The American Society of Mechanical Engineers

Reprinted From
PVP — Vol. 177, Quality Use of the Computer: Computational Mechanics, Artifical Intelligence, Robotics and Acoustic Sensing
Editors: J. F Cory, JR., J. L. Gordon, T. Miyoshi, and K. Suzuki
Book No. H00492 — 1989

A NEW METHOD TO LOCATE MULTIPLE SOUND SOURCES IN ONE-DIMENSIONAL SPACE BASED ON ESTIMATION OF EACH SOURCE SOUND

H. Kanai and K. Kido
Research Center for Applied Information Sciences
Tohoku University
Sendai, Japan

ABSTRACT

Several methods have been used to locate multiple sound sources on a line such as a tube or a pipe by using many sensor outputs. It is difficult, however, to locate sound sources accurately when the source sounds are correlated with each other and/or when the signal-to-noise ratio is low. By using the least mean square criterion, it is necessary to solve the under-determined normal equation to estimate the sound power and the waveform radiating from each assumed sound source. Therefore, we propose a new method to solve the equation by applying a tapering window to the singular-value-decomposition technique. These principles are confirmed experimentally. The proposed method is applicable to the diagnosis of piping by estimating the positions of holes from which streams of liquid or a gas are leaking.

1. INTRODUCTION

We propose a new method to locate multiple sound sources in a one-dimensional space such as in a tube or a pipe based on the estimation of the waveform radiated from each sound source by using a multi-channel observed signal detected by multiple sensors on the same line.

In the literature (Silvia, 1987 and Youn and Ahmed, 1984), several correlation-based methods have been proposed to locate a sound source using several sensor outputs. In these methods, the position of a sound source is determined from the time differences estimated by using the cross-spectrum or the correlation between each pair of sensor outputs. It is difficult, however, to locate multiple sound sources by such conventional methods when there are several sound sources correlated with each other, because these methods are not based on the direct estimation of the waveform radiating from each sound source.

Therefore, another method has been proposed (Silvia, 1984). In this proposed method, the position of a sound source is first assumed on the line, and the transfer function, $H_{ij}(\omega)$, from the i-th assumed sound source to the j-th sensor is calculated using the distance between them. After compensating for the transfer characteristics by multiplying the spectrum of the j-th sensor output $Y_j(\omega)$ by the inverse characteristics, $H_{ij}(\omega)^{-1}$ of $H_{ij}(\omega)$, the sound of the i-th assumed source is estimated. Then, the power of the sum, $\Sigma_j H_{ij}(\omega)^{-1}\cdot Y_j(\omega)$, is calculated for each assumed point i by using the cross-spectrum technique. The positions of the sound sources are determined from the peaks in the resultant power of each assumed point. However, if there are multiple sound sources, the estimated sound, $H_{ij}(\omega)^{-1}\cdot Y_j(\omega)$, of the i-th assumed point is affected by the signals radiating from the sound sources other than the i-th source. Moreover, if the sensor outputs are contaminated by noise, the power of the noise term is not negligible.

These difficulties result because the estimation of source signals with the previously proposed methods are not based on the least mean square (LMS) criterion. By using the LMS technique, if the position of each sound source is already known, each of the source sounds is estimated from the ordinary over-determined normal equation derived by minimizing the additional noise power (Yanagida et al., 1983). However, with regard to the problem of locating the multiple sound sources, since the positions of the sound sources are not known, it is necessary for the number of assumed sound sources on the line to be greater than the number of the sensors. Thus, the resultant normal equation derived by minimizing the noise power is under-determined, and the sound spectrum of each assumed sound source cannot be estimated by using ordinary methods. This is the difficulty encountered in estimating the locations of the signals of the multiple sound sources in low signal-to-noise ratio (SNR) cases.

Therefore, we propose a new method to locate the multiple sound sources as follows: We obtain the minimum norm solution to the under-determined normal equation by using the singular-value-decomposition (SVD) technique. However, it is still difficult to accurately determine the sound source locations because the positions of the sound sources become obscure due to the application of low order truncation in solving the under-determined normal equation. Thus, we use the new method to apply a tapering window in truncation of the order employed in the previously proposed SVD technique (Kanai and Kido, 1988). These principles are confirmed experimentally. The principles of the proposed method are applicable in many fields.

2. PRINCIPLE

It is assumed that there are N_0 actual sound sources and L sensors in a one-dimensional space as shown in Fig. 1 and that the sound radiated from each source propagates to the sensors through the line. To determine the positions of the

