PAPER

A new algorithm for blind estimation of common poles in multiple transmission paths based on linear prediction

Takafumi Hikichi* and Masato Miyoshi

NTT Communication Science Laboratories, NTT Corporation,
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan

(Received 19 December 2003, Accepted for publication 1 July 2004)

Abstract: This paper proposes a blind calculation method for the poles common to multiple signal transmission paths. In the field of room acoustics, the poles correspond to the mode frequencies that are determined by room size and shape, and they do not change when source and receiver locations change. Information on these acoustic poles is useful for many applications, including echo cancellation and sound field equalization in a room. Conventional pole estimation methods require *a priori* measurement of the room transfer functions. This paper proposes a new method for the blind calculation of the poles, where the poles are calculated solely from the observed signals. Simulation results show that the proposed algorithm provides precise estimates of the common poles.

Keywords: Common poles, Room transfer functions, Blind calculation, Linear prediction

PACS number: 43.55.Br, 43.55.Mc [DOI: 10.1250/ast.26.1]

1. INTRODUCTION

In the field of room acoustics, poles correspond to mode frequencies determined according to the room dimensions, and do not change even when source and receiver locations change. Therefore, not only zeros but also poles have been effectively used to model room transfer functions [1,2]. Information on these acoustic poles is also useful for other applications. For example, when designing an echo canceller [1,3], this information can be effectively utilized to reduce the number of parameters that represent echo paths, and hence to increase the convergence speed. As another example, common acoustical poles are also useful for multiple point equalization, which achieves flat sound pressure distribution [4]. The Suppression of reverberation has been investigated with the state feedback control technique, where the pole information is controlled [5].

However, conventional pole estimation methods require the measurement of transfer functions from source to receivers. This means that precise information on a source signal is necessary in those methods. Therefore, it is not applicable in the case where no such information is available. In order to obtain the room transfer functions, source and received signals have to be synchronized and

therefore systems tend to become complex. On the other hand, if the poles can be obtained solely from the received signals, the system can become much simpler and can be widely applied to some situations, such as on-line type equalization. Such processing that are based only on received signals is referred to as “blind” [6] hereafter. This paper proposes a new blind method for calculating the poles common to multiple room transfer functions, without *a priori* measurements of the transfer functions. This method only assumes that the frequency characteristics of the input signal is flat. Unlike the other blind methods, statistical characteristics such as Gaussianity are not assumed.

This paper is organized as follows. Chapter 2 describes the conventional and the proposed calculation methods, respectively. Chapter 3 uses simulations to compare the estimation performance obtained with the proposed and conventional methods. The dependence of the performance on certain parameters is also studied. The conclusion is given in Chap. 4.

2. CALCULATION OF COMMON POLES OF MULTIPLE TRANSFER FUNCTIONS

We deal with a single-source and two-microphone system as shown in Fig. 1. When the AR part that corresponds to acoustical poles is denoted as $1 - a(z)$, two transfer functions $g_1(z)$, $g_2(z)$ are described as [1],

*e-mail: hikichi@cslab.kecl.ntt.co.jp

Fig. 1 A single-source two-microphone acoustic system.

$$g_i(z) = \frac{h_i(z)}{1 - a(z)} \quad (i = 1, 2). \quad (1)$$

where $h_i(z)$ denotes J -th polynomials,
 $a(z) = a_1 z^{-1} + a_2 z^{-2} + \dots + a_K z^{-K}$.

We assume $h_1(z)$ and $h_2(z)$ have no common zeros [7].

2.1. Conventional Method for Calculating Common Poles

The conventional method for calculating common poles is described. This method has been used for multiple point sound pressure equalization [4], and was originally proposed as a way of estimating model parameters for the ARMA model [1]. When MA order is not zero, the stability

of the algorithm is not guaranteed.

