# Common Acoustical Pole and Zero Modeling of Room Transfer Functions

Yoichi Haneda, *Associate Member, IEEE*, Shoji Makino, *Member, IEEE*, and Yutaka Kaneda, *Member, IEEE*

*Abstract*—**A new model for a room transfer function (RTF) by using common acoustical poles that correspond to resonance properties of a room is proposed. These poles are estimated as the common values of many RTF's corresponding to different source and receiver positions. Since there is one-to-one correspondence between poles and AR coefficients, these poles are calculated as common AR coefficients by two methods: i) using the least squares method, assuming all the given multiple RTF's have the same AR coefficients and ii) averaging each set of AR coefficients estimated from each RTF. The estimated poles agree well with the theoretical poles when estimated with the same order as the theoretical pole order. When estimated with a lower order than the theoretical pole order, the estimated poles correspond to the major resonance frequencies, which have high $Q$ factors. Using the estimated common AR coefficients, the proposed method models the RTF's with different MA coefficients. This model is called the common-acoustical-pole and zero (CAPZ) model, and it requires far fewer variable parameters to represent RTF's than the conventional all-zero or pole/zero model. This model was used for an acoustic echo canceller at low frequencies, as one example. The acoustic echo canceller based on the proposed model requires half the variable parameters and converges 1.5 times faster than one based on the all-zero model, confirming the efficiency of the proposed model.**

## I. Introduction

A ROOM transfer function (RTF) expresses the transmission characteristics of a sound between a source and a receiver in a room. Modeling RTF's is a key technique for many applications that simulate RTF's. For example, an acoustic echo canceller (AEC) has an adaptive filter based on an RTF model. Another example is a sound field simulator, which simulates the sound field of rooms. An efficient model is required that can represent the RTF with few parameters because, as one example, the convergence speed of the adaptive filter depends on the number of parameters. An efficient model is also required to represent the many RTF's corresponding to different source and receiver positions with fewer parameters because, as another example, an efficient model can store all the RTF data in a small memory space.

The usual method of modeling an RTF is an all-zero model. The coefficients of this model correspond to the impulse response of the RTF in the time domain. The all-zero model can be implemented with an FIR filter. When the room has a long reverberation time, however, the all-zero model requires a large number of parameters to represent the RTF, i.e., many FIR filter coefficients are required. For example, when the reverberation time of the RTF is 500 ms, the FIR filter needs 4000 coefficients (8 kHz sampling) to represent the RTF with $-60$ dB accuracy. Furthermore, when the RTF varies due to, for example, changes in the source and receiver positions, all the parameters of the all-zero model change. These result in slow convergence in an acoustic echo canceller and require a large memory for a sound field simulator.

A pole/zero model (ARMA model) is also used as an RTF model [1]–[7]. From the physical point of view, poles represent resonances, and zeros represent time delays and anti-resonances. Because the poles can represent a long impulse response caused by resonances with fewer parameters than the zeros, the pole/zero model seems to match a physical room transfer function better than the all-zero model. In particular, a conventional pole/zero model has been studied to reduce the number of parameters. It has been reported that the order of the parameters of the pole/zero model is smaller than that of the all-zero model, especially at low frequencies [2], [3]. Schönle *et al.* reported that the pole/zero modeling of RTF's by multirate systems reduces the computational complexity for a sound field simulator [7]. It is possible to construct sound field simulators with pole/zero modeling at low frequency bands and all-zero modeling at high frequency bands using the subband technique.

In the conventional pole/zero model, however, both poles and zeros are estimated as variable parameters for the RTF variations, although acoustical (physical) poles which correspond to the resonance properties are invariant. Therefore, when the RTF changes, all the parameters of the pole/zero model also change, like in the conventional all-zero model.

A possible solution to this problem would be to estimate parameters that remain constant despite RTF variations due to, for example, changes in the source and receiver positions or the movement of people and to use them when modeling RTF's [8]–[11]. In this paper, we propose a new pole/zero model that has constant poles and variable zeros. In this new model, we use the estimated acoustical poles of a room as common constant poles. Acoustical poles correspond to the resonances of a room, and they do not change even if the source and receiver positions change or people move. Because the order of the acoustical poles in a room is so small that all of the common acoustical poles can be estimated at low frequencies, the proposed model is especially effective at low frequencies. These poles are first estimated from multiple RTF's. Then, the RTF's are represented by constant poles and by different sets of zeros. In the proposed model, variations in the RTF

Manuscript received December 3, 1992; revised August 2, 1993. The associate editor coordinating the review of this paper and approving it for publication was Prof. James F. Kates.

The authors are with the NTT Human Interface Laboratories, Tokyo, Japan.

IEEE Log Number 9215236.