M_e actual sound sources, it is necessary to estimate the sound signals, $x_i(n)$, $i=1,2,\ldots,M$, radiated from at the M assumed sound sources on the line ($M>M_e$). If there is actually a sound source at the i-th assumed point, the corresponding estimated signal, $x_i(n)$, coincides with the original sound radiated from the actual source and the estimated signal has large power. Let M dimensional vector, $x(n)$, denote an M-channel N-length input signal radiated from the assumed sound sources at sample index n,

$$x(n) = [\ x_1(n),\ x_2(n),\ \ldots\ ,\ x_M(n)\ ]^T, \qquad (0 \le n < N)$$

where the superscript T indicates transposition. If L dimensional vector, $y(n)$, represents the L-channel output signal received by the L sensors, then the linear shift-invariant system convolution that relates $x(n)$ to $y(n)$ is given by

$$y(n) = \begin{pmatrix} \sum_{i=1}^{M} \sum_{k=0}^{N-1} x_i(n-k)h_{i1}(k) \\ \sum_{i=1}^{M} \sum_{k=0}^{N-1} x_i(n-k)h_{i2}(k) \\ \cdots\cdots \\ \sum_{i=1}^{M} \sum_{k=0}^{N-1} x_i(n-k)h_{iL}(k) \end{pmatrix} + n(n)$$

$$\triangleq h(n) * x(n) + n(n), \tag{1}$$

where the M-input L-output L-by-M multichannel impulse response matrix is given by

$$h(n) = \begin{pmatrix} h_{11}(n) & h_{12}(n) & \ldots & h_{1M}(n) \\ h_{21}(n) & h_{22}(n) & \ldots & h_{2M}(n) \\ \cdot & \cdot & \cdots & \cdot \\ h_{L1}(n) & h_{L2}(n) & \ldots & h_{LM}(n) \end{pmatrix}.$$

The impulse response of the transfer system from i-th sound source to the j-th sensor is denoted by $h_{ij}(n)$, and the L-dimensional vector $n(n)$ denotes an L-channel mutually uncorrelated noise included in $y(n)$,

$$n(n) = [\ n_1(n),\ n_2(n),\ \ldots\ ,\ n_L(n)\ ]^T. \qquad (0 \le n < N)$$

The L-by-M multi-channel discrete Fourier transform matrix, $H(k)$, is defined as

$$H(k) = \begin{pmatrix} H_{11}(k) & H_{12}(k) & \ldots & H_{1M}(k) \\ H_{21}(k) & H_{22}(k) & \ldots & H_{2M}(k) \\ \cdot & \cdot & \cdots & \cdot \\ H_{L1}(k) & H_{L2}(k) & \ldots & H_{LM}(k) \end{pmatrix}, \qquad (k=0,1,\ldots,N-1)$$

where $H_{ij}(k)$ is the N-point discrete Fourier transform of $h_{ij}(n)$:

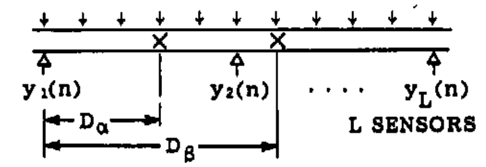

Fig. 1 The relation between the M_e actual sound sources, the L sensors, and the M assumed sound sources in the one-dimensional space. They are represented by 'X', 'Δ', and '↓', respectively. In the figure, D_k, (k=α and β), denotes the distance between the k-th actual sound source and the left edge of the pipe.

$$H_{ij}(k) = \sum_{n=0}^{N-1} h_{ij}(n)\ \exp(-j2\pi kn/N). \tag{2}$$

Let $X_i(k)$, $N_j(k)$, and $Y_j(k)$ denote the Fourier transforms of the signals, $x_i(n)$, $n_j(n)$, and $y_j(n)$, respectively, and let $X(k)$, $N(k)$, and $Y(k)$ denote the vectors of these spectra as follows:

$$X(k) = [\ X_1(k),\ X_2(k),\ \ldots\ ,\ X_M(k)\ ]^T,$$

$$N(k) = [\ N_1(k),\ N_2(k),\ \ldots\ ,\ N_L(k)\ ]^T,$$

and

$$Y(k) = [\ Y_1(k),\ Y_2(k),\ \ldots\ ,\ Y_L(k)\ ]^T.$$

Thus, the multi-channel convolution in Eq. (1) is described in the frequency domain as

$$Y(k) = H(k)\cdot X(k) + N(k), \qquad (k=0,1,2,\ldots,N-1) \tag{3}$$

In this paper, suppose the characteristic of the transfer system of the each element, $H_{ij}(k)$, of $H(k)$ to be approximately determined by the distance, r_{ij}, between the assumed i-th sound source and the j-th sensor as follows:

$$H_{ij}(k) = \exp(-\alpha r_{ij})\ \exp(-j2\pi r_{ij}\cdot f_s k/cN), \tag{4}$$

where α, f_s, and c denote an attenuation constant, the sampling frequency, and the sound velocity, respectively.

To locate M_e sound sources from the L-channel observed signal, $y(n)$, or the spectrum vector, $Y(k)$, it is necessary to determine the M_e positions radiating large sound power in the M assumed sound sources ($M_e<M$). After describing the ordinary approaches, we propose a new method in the following section.

