\,(t,\omega)$ and $X_l\,(t,\omega)$ of observation signals $x_k\,(t)$ and $x_l\,(t)$ respectively can be represented

$$X_k(t,\omega) = \sum_{j=1}^{N} a_{kj} S_j(t - c_{kj}, \omega)\,. \tag{6}$$

$$X_l(t,\omega) = \sum_{j=1}^{N} a_{lj} S_j(t - c_{lj}, \omega)\,. \tag{7}$$

Then, complex valued quotient function $Q(t_k,t_l,\omega)$ is introduced and defined by

$$Q(t_k,t_l,\omega) = \frac{X_k(t,\omega)}{X_l(t,\omega)} = \frac{\sum_{j=1}^{N} a_{kj} S_j(t - c_{kj}, \omega)}{\sum_{j=1}^{N} a_{lj} S_j(t - c_{lj}, \omega)}\,. \tag{8}$$

By the assumption of independency (Assumption 1), if $(\tilde{t},\omega) \in E_{kj}$, then

$$X_k(\tilde{t},\omega) = a_{kj} S_j(\tilde{t} - c_{kj}, \omega) \tag{9}$$

and if $(\hat{t},\omega) \in E_{lj}$, then

$$X_l(\hat{t},\omega) = a_{lj} S_j(\hat{t} - c_{lj}, \omega)\,. \tag{10}$$

Moreover in this case, by the assumption of independency (Assumption 2), if $(\tilde{t},\omega), (\hat{t},\omega) \in G_j$, and $\hat{t} = \tilde{t} - (c_{kj} - c_{lj})$ are satisfied, then

$$Q(\tilde{t},\hat{t},\omega) = \frac{a_{kj}}{a_{lj}} \in \mathbf{R} \ \text{(Real value)}. \tag{11}$$

Therefore, if $\hat{t} = \tilde{t} - (c_{kj} - c_{lj})$ is satisfied at least in the domain G_j, the quotient function *Q* takes same constant real value a_{kj}/a_{lj}. Because *Q* is complex valued function in general, when $\tilde{t}$, $\hat{t}$ and ω vary all time-frequency domain, and if the region that takes same constant real value has measurable(some area), the number of same constant real values coincides the number *N* of source signals. The possibility that the quotient function *Q* takes other real values still remain, but in that case, the existence of the region which has measurable is very rare.

However, it is still uncertain which real values a_{kj}/a_{lj} correspond to the source signals. Here, we calculate the quotient function *Q* about each *k* component to make *l* component fix. If it can be possible to estimate independent domain E_{kj} on time-frequency domain, we can found the correspondence of them. First, we try memorizing *Q*(real value) and $(\tilde{t},\omega)$ when the quotient function takes real value a_{kj}/a_{lj}, and mark every real value a_{kj}/a_{lj} on time-frequency domain $X_k\,(t,\omega)$. Then, the marked areas of source signals, that is, independent domains E_{kj} are illustrated on time-frequency domain $X_k\,(t,\omega)$. If time lag c_{kj} is not so large, then the domains G_j overlapped with each E_{kj} exist. In these circumstances, the real values prove to corresponding with the domain G_j. Then the problem which real values correspond to source signals is solved.

2.3 Specification of the Location of Source Signals

When Eq.(11) holds, then

$$\hat{t} - \tilde{t} = c_{lj} - c_{kj}\,. \tag{12}$$

It means that the relative time lags between the distance $\overline{s_j x_k}$ and $\overline{s_j x_l}$ is $\hat{t} - \tilde{t}$. Where $\overline{s_j x_k}$ means the distance between the location of the source point s_j and observation point x_k. (In this paper, the source signals and the location of source points are represented the same symbol $s_j(t)$, s_j respectively.)