The common AR coefficients that correspond to the common poles can be estimated as those that minimize the mean squares cost function ϵ as,

$$\begin{aligned} \epsilon &= \sum_{i=1}^M \sum_{k=0}^{\infty} e_i^2(k), \\ &= \sum_{i=1}^M \sum_{k=0}^{\infty} \left[g_i(k) - \sum_{n=1}^K a_n g_i(k-n) \right]^2. \end{aligned} \quad (2)$$

where

$g_i(n)$ is a measured impulse response obtained by i -th source-microphone pairs ($1 \leq i \leq M$),

a_n are common AR parameters, and

M is the number of source-microphone pairs.

We deal with the case $M = 2$ hereafter. To minimize this cost function, all the partial derivatives of the cost function with respect to a_n must be zero. Then, the common AR coefficients a_n can be expressed as

$$\mathbf{a} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{v}, \quad (3)$$

where

$$\begin{aligned} \mathbf{a} &= [a_1, a_2, \dots, a_K]^T, \\ \mathbf{W} &= [\mathbf{G}_1^T, \mathbf{G}_2^T]^T, \\ \mathbf{v} &= [\mathbf{g}_1^T, \mathbf{g}_2^T]^T, \\ \mathbf{g}_i &= [g_i(1), g_i(2), \dots, g_i(N-1), 0, 0, \dots, 0]^T, \\ \mathbf{G}_i &= \begin{bmatrix} g_i(0) & 0 & \dots & 0 \\ g_i(1) & g_i(0) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ g_i(K-1) & g_i(K-2) & \dots & g_i(0) \\ \vdots & \vdots & & \vdots \\ g_i(N-1) & g_i(N-2) & \dots & g_i(N-K) \\ 0 & g_i(N-1) & \dots & g_i(N-K-1) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & g_i(N-1) \end{bmatrix} \quad (N+K-1). \end{aligned}$$

$\underbrace{\hspace{15em}}_K$

Since this method uses actual impulse responses to estimate the poles, it is necessary to measure them in advance.

2.2. Proposed Method

The proposed method blindly calculates common poles without measuring transfer functions $g_1(z)$ and $g_2(z)$, assuming that a stationary white noise is input into the system. Furthermore, common poles are expected to be

estimated precisely even when using fewer transfer functions than required with the conventional method.

The signals observed by the two microphones are denoted as $x_1(n)$ and $x_2(n)$. Here, we introduce the two channel linear prediction matrix $\mathbf{Q}$ defined by the following equation,

$$\mathbf{x}_n^T = \mathbf{x}_{n-1}^T \mathbf{Q}, \quad (4)$$

Fig. 2 A block diagram that represents the input-output relationship of the acoustic system shown in Fig. 1. The transfer function $g_i(z)$ is divided into $h_i(z)$ and $1/(1-a(z))$.

where

$$\mathbf{x}_n = [x_1(n), \dots, x_1(n-m), x_2(n), \dots, x_2(n-m)]^T.$$

By multiplying $\mathbf{x}_{n-1}$ with this equation from the left-hand side and employing expectation, the above equation can be written as,

$$E\langle \mathbf{x}_{n-1} \mathbf{x}_n^T \rangle = E\langle \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T \rangle \mathbf{Q}, \quad (5)$$

where $E\langle \cdot \rangle$ is an expectation operator.

Here, it should be noted that the input and output relationship can be represented by the block diagram shown in Fig. 2. In Fig. 2, signal $u(n)$ produced by AR process $1/(1-a(z))$ is input into the two transfer functions $h_1(z)$ and $h_2(z)$. Then, signal $u(n)$ is expressed by the

following equation [8].

$$\mathbf{u}_n = \mathbf{C}^T \mathbf{u}_{n-1} + \mathbf{e}_n, \quad (6)$$

where

$$\mathbf{u}_n = [u(n), u(n-1), \dots, u(n-m-J+1)]^T,$$

$$\mathbf{e}_n = [e(n), 0, \dots, 0]^T,$$

$$\mathbf{C} = \begin{bmatrix} a_1 & a_2 & \dots & a_K & 0 & 0 \\ 1 & 0 & \dots & 0 & \vdots & \vdots \\ 0 & 1 & & 0 & \vdots & \vdots \\ & & \ddots & & \vdots & \vdots \\ \vdots & & & 1 & \vdots & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 \end{bmatrix}^T \quad (m+J),$$

$e(n)$ is stationary white noise with unit variance, and $m+J > K$.