2.1 LOCATION OF SOUND SOURCES USING CROSS-SPECTRUM BASED METHOD.

The spectrum, $X_i(k)$, of the i-th assumed source is estimated by multiplying the spectrum, $Y_j(k)$, of the j-th sensor output by the inverse characteristic, $H_{ij}(k)^{-1}$, of the transfer function, $H_{ij}(k)$, from the i-th assumed sound source to the j-th sensor and summing up the L compensated spectra, { $H_{ij}(k)^{-1}\cdot Y_j(k)$ }, (j=1,2,...,L) as follows:

$$\hat{X}(k) = H(k)^{-}\cdot Y(k), \tag{5}$$

where $H(k)^{-}$ denotes the M-by-L inverse characteristic transfer matrix:

$$H(k)^{-} = \begin{pmatrix} H_{11}(k)^{-1} & H_{12}(k)^{-1} & \ldots & H_{1L}(k)^{-1} \\ H_{21}(k)^{-1} & H_{22}(k)^{-1} & \ldots & H_{2L}(k)^{-1} \\ \cdot & \cdot & \cdots & \cdot \\ H_{M1}(k)^{-1} & H_{M2}(k)^{-1} & \ldots & H_{ML}(k)^{-1} \end{pmatrix}.$$

The (j,i)-element, $H_{ij}(k)^{-1}$, of $H(k)^{-}$ is the Fourier transform of the inverse characteristics, $h_{ji}(n)^{-1}$, of the impulse response, $h_{ij}(n)$, of the transfer system $H_{ij}(k)$ from the i-th assumed sound source to that of the j-th sensor:

$$h_{ji}(n)^{-1} * h_{ij}(n) = \delta(n), \qquad (1\le i\le M,\ 1\le j\le L)$$

and

$$\delta(n) = \begin{cases} 1 & n=0 \\ 0 & n\neq 0. \end{cases}$$

If there is an actual source at the i-th assumed sound point, the estimated source sound spectrum, $\hat{X}_i(k)$, i.e., the i-th element of $\hat{X}(k)$, has large power. Thus, the method reported by Shima et al. (1988) identified the location of the actual sound source. The method obtains the approximate power, $W_i(k)$, radiated from i-th assumed sound source by calculating the cross-spectrum between $H_{ji}(k)^{-1}\cdot Y_j(k)$ and $H_{j'i}(k)^{-1}\cdot Y_{j'}(k)$ and then by averaging them for every pair of sensor points j and j', $(1 \leq j, j' \leq L,\ j \neq j')$ as:

$$W_i(k) = \sum_{j=1}^{L} \sum_{\substack{j'=1 \\ (j \neq j')}}^{L} [H_{ji}(k)^{-1}\cdot Y_j(k)]^{*} \cdot [H_{j'i}(k)^{-1}\cdot Y_{j'}(k)]$$

$$= |\underline{X}_i(k)|^2 - \sum_{j=1}^{L} |H_{ji}(k)^{-1}\cdot Y_j(k)|^2, \qquad (k=0,1,2,\ldots,N-1) \tag{6}$$

where * denotes the complex conjugate. By summing up the spectrum, $W_i(k)$, estimated for each frequency, k, the power, P_i, of the sound radiated from the i-th assumed sound source is obtained as follows:

$$P_i = \sum_{k} W_i(k), \qquad (i=1,2,\ldots,M) \tag{7}$$

The position of the sound source is determined from the power, $\{P_i\}$, of each assumed source sound. However, by substituting $Y(k)$ of Eq. (3), the estimate, $\underline{X}(k)$, of Eq. (5) is described as

$$\underline{X}(k) = H(k)^{-}H(k)X(k) + H(k)^{-}N(k). \tag{5'}$$

The first term, $H(k)^{-}H(k)$, is ordinarily not equal to the unit matrix, and thus the estimate, $\underline{X}(k)$, does not coincide with the original vector, $X(k)$, even in a noiseless case. Therefore, the estimated signal, $\underline{x}_i(n)$, of the i-th assumed sound source is affected by the convolution between the inverse transfer characteristic and the source sound signals other than the i-th sound source. Moreover, in low SNR cases, the second term corresponding to $\Sigma_j\ h_{ji}(n)^{-1} * n_j(n)$ due to noise cannot be negligible. Therefore, the method cannot be used to locate sound sources when there are multiple sources and/or the SNR is low.

In the literature, correlation-based methods have been proposed to estimate the time delay (Silvia, 1987 and Youn et al., 1984). However, if there is a correlation between the source sounds, $\{x_i(n)\}$, the position of each sound source cannot be reliably estimated by these methods. Moreover, these methods cannot estimate the waveform or the spectrum of each assumed sound source.

2.2 LOCATION OF ACTUAL SOUND SOURCES USING SVD OF H(k)

By minimizing the power of noise term, $N(k)$, of Eq. (3), the least mean square solution of $\underline{X}(k)$ of $X(k)$ is obtained as follows:

$$\underline{X}(k) = [H(k)^{H}H(k)]^{-1}H(k)^{H}Y(k), \tag{8}$$

where the superscript H denotes the conjugate transposition. In general least mean square problems, the dimension L of the observed vector, $Y(k)$, is ordinarily greater than the dimension M of estimated parameter vector, $X(k)$, that is, the normal equation is over-determined (Cadzow, 1982 and Oppenheim et al., 1975). However, in the determination of the positions of the sound sources, the number of the assumed sound sources, M, is greater than the number of the sensors, L. Thus, the rank of the M-by-M matrix $H(k)^{H}H(k)$ of Eq. (5) is equal to or less than L (L<M) and the Moore-Penrose generalized inverse matrix, $[H(k)^{H}H(k)]^{-1}H(k)^{H}$, cannot be calculated directly. It is difficult to estimate the spectrum, $\underline{X}(k)$, of each assumed sound source in such an under-determined case.