Therefore, when the source signal s_j is fixed, and propagation velocity represents v (≒330 m/sec in case of sound), and let $d_{jkl} = v(\hat{t} - \tilde{t})$, then d_{jkl} can be shown

$$|s_j - x_k| - |s_j - x_l| = d_{jkl}\,. \tag{13}$$

So, the relative distance d_{jkl} between $\overline{s_j x_k}$ and $\overline{s_j x_l}$ can be known. In the case of 2 dimensional space (3-dim), if three (four) relative distance can be known, the location of source signal is specified.

2.4 Specification of Source Signals

Let the constant matrix **B** be

$$\mathbf{B} = (b_{kj}) \ , \ b_{kj} = a_{kj} / a_{lj} \ \ (\ l \text{ is fixed}). \tag{14}$$

The matrix **B** which is composed of b_{kj} is specified at **2.2**, and so is c_{kj} at **2.3**.

Let $\tilde{s}_j(t)$ be

$$x_k(t) = \sum_{j=1}^{N} b_{kj} \tilde{s}_j(t - c_{kj})\,. \tag{15}$$

Here, we rewrite Eq.(15) without no confusion as follows.

$$\mathbf{x}(t) = \mathbf{B}\tilde{\mathbf{s}}(t - c_{kj})\,. \tag{16}$$

The difference between $\mathbf{s}(t)$ and $\tilde{\mathbf{s}}(t)$ is a multiple of constant. The representation of Fourier domain of Eq.(16) is

$$\hat{\mathbf{x}}(\omega) = \mathbf{B}\hat{\tilde{\mathbf{s}}}(\omega, c_{lj})\,. \tag{17}$$

Where the symbol ^ means Fourier domain.

The *k* –th component of the Eq.(17) is rewritten by

$$\hat{x}_k(\omega) = \sum_{j=1}^{N} b_{kj} e^{-i\omega c_{kj}} \hat{\tilde{s}}_j(\omega)\,. \tag{18}$$

The Eq.(18) is linear equation when ω is fixed, therefore if each ω is fixed, and solve the Eq.(18), then $\hat{\tilde{s}}_j(\omega)$ $(1 \le j \le M)$ can be calculated. Finally, $\tilde{s}_j(t)$ can be calculated by Inverse Fourier Transform (F^{-1}).

$$\tilde{s}_j(t) = F^{-1}[\hat{\tilde{s}}_j(\omega)]\,. \tag{19}$$

In this paper, $\tilde{\mathbf{s}}(t)$ can be calculated instead of $\mathbf{s}(t)$.

3 Numerical Test

3.1 Problem Setting

Numerical test can be conducted to confirm our method. In this example, 2 dimensional space is assumed. The locations of source points ($N = 5$) and observation points ($M = 5$) are illustrated in Fig.2. The propagation velocity sets 330 m/sec. The source signals are shown in Fig.3. The $s_1(t)$ and $s_2(t)$ are male voice in English. The $s_3(t)$ and $s_4(t)$ are female voice in English. The $s_5(t)$ is an traffic noise recorded at the avenue. Sampling frequency is 44100[Hz], total duration time is 11.88[sec], and number of total step is 524288(=2^{19}). Damping matrix $\mathbf{A}$ is constructed inversely proportional to the distance between s_j and x_k.

$$\mathbf{A} = \begin{pmatrix} 0.0720 & 0.0909 & 0.0614 & 0.1111 & 0.1414 \\ 0.2000 & 0.1240 & 0.1562 & 0.1394 & 0.3536 \\ 0.2425 & 0.2774 & 0.1085 & 0.0698 & 0.1213 \\ 0.1857 & 0.0808 & 0.3333 & 0.0698 & 0.0877 \\ 0.1000 & 0.0677 & 0.1414 & 0.1768 & 0.1240 \end{pmatrix}. \quad (20)$$

$$\mathbf{B} = \begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 2.779 & 1.364 & 2.542 & 1.236 & 2.500 \\ 3.369 & 3.051 & 1.766 & 0.629 & 0.858 \\ 2.580 & 0.889 & 5.426 & 0.629 & 0.620 \\ 1.389 & 0.745 & 2.302 & 1.591 & 0.877 \end{pmatrix}. \quad (21)$$