Then, $\mathbf{x}_{n-1}$ can be written as,

$$\mathbf{x}_{n-1}^T = \mathbf{u}_{n-1}^T \mathbf{H}, \quad (7)$$

where

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2],$$

$$\mathbf{H}_i = \begin{bmatrix} h_i(0) & 0 & \dots & 0 \\ h_i(1) & h_i(0) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ h_i(m) & h_i(m-1) & \dots & h_i(0) \\ \vdots & \vdots & & \vdots \\ h_i(J-1) & h_i(J-2) & \dots & h_i(J-m-1) \\ 0 & h_i(J-1) & \dots & h_i(J-m) \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & h_i(J-1) \end{bmatrix} \quad (m+J),$$

(m+1)

It should be noted that matrix $\mathbf{H}$ has more columns than rows, i.e., $2(m+1) > m+J$. Matrix $\mathbf{Q}$ in Eq. (5) is calculated, in a minimum-mean-square-error (MMSE) sense, as follows.

$$\begin{aligned} \mathbf{Q} &= (E\langle \mathbf{x}_{n-1} \mathbf{x}_{n-1}^T \rangle + \delta^2 \mathbf{I})^{-1} E\langle \mathbf{x}_{n-1} \mathbf{x}_n^T \rangle, \\ &= (E\langle \mathbf{H}^T \mathbf{u}_{n-1} \mathbf{u}_{n-1}^T \mathbf{H} \rangle + \delta^2 \mathbf{I})^{-1} \\ &\quad E\langle \mathbf{H}^T \mathbf{u}_{n-1} \mathbf{u}_n^T \mathbf{H} \rangle, \end{aligned} \quad (8)$$

$$= (\mathbf{H}^T E\langle \mathbf{u}_{n-1} \mathbf{u}_{n-1}^T \rangle \mathbf{H} + \delta^2 \mathbf{I})^{-1}$$

$$\mathbf{H}^T E\langle \mathbf{u}_{n-1} \mathbf{u}_n^T \rangle \mathbf{C} \mathbf{H},$$

where $\mathbf{I}$ is the unit matrix, and δ is a small positive number.

Since $E\langle \mathbf{u}_{n-1} \mathbf{u}_{n-1}^T \rangle$ can be assumed to be positive-definite, it can be replaced by $\mathbf{U}^T \mathbf{U}$, where $\mathbf{U}$ is a matrix. Then, matrix $\mathbf{Q}$ can be written as,

$$\mathbf{Q} = (\mathbf{H}^T \mathbf{U}^T \mathbf{U} \mathbf{H} + \delta^2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{U}^T \mathbf{U} \mathbf{C} \mathbf{H}. \quad (9)$$

If we assume that matrix $\mathbf{H}$ has a full row-rank [9], using the definition of the Moore-Penrose inverse matrix [10], we can rewrite the above equation as,

$$\begin{aligned}\mathbf{Q} &= \lim_{\delta \rightarrow 0} (\mathbf{H}^T \mathbf{U}^T \mathbf{U} \mathbf{H} + \delta^2 \mathbf{I})^{-1} \mathbf{H}^T \mathbf{U}^T \mathbf{U} \mathbf{C} \mathbf{H}, \\ &= (\mathbf{U} \mathbf{H})^+ \mathbf{U} \mathbf{C} \mathbf{H}, \\ &= \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{U} \mathbf{C} \mathbf{H}, \\ &= \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{H},\end{aligned}\quad (10)$$

where $\mathbf{A}^+$ denotes the Moore-Penrose inverse matrix of $\mathbf{A}$ defined as,

$$\mathbf{A}^+ = \lim_{\delta \rightarrow 0} (\mathbf{A}^T \mathbf{A} + \delta^2 \mathbf{I})^{-1} \mathbf{A}^T.$$