Therefore, we propose a reliable method to estimate, $\underline{X}(k)$, by using the Singular-value-decomposition (SVD) technique (Lancaster et al., 1985) as follows: The L-by-M matrix, $H(k)$, of the k-th frequency component is broken down as follows:

$$H(k) = Q_1(k)\ \Sigma(k)\ Q_2(k)^{H}, \tag{9}$$

where $Q_1(k)$ and $Q_2(k)$ denote L-by-L and M-by-M orthonormal matrices, which are composed of the orthonormal eigenvectors of $H(k)^{H}H(k)$ and $H(k)H(k)^{H}$, respectively, and $\Sigma(k)$ denotes the L-by-M diagonal matrix described as follows:

$$\Sigma(k) = \begin{pmatrix} \sigma_1(k) & 0 & \cdots & 0 & 0 \cdots 0 \\ 0 & \sigma_2(k) & 0 \cdots & 0 & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \cdots \vdots \\ 0 & & \cdots 0 & \sigma_L(k) & 0 \cdots 0 \end{pmatrix}$$

$$= \mathrm{diag}[\ \sigma_1(k),\ \sigma_2(k),\ \ldots,\ \sigma_L(k)\ ],$$

where $\{\ \sigma_i(k)\ \}$ denotes the singular values of $H(k)$. Since the ratio of the minimum to the maximum singular value is low, the rank of $H(k)$ is approximated by the order R, where R is smaller than L. Thus, the L-by-M diagonal matrix obtained by truncating the small singular values, denoted by $\Sigma_R(k)$, is described as:

$$\Sigma_R(k) = \mathrm{diag}\ [\ \sigma_1(k),\ \sigma_2(k),\ \ldots,\ \sigma_R(k),\ 0,\ \ldots,\ 0\ ].$$

When the M-by-L inverse matrix, $\Sigma_R(k)^{-1}$, of $\Sigma_R(k)$ is defined as

$$\Sigma_R(k)^{-1} = \begin{pmatrix} 1/\sigma_1(k) & 0 & \cdots 0 & \cdots 0 \\ 0 & 1/\sigma_2(k) & 0 \cdots \vdots & \cdots 0 \\ \vdots & \vdots & \cdots\ 0 & \cdots 0 \\ 0 & & \cdots 0\ 1/\sigma_R(k) & 0 \cdots 0 \\ 0 & & \cdots\ 0 & \cdots 0 \\ \vdots & & \cdots\ \vdots & \cdots \vdots \\ 0 & & \cdots\ 0 & \cdots 0 \end{pmatrix}$$

$$= \mathrm{diag}[\ 1/\sigma_1(k),\ 1/\sigma_2(k),\ldots,\ 1/\sigma_R(k),\ 0,\ldots,0\ ].$$

the estimate, $\underline{X}(k)$, of Eq. (8) is obtained by

$$\underline{X}(k) = Q_2(k)\ \Sigma_R(k)^{-1}\ Q_1(k)^{H}\ Y(k). \tag{10}$$

The physical meaning of the estimate, $\underline{X}(k)$, is considered as follows: By substituting $Y(k)$ of Eq. (3) and $H(k)$ of Eq. (9), the estimate, $\underline{X}(k)$, of Eq. (10) is described as

$$\underline{X}(k) = Q_2(k)\Delta_R(k)Q_2(k)^{H}X(k) + Q_2(k)\Sigma_R(k)^{-1}Q_1(k)^{H}N(k), \tag{10'}$$

where M-by-M matrix, $\Delta_R(k) = \Sigma_R(k)^{-1}\Sigma(k)$, is defined as follows:

$$\Delta_R(k) = \mathrm{diag}[\ \delta_1(k),\ \delta_2(k),\ldots,\ \delta_M(k)\ ],$$

and

$$\delta_i(k) = \begin{cases} 1 & \text{for } 1 \leq i \leq R \\ 0 & \text{for } R < i \leq M. \end{cases}$$

The first term, $[Q_2(k)^{H}]^{-1}\Delta_R(k)Q_2(k)^{H}$ of Eq. (10') denotes the "generalized" low-order-passing filter (Kanai and Kido, 1988). The second term, $Q_2(k)\Sigma_R(k)^{-1}Q_1(k)^{H}$, denotes the "generalized" low-order-suppression filter. Thus, by estimating $\underline{X}(k)$ based on Eq. (10), the noise term is suppressed even in low SNR cases. For example, the truncation order, R, is determined by comparing the following normalized ratio, $\rho(R)$, with the threshold, T_0:

$$\rho(R) = \left(\sum_{i=1}^{R} |\delta_i|^2 \Big/ \sum_{i=1}^{L} |\delta_i|^2 \right)^{1/2}, \tag{11}$$

where T_0 will depend on the particular application (Cadzow, 1982).

However, at the same time, the sharp truncation due to the generalized low-order passing filter produces a ripple around each actual sound source position in a manner similar to the well-known Gibbs phenomenon (Oppenheim et al., 1975) caused by ordinary low-pass filtering. Therefore, we have proposed an approach using a tapering window in the above SVD technique to suppress the ripple and clarify the each actual sound source position outlined below (Kanai and Kido, 1988).