The relative time lag matrix ($l = 1$) is

$$\mathbf{C}_{\mathrm{R}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1189 & 393 & 1319 & 230 & 567 \\ 1306 & 988 & 943 & -710 & -157 \\ 1137 & -183 & 1774 & -710 & -579 \\ 521 & -503 & 1230 & 447 & -132 \end{pmatrix}. \quad (22)$$

Where we write the entries of $\mathbf{C}_{\mathrm{R}}$ as lag step $(c_{lj} - c_{kj}) \times 44100$ instead of actual $(c_{lj} - c_{kj})$. The synthesized observation signals $x_k(t)$ are shown in Fig.4.

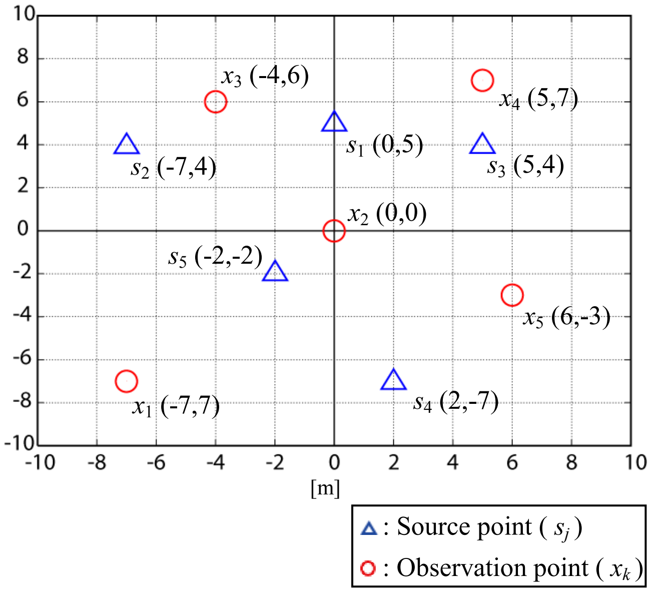

Fig.2 The locations of source points ($N = 5$) and observation points ($M = 5$).

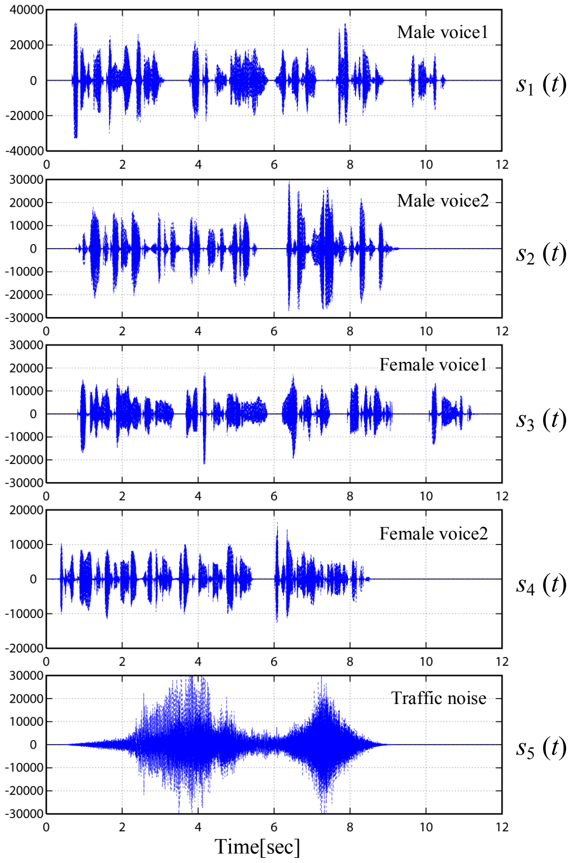

Fig.3 Source signals ($N = 5$).