The following relation holds for the non-zero eigenvalues of matrix $\mathbf{Q}$ [10],

$$\begin{aligned}\lambda(\mathbf{Q}) &= \lambda(\mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{H}), \\ &= \lambda(\mathbf{H} \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C}), \\ &= \lambda(\mathbf{C}),\end{aligned}\quad (11)$$

where $\lambda(\mathbf{A})$ are the non-zero eigenvalues of matrix $\mathbf{A}$. As a result, the following relationship is derived.

$$f_c(\mathbf{Q}) \equiv f_c(\mathbf{C}) \equiv 1 - a(z), \quad (12)$$

where $f_c(\mathbf{A})$ is the characteristic polynomial of matrix $\mathbf{A}$. That is, the characteristic polynomial of matrix $\mathbf{Q}$ is equivalent to the AR polynomial $1 - a(z)$ [8,11]. Hence, the AR polynomial is blindly calculated from observed signals.

The proposed algorithm is summarized as follows.

- (1) The two-channel linear prediction matrix $\mathbf{Q}$ with $2(m+1) \times 2(m+1)$ is calculated using output signal vectors $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$ by Eq. (8), where m satisfies $m \geq J-1$ and $m > K-J$.
- (2) The characteristic polynomial of matrix $\mathbf{Q}$ is calculated to obtain the AR polynomial $1 - a(z)$ that corresponds to the common poles.

3. SIMULATIONS

3.1. Conditions

Simulations were performed to show the validity of the proposed algorithm. We assumed the simulation setup as shown in Fig. 3, where a simple sound field is surrounded by three reflectors. From this situation, the transfer functions $g_i(z)$ in Eq. (1) were determined as follows. As the MA part, $h_i(z)$, we simulated the initial reflections by the image method [12]. As the AR part, two transfer functions $g_i(z)$ are expected to have the theoretical resonance frequencies calculated by $f_n = nc/2L$, where L is the distance between two opposing walls, c is sound velocity, and n is an integer (e.g. [13]). The AR part, $1 - a(z)$, was determined to have this theoretical frequen-

Fig. 3 Simulation setup.

Table 1 Simulation conditions.

J	Order of zeros	8
K	Order of poles	6
f_1	Lowest resonance frequency	171.5 Hz
f_n	Resonance frequencies	$f_1, 2f_1, 3f_1$
r_n	Radii of complex poles	0.9, 0.85, 0.8
N	Order of impulse responses	80
f_s	Sampling frequency	2 kHz

cies. Generally, the higher the resonant frequencies, the smaller the power of the resonances. Therefore, the radii of the poles corresponding to the resonances were weighted so as to attenuate the poles of higher frequencies more. As a result, the effect of the 4th pole on the frequency characteristics became negligible small, hence this was omitted. These conditions are shown in Table 1.

Figure 4(a) shows the impulse responses of $g_1(z)$ and $g_2(z)$, and (b) shows their frequency responses. The impulse responses are truncated at the point where the magnitude of the response seems sufficiently small. from the beginning. We used these two impulse responses to estimate the common AR parameters by the conventional method.

The Spectral Distortion (SD) [14], shown in the following, was used to evaluate the results.

$$SD = \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} \{20 \log |P(f)| - 20 \log |\hat{P}(f)|\}^2}, \quad (13)$$

where

$P(f) = 1/(1 - a(e^{j\pi f/F}))$ is the given all-pole spectrum, $\hat{P}(f) = 1/(1 - \hat{a}(e^{j\pi f/F}))$ is the estimated all-pole spectrum.