The sequence, $\{\lambda_i(k)\}$, is calculated from the convolution between the truncated diagonal element, $\{\delta_i(k)\}$, of the matrix, $\Delta_R(k)$, and a tapering window function, $\{w_i\}$. The M-by-M diagonal matrix, $\Lambda(k)$, is defined by the resultant term, $\lambda_i(k)$, in the (i,i)-position ($1 \leq i \leq M$) and by zeros elsewhere. The source sound, $X(k)$, of the k-th frequency is estimated using the matrix, $\Lambda(k)\cdot\Sigma^{-1}(k)$, instead of $\Sigma_R(k)^{-1} = \Delta_R(k)\cdot\Sigma(k)^{-1}$ of Eq. (10) as

$$\hat{X}(k) = Q_2(k)\ \Lambda(k)\ \Sigma(k)^{-1}\ Q_1(k)^H\ Y(k), \qquad (12)$$

where

$$\Lambda(k) = \mathrm{diag}[\ \lambda_1(k),\ \lambda_2(k),\ \ldots,\ \lambda_M(k)\ ].$$

For example, the window function is defined by the following Hamming window (Oppenheim et al.,1975):

$$w_i = 0.54 + 0.46\cos(2\pi i/2N_w), \quad (-N_w \leq i \leq N_w) \qquad (13)$$

where the width of the window is equal to $2N_w+1$ points.

The spectrum, $\hat{X}(k)$, of M assumed sound sources is estimated for each frequency, k, by using Eq. (10) or Eq. (12). By summing up the power of the spectrum estimates, $\{ \hat{X}(k) \}$, obtained for each frequency, $k=0,\ 1,\ \ldots,\ N-1$, the sound power, P_i, radiated from the i-th assumed position, $1 \leq i \leq M$, is calculated as follows:

$$P_i = \sum_k |\hat{X}_i(k)|^2. \qquad (14)$$

Since the additive noise term is suppressed by the order truncation and the spectrum of each assumed source sound is estimated using the generalized inverse characteristic of the multi-channel transfer function and the above tapering window function, the positions and the signals of the sound sources are accurately estimated by the proposed method.

3. SIMULATION RESULTS

In order to illustrate the characteristics of the cross-spectrum based method and the proposed method, we have chosen the following three examples. These methods were implemented on an IBM3081-KX6 (1 word=32bit) computer using double precision arithmetic. In the following simulation experiments, the sound velocity, c, is 400 m/s, the attenuation constant α of Eq. (4) is 0.01 1/m (Shima et al., 1988), and the sampling frequency, f_s, is 2.4 kHz. The total length, N, of each of the L signals received at the sensors is 64 points. The length of a pipe in a one-dimensional space as shown in Fig. 1 is equal to 10m.

<u>First Example</u>: In order to compare the inverse characteristic of $H(k)^-$ of Eq. (5) and that of $Q_2(k)\Sigma_R(k)^{-1}Q_1(k)^H$ of Eq. (10), the following simple impulse sequences are chosen as the two source sound signals:

$$x_\alpha(n) = \delta(n-10)$$

and

$$x_\beta(n) = -\delta(n-15). \quad (n=0,1,2,\ldots,63) \qquad (15)$$

These 64-point length signals are shown in Fig. 2(a). The positions of the two actual sound soufces are shown by the mark "▼" in the resultant figures of the estimated power, $\{ P_i \}$. Five sensors are used and they are arranged at equal intervals along the surface of the pipe. The noise signal, $\{ n_j(n) \}$, that contaminates at the j-th sensor output ($j=1,2,\ldots,5$) is mutually uncorrelated white noise. The SNR is 20 dB. The number M of the assumed sound sources is 11. They are also arranged at equal intervals along the line and the source positions of the two actual sounds, $x_\alpha(n)$ and $x_\beta(n)$, coincide with the positions of the second and ninth assumed sources, respectively. The power, $\{P_i\}$, of the estimated spectrum, $\hat{X}(k)$, is averaged for the all frequencies, $k=1,\ 2,\ \ldots,\ 31$, other than the d.c. component (k=0).

Figures 2(b) and 2(c) show the waveform estimates, $\{\hat{x}_i(n)\}$, ($i=1,2,\ldots,11$), of each assumed sound source and the radiating sound power, $\{ P_i \}$, obtained by the cross-spectrum based method and the proposed method, respectively. Both methods accurately estimate the positions of the actual sounds from the radiating power, $\{ P_i \}$. However, even in the case in which noise is absent in Fig. 2(b-1), there are many impulses in the estimated signals of the assumed points other than the actual sound sources, (i=2 and i=9). These pulses show the difficulty of removing the transfer characteristics from each sensor output by using the simple inverse