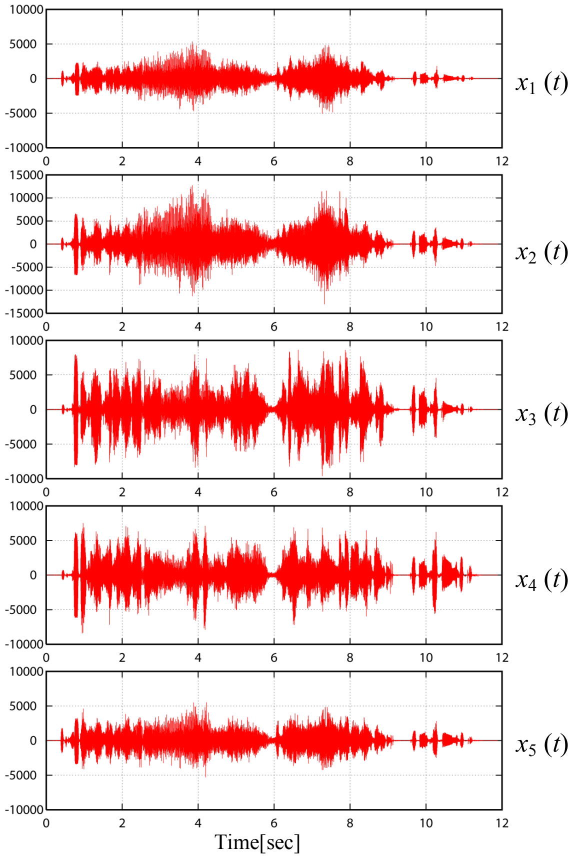

Fig.4 Observation signals ($M = 5$).

Complex mother wavelet which consists of Mayer wavelet (as real part) and Hilbert transform of Mayer wavelet (as imaginary part) is adopted to obtain time-frequency information of observation signals.

3.2 Specification of the Number of Source Signals

From now, observation signals $x_k(t)$ and their locations are only known data. Fig.5 is a histogram of $X_k(t,\omega)/X_l(t,\omega)$ ($l=1$) that the quotient function takes real value. Five peaks can be seen in each histogram. Therefore, the number of source points can be estimated as 5(=*N*). The computed relative damping (b_{kj}) and relative time lag ($c_{lj} - c_{kj}$) correspond to the coordinate of Quotient and Time-shift of each peaks' value in Fig.5. The computed relative damping matrix $\tilde{\mathbf{B}}$ and relative time lag matrix $\tilde{\mathbf{C}}_\mathrm{R}$ is

$$\tilde{\mathbf{B}} = \begin{pmatrix} 1.000 & 1.000 & 1.000 & 1.000 & 1.000 \\ 2.779 & 1.364 & 2.540 & 1.236 & 2.500 \\ 3.370 & 3.049 & 1.765 & 0.628 & 0.857 \\ 2.579 & 0.889 & 5.423 & 0.628 & 0.620 \\ 1.389 & 0.745 & 2.301 & 1.591 & 0.877 \end{pmatrix}. \tag{23}$$