3.2. Comparison with Conventional Method

Figure 5(a) shows the given and estimated all-pole spectra provided by the conventional method. Here, the AR order is set to match the actual AR order, i.e., $K = 6$. The figure shows the discrepancy between the estimated and reference responses. In this case, $SD = 3.8$ [dB] was obtained.

Fig. 4 Time and frequency responses of $g_1(z)$ and $g_2(z)$.

Fig. 5 Estimated all-pole spectra (The reference spectrum is shown by a dashed line in each figure).

Next, we tested the proposed calculation under the same conditions. We used a stationary white noise signal with a normal distribution, zero mean, and unit variance.

Using this random signal as the input, we estimated the AR coefficients from the two channel output signals. Since this input signal is not perfectly random, it is necessary to suppress the effect of imperfect randomness on the estimation precision. To achieve this, we used input signals with sufficiently long samples (5,000 samples), and we obtained the estimated AR parameters by averaging over 20 repetitions. The observed signals were normalized by their root-mean-square values then matrix $\mathbf{Q}$ was calculated with Eq. (8).

Figure 5(b) shows the given and estimated all-pole spectra obtained with the proposed method. The output signal length m was adjusted to 7 to meet the above mentioned requirement, i.e., $m \geq J - 1$. It is shown that the estimated spectrum agrees quite well with the actual response, and the obtained SD value is much smaller than the conventional result. In this case, the estimated AR order is $\hat{K} = 14$. Here, the parameter δ^2 , referred to as the precision parameter, was adjusted so as to get the smallest distortion ($\delta^2 = 10^{-9}$ was used).

The pole location was calculated from the estimated AR coefficients. Figure 6 shows the estimated and given poles. The vertical axis shows the absolute values of the complex poles. This figure also shows the validity of the proposed method.

3.3. Effect of the Parameters in the Proposed Method

The proposed algorithm calculates $2(m+1)$ -dimensional correlation matrices of the output signals, and calculates matrix $\mathbf{Q}$ using a precision parameter δ^2 . Then, the desired AR polynomial can be obtained by calculating the characteristic polynomial of matrix $\mathbf{Q}$. The accuracy of

Fig. 6 Comparison of estimated and given poles.

Fig. 7 Estimated all-pole spectra with different m values (The reference spectrum is shown by a dashed line in each figure).

the estimated AR polynomial depends on the parameters m and δ^2 . The effect of these parameters is investigated here. Different from the proposed method, the conventional method has no such parameters that can affect the accuracy of the estimated poles.

Figure 7 shows the estimated all-pole spectra with $\delta^2 = 10^{-9}$, where the calculated AR order is varied in three cases, namely $m = 3$, 7, and 20. In the first case ($m = 3$), the response exhibits degradation. It should be noted that this does not satisfy the above requirement. In contrast, in the second ($m = 7$) as shown in Fig. 7(b) which is a replot of Fig. 5(b), and the third case ($m = 20$), the response was estimated precisely.

Figure 8 shows the spectral distortion for the same δ^2 parameter, where m varied from 3 to 42. The distortion decreased when m reached 7. This corresponds to $J = 8$, which coincides with the actual order of zeros, as shown in Table 1. Even when the order is further increased, the SD value remains almost constant. It may therefore be concluded that when a sufficiently long AR order is given, the poles can be estimated precisely. Figure 8 also shows the spectral distortion for the parameter $\delta^2 = 10^{-7}$. The figure shows that the δ^2 value determines the upper limit of

Fig. 8 Effect of the order on the spectral distortion with $\delta^2 = 10^{-7}$ and 10^{-9} .

Fig. 9 Estimated spectra and the corresponding poles with $\delta^2 = 10^{-7}$ and 10^{-9} .

the estimation accuracy. The accuracy improves as δ^2 becomes smaller.

This dependence of the accuracy on the δ^2 parameter is shown in terms of the all-pole spectra and the corresponding poles in Fig. 9. When δ^2 becomes large, the spectrum shows a discrepancy, and the poles move away from the reference locations. In contrast, when the δ^2 value was too small, the inverse matrix in Eq. (8) was not correctly calculated, and the performance suffered from numerical errors.