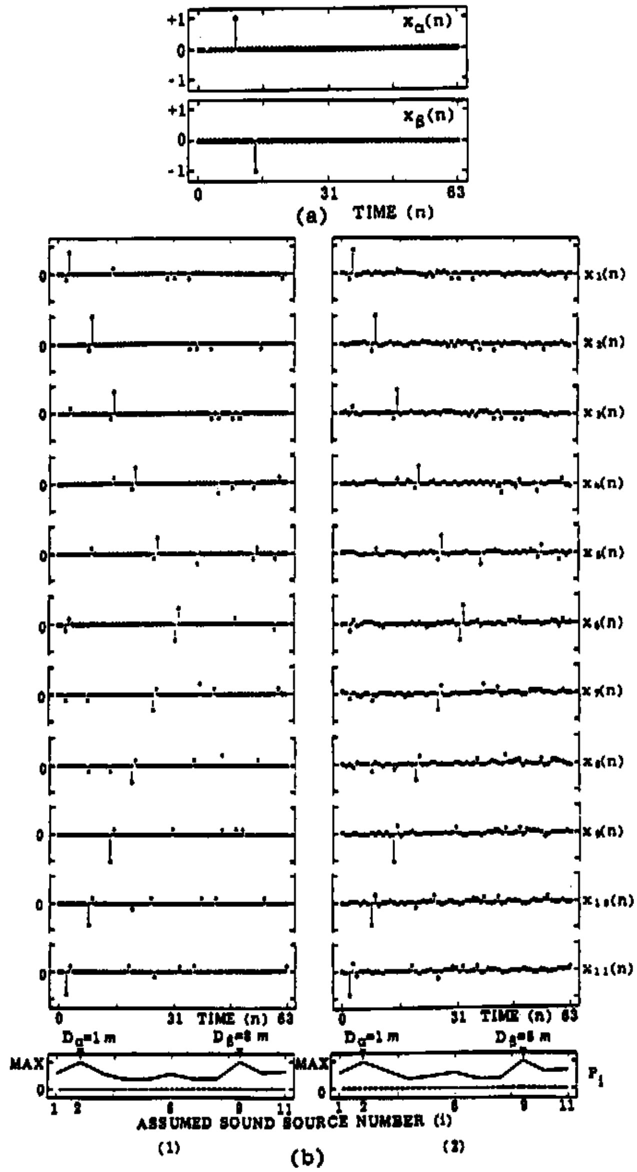

Fig. 2 The estimated signals, $\hat{x}_i(n)$, and the radiating power, P_i, of the i-th assumed sound sources for the first example. In the graphs of the estimated signals, the vertical axis scale is as follows: 1 unit equals 0.5. ($M_s=2$, $L=11$, $M=21$, and $N_w=0$.) (1) SNR=∞, (2) SNR= 20dB.

(a) The original signals, $x_\alpha(n)$ and $x_\beta(n)$, of the two actual source sounds. (b) Results obtained by the cross-spectrum based method in section 2.1.

characteristics, $H(k)^-$, of Eq. (5). Contrastively, by using the proposed method, the original waveform of Eq. (15) radiated from each actual source is estimated accurately, even when the SNR is equal to 20 dB as shown in Fig. 2(c-2).

As shown in Fig. 2(d-2) in the case of the SNR=20dB, the sound position cannot be estimated from the power, $\{P_i\}$, estimated from the proposed method when the threshold, T_e, used in Eq. (11) is equal to 100 %, that is, when all singular values of $H(k)$ are used for the estimation. This is due to the second term, $Q_s(k)\Sigma_R(k)^{-1}Q_s(k)^H N(k)$, of Eq. (10') caused by the additive noise. However, by truncating the small singular values (T_e=99.9%), more accurate estimates are obtained by the proposed method as shown in the power, $\{P_i\}$, of Fig. 2(c-2).

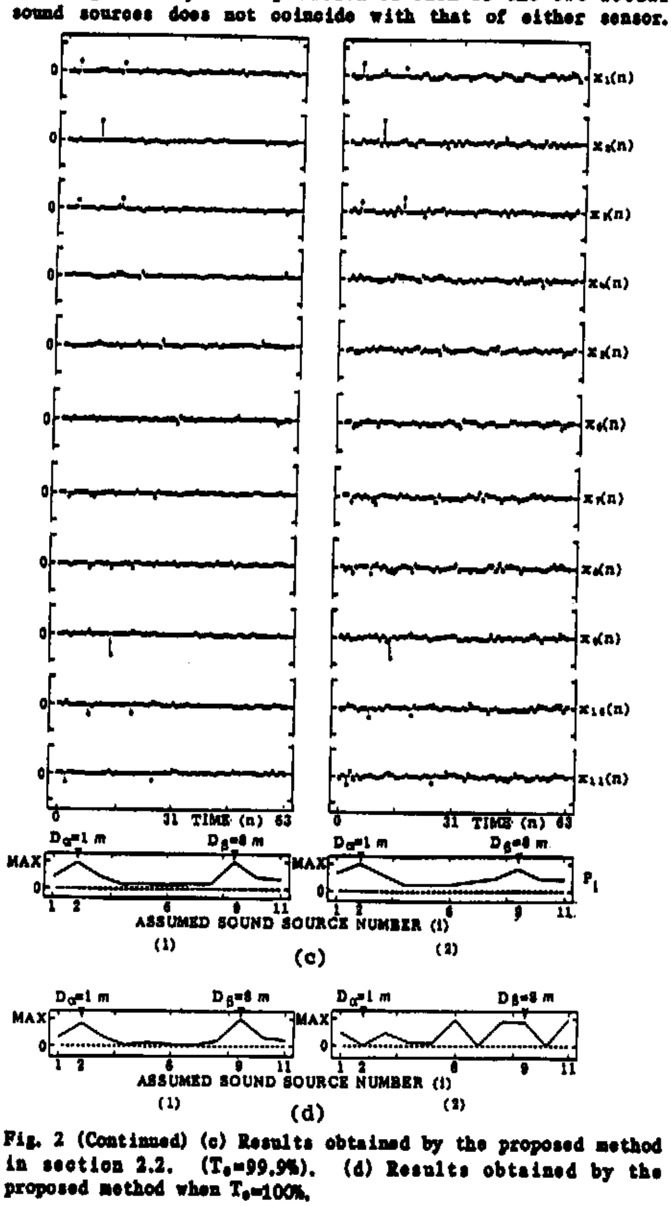

Fig. 2 (Continued) (c) Results obtained by the proposed method in section 2.2. (T_e=99.9%). (d) Results obtained by the proposed method when T_e=100%.