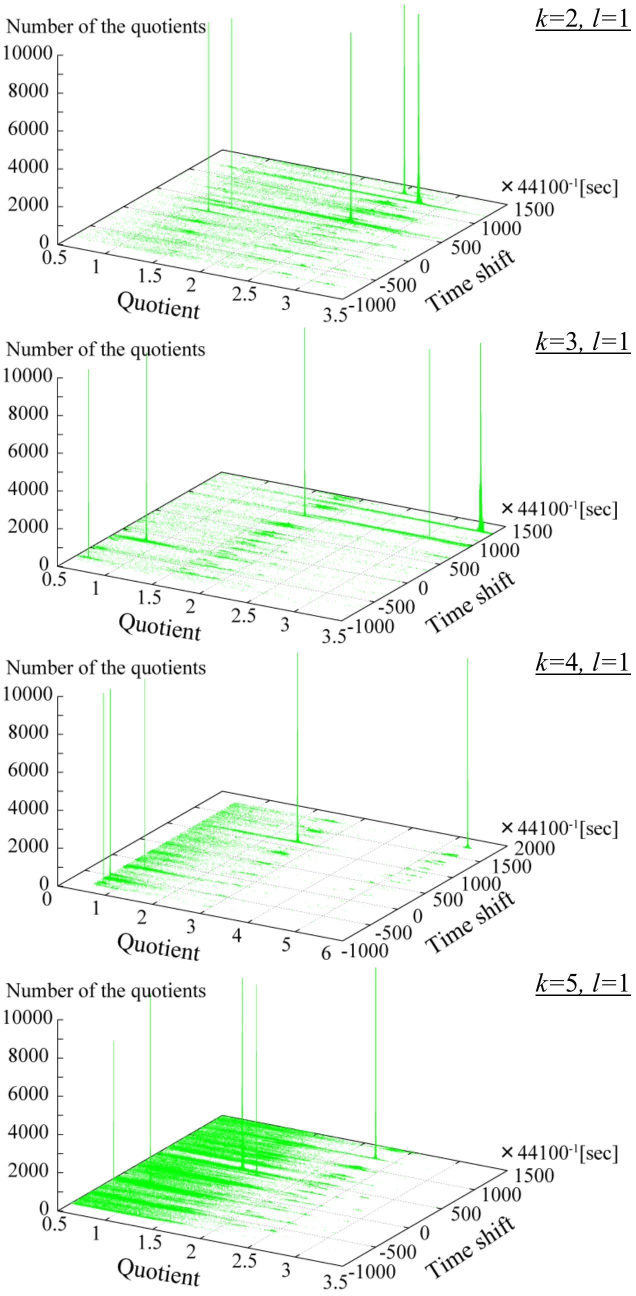

Fig.5 Histgram of $X_k(t,\omega)$ / $X_l(t,\omega)$ (Real value).

$$\tilde{\mathbf{C}}_\mathrm{R} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1189 & 393 & 1319 & 230 & 567 \\ 1306 & 988 & 943 & -710 & -157 \\ 1137 & -183 & 1774 & -710 & -579 \\ 521 & -503 & 1230 & 447 & -132 \end{pmatrix}. \tag{24}$$

Compared with the components of matrix **B** (Eq.(21)), and $\tilde{\mathbf{B}}$ (Eq.(23)), there is an error of less than 0.06 %. Compared with the components of matrix $\mathbf{C}_\mathrm{R}$(Eq.(22)) and $\tilde{\mathbf{C}}_\mathrm{R}$ (Eq.(24)), $\tilde{\mathbf{C}}_\mathrm{R}$ is completely coincident with $\mathbf{C}_\mathrm{R}$.

3.3 Specification of the Location of Source Signals

The relative distance d_{jkl} between $\overline{s_j x_k}$ and $\overline{s_j x_l}$ can be known after calculating relative time lag at section 3.2. Now, there are five observation points so that four relative distance d_{jkl} can be found by each source point. Fig.6 is hyperbolic curves from five source points described in d_{jkl}. In this case, a source point must be on four curves to reduce an error in calculation, though in the case of 2 dimensional space, it is enough to be on at least three curves. Therefore, each location of five source points is specified to compute the point of intersection. The result of identified source points is shown in Fig.7. It is very good accuracy. There is an error of less than 10 mm (0.07% of the max range of the x-y coordinate).