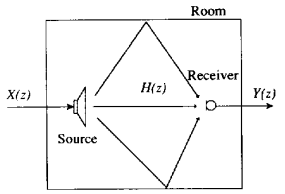


Fig. 1. RTF between a source and a receiver in a room.

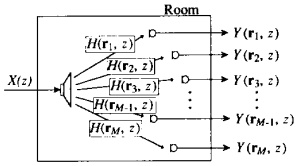


Fig. 2. RTF's corresponding to different source and receiver positions $\mathbf{r}_j$ $(j = 1, 2, \cdots, M)$.

require far fewer variable parameters to represent them than the previous conventional pole/zero model. Therefore, this model can reduce the memory space needed to store the coefficients of many modeled RTF's.

This paper is organized as follows. Section II reviews the conventional modeling of the RTF's. Section III proposes the common-acoustical-pole and zero model. We also propose two methods for estimating the common acoustical poles from multiple RTF's. The estimation results of common acoustical poles are discussed in Section IV. In Section V, the proposed model is applied to an AEC, and computer simulations demonstrate its advantages.

## II. Conventional Models of Room Transfer Functions

An RTF describes the sound transmission characteristics between a source and a receiver. When the source signal and the RTF are denoted by $X(z)$ and $H(z)$, the output signal $Y(z)$ of the receiver is expressed by

$$Y(z) = H(z)X(z), \tag{1}$$

as shown in Fig. 1. The RTF $H(z)$ includes the characteristics of the direct sound and all reflected sounds in the room.

Now, we consider the modeling of the multiple RTF's $H(\mathbf{r}_j, z)$ $(j = 1, \ldots, M)$ shown in Fig. 2, where $\mathbf{r}_j$ represent the source and receiver positions, and $Y(\mathbf{r}_j, z)$ represent the received signals at each receiver position of $\mathbf{r}_j$. Since the arrival times and the amplitudes of direct and reflected sounds are different for each $\mathbf{r}_j$, these RTF's $H(\mathbf{r}_j, z)$ are different from each other.

### A. All-Zero Model

The RTF can be modeled by the conventional all-zero model, which can be represented with either zeros $q_i(\mathbf{r}_j)$'s or MA coefficients $b_i(\mathbf{r}_j)$'s as

$$\hat{H}(\mathbf{r}_j, z) = C\ z^{-Q_1} \prod_{i=1}^{Q_2} (1 - q_i(\mathbf{r}_j)\ z^{-1}) = \sum_{i=0}^{Q} b_i(\mathbf{r}_j) z^{-i}, \tag{2}$$

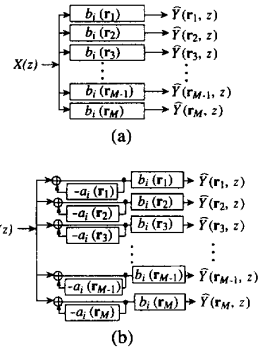


Fig. 3. Block diagram of multiple RTF modeling based on conventional models: (a) All-zero model; (b) pole/zero model.

where

| | |
|---|---|
| $\hat{H}(\mathbf{r}_j, z)$ | all-zero (MA) modeled RTF corresponding to each $\mathbf{r}_j$, |
| $\mathbf{r}_j$ | positions of the source and receiver, $j = 1, 2, \cdots, M$, |
| $M$ | number of RTF's, |
| $C$ | gain constant, |
| $Q, Q_1, Q_2$ | order of zeros, $(Q = Q_1 + Q_2)$, |
| $q_i(\mathbf{r}_j)$ | zero for each $\mathbf{r}_j$, |
| $b_i(\mathbf{r}_j)$ | MA coefficient for each $\mathbf{r}_j$. |

The MA coefficients correspond to the impulse response coefficients of the RTF.

Fig. 3(a) illustrates the modeling of the RTF's using the all-zero model. The boxes in Fig. 3(a) represent FIR filters with the coefficients $b_i(\mathbf{r}_j)$ $(i = 0, \cdots, Q;\ j = 1, \cdots M)$. To model $M$ RTF's, i.e., $H(\mathbf{r}_j, z)$ $(j = 1, \cdots M)$, the all-zero model requires $M$ FIR filters, each of which has $Q + 1$ coefficients. Therefore, the total number of different coefficients needed to represent $M$ RTF's is $M \times (Q + 1)$.