Second Example: Here we present another simple example in the case where two source sounds, $x_\alpha(n)$ and $x_\beta(n)$, are mutually uncorrelated white noise when there are two actual sound sources. Such uncorrelated source sounds are also used in the succeeding simulation experiments. The number of the assumed sounds, M, and the number of the used sensors, L, are 21 and 11, respectively. The position of each of the two actual sound sources does not coincide with that of either sensor. The SNR is equal to 20 dB. By using the cross-spectral based method in section 2.1, it is difficult to detect the position of the second actual sound source as shown in Fig. 3(a). However, as shown in Fig. 3(b), the proposed method reliably estimates the position of the each sound source.

In order to determine the optimum value of each parameter used in the proposed method, we chose the following three simulation experiments. Figure 4 shows the average estimation accuracy, γ, obtained in 32 independent trials for various values of the threshold, T_e, used in Eq. (11) and the width, N_w, of the tapering window. The estimation accuracy, γ, is defined as follows:

$$\gamma = \frac{E\left(\sum_{i \text{ for the actual sources}} P_i\right)/M_e}{E\left(\sum_{i \text{ for all assumed sources}} P_i\right)}, \qquad (16)$$

where $E(\cdot)$ denotes the average operation. The number of the actual sound sources, M_e, and the number of the sensors, L, are equal to 2 and 11, respectively. In each trial, the position of these two sound sources are selected randomly from the positions of the 21 assumed source points on the line. The SNR is equal to 20 dB. As shown in Fig. 4, the optimum values of the parameters, T_e and N_w, are about 99.9% and 4, respectively, and these values are also used in the succeeding simulation experiments. As seen from the figure, it is effective to use the tapering window.

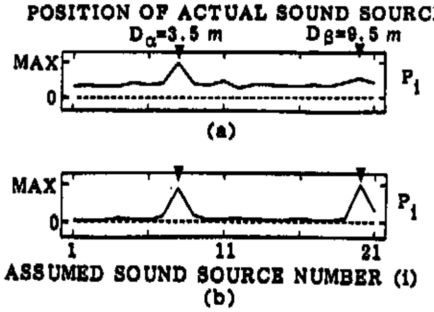

Fig. 3 The estimated radiating power, $\{P_i\}$, of the assumed sound sources when the source sound signal are mutually uncorrelated white noise for the second example. (SNR=20dB, M_e=2, L=11, and M=21).
(a) Estimated by the cross-spectrum based method in section 2.1. (b) Estimated by the proposed method in section 2.2. (T_e=99.9%, and N_w=4).

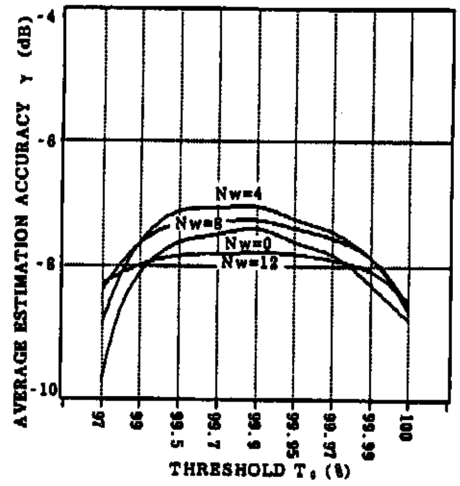

Fig. 4 The relation between the threshold, T_e, and the average estimation accuracy, γ, for various values of the width, N_w, of the tapering window. (SNR=20dB, M_e=2, L=11, and M=21.)

Next, the average estimation accuracy, γ, is calculated for various values of the number of sensors, L, and the number of the actual sound sources, M_e, as shown in Fig. 5. The number of the assumed sound sources, M, is equal to 21. From Fig. 5, the estimation accuracy, γ, is in proportion to the logarithm of the number, L.

Figure 6 shows the average estimation accuracy, γ, obtained for various values of the SNR and the number of the sensors, L. In the experiments, two actual sound sources are used and the number of the assumed sound source, M is 21. Enough estimation accuracy, γ, is obtained when the SNR is larger than 20 dB.

Third Example: We chose a slightly more complicated example as follows: There are four actual sound sources. The positions of these sources are shown in Fig. 7. The number, M, of the assumed sound source points is equal to 61, and 21 sensors are used. The SNR is 20 dB. As shown in Fig. 7(b), the proposed method accurately estimate the positions of all four actual sound sources from the short length signals received by the multiple sensors.