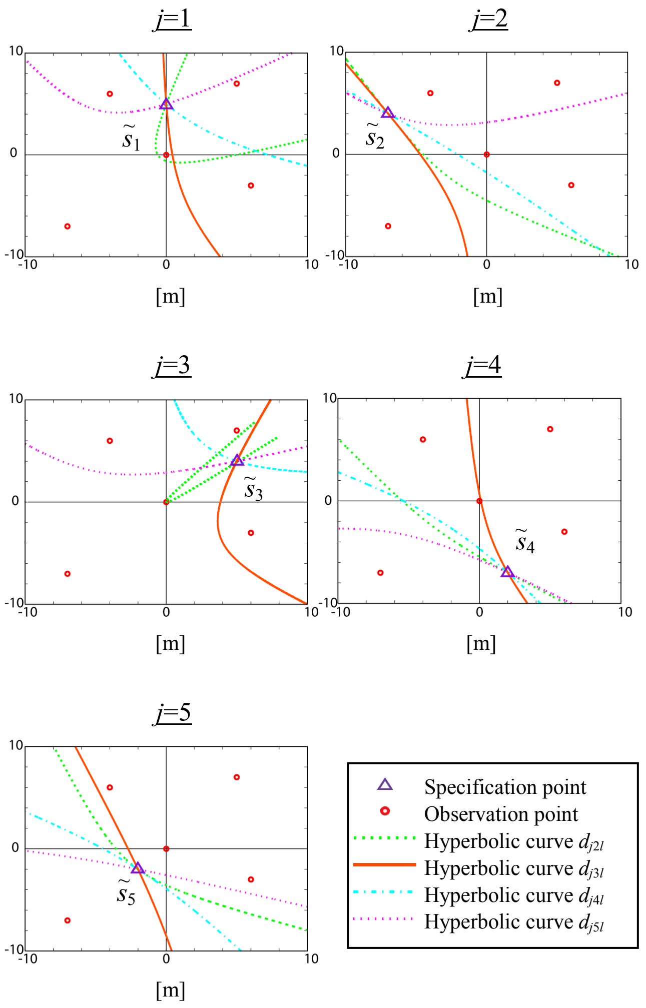

Fig.6 Hyperbolic curves to specify source points.

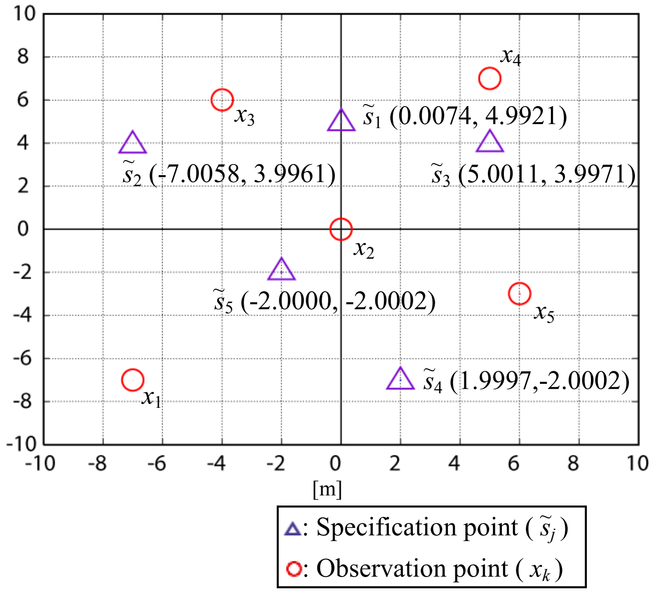

Fig.7 Specification of source points.

3.4 Specification of Source Signals

The relative damping matrix $\tilde{\mathbf{B}}$ has been computed at section 3.2, and the time lag matrix $\tilde{\mathbf{C}}$ ($\tilde{c}_{kj}$) has been computed to specify the location of source points at section 3.3. Namely, each of unknown variable except of $\hat{\tilde{s}}_j(\omega)$ at Eq.(18) has turned out to be known. So $\hat{\tilde{s}}_j(\omega)$ can be calculated to solve a linear Eq.(18) about every frequency. Therefore $\tilde{s}_j(t)$ can be estimated by Inverse Fourier Transform of $\hat{\tilde{s}}_j(\omega)$ in Eq.(19).