4. CONCLUSIONS

This paper proposed a blind calculation method for estimating common poles based on linear predictions that does not require *a priori* measurement of the transfer functions. Furthermore, by utilizing the characteristic polynomial of the linear prediction matrix, the AR

polynomial is directly estimated, which means that the pole estimation precision is unaffected by the zeros that may exist in the transfer functions. Simulation results showed that the proposed algorithm provides precise estimates of the common poles.

The proposed method has a few parameters that must be appropriately chosen. The effect of these parameters was examined. It was concluded that (1) when a sufficiently large order is used, the poles are estimated correctly, and (2) the precision parameter determines the upper limit of the performance.

ACKNOWLEDGEMENTS

The authors express their gratitude to Dr. Frank K. Soong, Invited Researcher of ATR Spoken Language Translation Research Laboratories, for his invaluable comments.

REFERENCES

- [1] Y. Haneda, S. Makino and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, **2**, 320–328 (1994).
- [2] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vib.*, **146**, 281–302 (1991).
- [3] G. Long, D. Shwed and D. D. Falconer, "Study of a pole-zero adaptive echo canceller," *IEEE Trans. Circuits Syst.*, **34**, 765–769 (1987).
- [4] Y. Haneda, S. Makino and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Process.*, **5**, 325–333 (1997).
- [5] T. Samejima and M. Yasuoka, "Active suppression of reverberation in a sound field through state feedback control," *J. Acoust. Soc. Jpn. (J)*, **55**, 12–22 (1999).
- [6] S. Haykin, *Unsupervised Adaptive Filtering*, Vol. II: Blind deconvolution (John Wiley & Sons, New York, 2000).
- [7] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust. Speech Signal Process.*, **36**, 145–152 (1988).
- [8] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation* (Prentice-Hall, New Jersey, 2000).
- [9] S. Haykin, *Adaptive Filter Theory* (Prentice-Hall, New Jersey, 1996).
- [10] S. Kodama and N. Suda, *Matrix theory for Systems and Control* (Corona Publishing Co., Tokyo, 1978).
- [11] M. Miyoshi, "Estimating AR parameter-sets for linear-recurrent signals in convolutive mixtures," *Proc. ICA 2003*, pp. 585–589 (2003).
- [12] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, **65**, 943–950 (1979).
- [13] H. Kuttruff, *Room Acoustics* (Elsevier Applied Science, London, 1973).
- [14] W. B. Klein and K. K. Paliwai, Eds., *Speech Coding and Synthesis* (Elsevier, Amsterdam, 1995).

Takafumi Hikichi was born in Nagoya, in 1970. He received his Bachelor and Master of Electrical Engineering degrees from Nagoya University in 1993 and 1995, respectively. In 1995, he joined the Basic Research Laboratories of NTT. He is now working at the Signal Processing Research Group of the Communication Science Laboratories, NTT. His research interests include physical modeling of musical instruments, room acoustic modeling, and signal processing for speech enhancement and dereverberation. He was honored to receive the Kiyoshi-Awaya Incentive Awards from the ASJ in 2000. He is currently a visiting associate professor of the Graduate School of Information Science, Nagoya University. He is a member of IEEE, ASA, ASJ, IEICE, and IPSJ.

Masato Miyoshi received M. E. from Doshisha University in Kyoto in 1983. Since joining NTT as a research staff member this year, he has been engaging on the research and development of acoustic signal processing technologies. Currently, he is in charge of a group leader of the Media Information Lab. of NTT Communication Science Labs. He was honored to receive 1988 IEEE ASSP Senior Awards, 1989 ASJ Kiyohi-Awaya Incentive Awards, and 1990 ASJ Satoh Paper Awards. He also received Ph. D. from Doshisha Univ. in 1991. He is a member of IEEE, AES, ASJ and IEICE.