4. CONCLUDING REMARKS

This paper proposes a new method of locating multiple sound sources in one-dimensional space by applying tapered singular-value-decomposition to an under-determined normal equation derived from estimating the spectrum of each assumed sound source from the observed multiple sensor outputs. From the experiments with numerical examples, the multiple actual sound sources are estimated accurately using the proposed method. The relation between the accuracy of the estimated sound positions and the parameters used in the proposed method is also examined.

The usefulness of the proposed method should not only be confirmed the above computer simulation but also by experiments using pipes. When there are reflecting points in the pipe, it is more difficult to estimate the position and the waveform of each sound source. The principle of the proposed method is applicable in the estimation of the position and the waveform of the sound sources in a two- or three-dimensional space. However, in these cases, it is necessary to use a much larger number of the assumed source. These three important issues are currently under investigation.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dr. Noriyoshi Chubachi, Dr. Shozo Makino, Dr. Yoshiyuki Kawazoe, and Dr. Masato Abe of Tohoku university.

REFERENCES

Silvia, M. T., 1987, "Handbook of Digital Signal Processing Engineering Applications," Edited by D. F. Elliot. Academic Press. (Chapter 10: Deconvolution and Chapter 11: Time delay estimation).

Youn, D. H., and Ahmed, N., 1984, "Time Delay Estimation via Coherence: An Adaptive Approach, " J. Acoust. Soc. Am., Vol. 75, No. 2, February.

Shima, A., Liu, J., Noto, H., Abe, M., and Kido, K., 1988, "A New Cross-Spectral Method to Locate Sound Sources in One-Dimensional Space," J. Acoust. Soc. Jpn. Vol. (E)-9, No. 1, pp. 1-11.

Kanai, H., and Kido, K., 1988, "Estimation of Multipulse Series Driving of an All-pole Transfer System," Trans. of Jpn. Society of Mechanical Engineers (in Japanese). Vol. 54, No. 508, pp. 2907-2914.

Yanagida, M., Kakusho, O., and Nomura, Y., 1983, "Least-square Method from Source Sound Separation in Multi-source Environment," Proceedings of 11th International Conference on Acoustics, Lyon-Touluse, Paris, pp. 167-170.

Lancaster, P. and Tismenetsky, M., 1985, "The Theory of Matrices," Academic press, United Kingdom.

Cadzow, J. A., 1982, "Spectral Estimation: An Over-determined Rational Model Equation Approach," Proc. IEEE, Vol. 70, No. 9, pp. 907-939.

Oppenheim, A. V. and Schafer, R. W., 1975, "Digital Signal Processing," Prentice-Hall, Englewood Cliffs, New-Jersey.

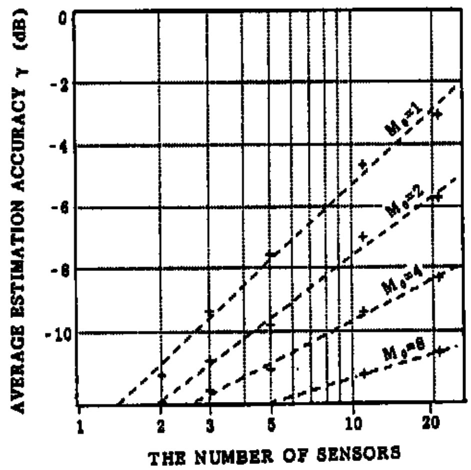

Fig. 5 The relation between the number of the sensors, L, and the estimation accuracy, γ, for various numbers of actual sound sources, M_e . (SNR=20dB, M=21, T_e=99.9%, and N_w=4.)

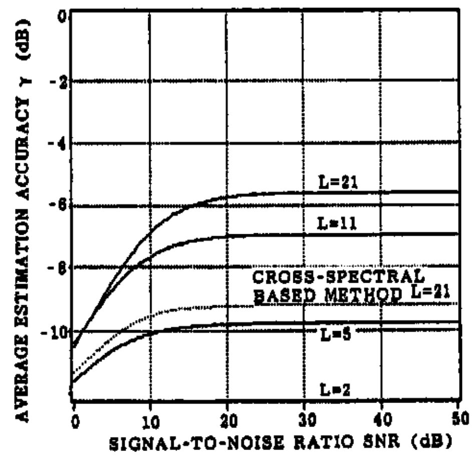

Fig. 6 The relation between the SNR and the estimation accuracy, γ, for various numbers of the sensors, L. (M_e=2, M=21, T_e=99.9%, and N_w=4.)
The solid line: the proposed method (L=2, 5, 11, 21).
The dotted line: the previous method by Shima (1988) (L=21).

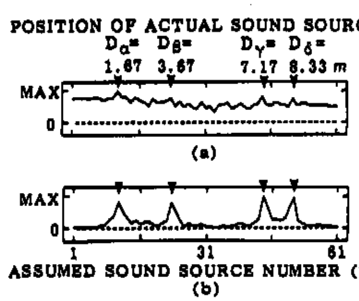

Fig. 7 The estimated radiating power (P_i) of the assumed sound sources when there are 4 actual sound sources for the third example. (SNR=20dB, M_e=4, L=21, and M=61).
(a) Estimated by the cross-spectrum based method in section 2.1. (b) Estimated by the proposed method in section 2.2. (T_e=99.9%, and N_w=4).