However, the calculated source signals $\tilde{s}_j(t)$ don't coincide with actual source $s_j(t)$ concerning with the size of amplitude as they are. If we assume that the damping coefficient is in inverse proportion to the propagation distance, the damping (a_{lj}) ($l=1$) can be calculated as

$$(a_{lj}) = (0.0720 \quad 0.0909 \quad 0.0614 \quad 0.1111 \quad 0.1414)\,. \qquad (25)$$

Through the multiplication $\tilde{s}_j(t)$ by them, we finally get $\tilde{s}_j(t)$ which is coinciding with $s_j(t)$ about the size too. Fig.8 shows calculated signals $\tilde{s}_j(t)$. Compared with actual source signals $s_j(t)$ (Fig.3), there are very good coincidence and errors of less than 0.06%.

4 Conclusion

The method for the separation of observation signals and the specification of location of source signals was formulated through the consideration of time lag. Using this method, the separation and the specification of location can be conducted, under the assumption of some independency of source signals. It is confirmed through the numerical test that the specification of location has done almost completely and the separation has done with high accuracy.

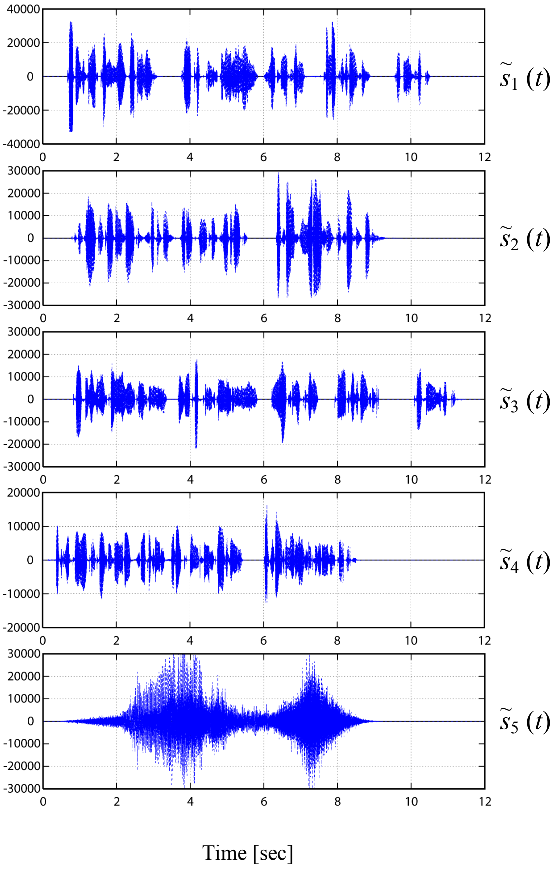

Fig.8 Calculated source signals.

References

[1] R. Balan and J. Rosca, Statistical properties of STFT ratios for two channel systems and applications to blind source separation, Proceedings ICA2000, 19-22, Helsinki, Finland (2000)

[2] D. Napoletani, C.A. Berenstein and P.S. Krishnaprasad, Quotient signal decomposition and order estimation, TECHNICAL RESEARCH REPORT of University of Maryland, http://www.isr.umd.edu (2002)

[3] Keiko Fujita, Yoshitsugu Takei, Akira Morimoto, Ryuichi Ashino and Mitsuo Morimoto, Blind source separation on a time-frequency space -mathematical background-, IEICE Technical Report EA2005-12, 37-42 (2005) (in Japanese with English abstract)

[4] Akira Morimoto, Keiko Fujita and Ryuichi Ashino, Blind source separation on a time-frequency space, IEICE Technical Report EA2005-11, 31-36 (2005) (in Japanese with English abstract)