### B. Pole/Zero Model

The RTF can be also modeled by the pole/zero model. This model is equivalent to the ARMA model. It can be represented with either poles $p_i(\mathbf{r}_j)$ and zeros $q_i(\mathbf{r}_j)$ or AR coefficients $a_i(\mathbf{r}_j)$ and MA coefficients $b_i(\mathbf{r}_j)$ as

$$\hat{H}(\mathbf{r}_j, z) = \frac{C z^{-Q_1} \prod_{i=1}^{Q_2} (1 - q_i(\mathbf{r}_j) z^{-1})}{\prod_{i=1}^{P} (1 - p_i(\mathbf{r}_j) z^{-1})} = \frac{\sum_{i=0}^{Q} b_i(\mathbf{r}_j) z^{-i}}{1 + \sum_{i=1}^{P} a_i(\mathbf{r}_j) z^{-i}}, \tag{3}$$

where

| | |
|---|---|
| $\hat{H}(\mathbf{r}_j, z)$ | pole/zero (ARMA) modeled RTF corresponding to each $\mathbf{r}_j$, |
| $Q, Q_1, Q_2$ | order of zeros, $(Q = Q_1 + Q_2)$, |
| $P$ | order of poles, |
| $p_i(\mathbf{r}_j)$ | pole for each $\mathbf{r}_j$, |
| $a_i(\mathbf{r}_j)$ | AR coefficient for each $\mathbf{r}_j$. |

The pole/zero model can be implemented with an IIR filter. Fig. 3(b) shows pole/zero modeling of RTF's using IIR filters. The boxes in Fig. 3(b) represent recursive filter sections with the AR coefficients $a_i(\mathbf{r}_j)$ and nonrecursive (or FIR) filters

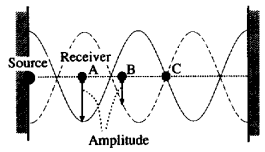


Fig. 4. Standing wave in a 1-D room. Since resonance can be observed at any source and receiver position except node points (such as point $C$), all the RTF's have common acoustical poles that reflect the information of the resonance frequencies. The amplitudes are different at different points, e.g., $A$, $B$, and $C$.

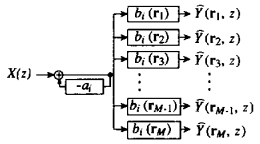


Fig. 5. Block diagram of multiple RTF modeling based on the proposed CAPZ model.

with the MA coefficients $b_i(\mathbf{r}_j)$. In the conventional pole/zero model, the poles and zeros, that is, the AR coefficients and the MA coefficients, are obtained for each RTF with $\mathbf{r}_j$. Therefore, all the AR and MA coefficients depend on $\mathbf{r}_j$, and the conventional pole/zero model requires $M \times (P + Q + 1)$ coefficients to represent $M$ RTF's, $H(\mathbf{r}_j, z)$ $(j = 1, \cdots M)$.

Thus, all the parameters (coefficients) are dependent on $\mathbf{r}_j$ $(j = 1, \cdots M)$, and many different coefficient values are required to represent many RTF's in the conventional all-zero model and pole/zero model.

## III. Proposed Model

### A. Common-Acoustical-Pole and Zero Model

To model the RTF's efficiently with few parameters, we propose a new pole/zero model that has invariable (or constant) parameters for the RTF variations. This new model uses acoustical poles as the invariables. The RTF variations are caused by, for example, changes in the source and receiver positions, or moving people. Here, we will consider the RTF variations corresponding to the source and receiver positions.

The background of the proposed model is that the acoustical poles of the RTF's are physically common to all RTF's. As is known, the general form of the pressure response between a source and a receiver in a room is given by the Green's function of the room, which is dependent on the coordinates of these two points. Assuming the ideal case when the source pressure is not a sinusoidal signal but instead has the properties of a delta function, then the Green's function will represent the RTF. The "eigenfrequencies (resonance frequencies)" and the "damping constant (corresponding to $Q$ factors)" of this function do not depend on source and receiver positions [12]. It can be assumed that the poles of the modeled RTF correspond to the "eigenfrequencies" and "damping constants" of the RTF; therefore, they do not depend on the source and receiver positions in a room. These resonance frequencies and those $Q$ factors of each the room are referred to as common acoustical poles.

Considering standing waves helps us to understand that all the RTF's have common acoustical poles. Standing waves occur at the resonance frequencies, and the resonance can be observed at any receiver point except node points, as shown in Fig. 4. This shows that all the RTF's between any source point and any receiver point include the information of the resonance frequencies, i.e., all the RTF's have common acoustical poles, which correspond to the resonance characteristics. However, the amplitudes are different for different receiver positions, as shown in Fig. 4. The difference in the amplitudes is reflected in the zeros of the RTF's.

Considering that the common acoustical poles do not depend on the source and receiver positions, we propose the common-acoustical-pole and zero (CAPZ) model. This can be represented in both pole/zero and ARMA forms:

$$\hat{H}(\mathbf{r}_j, z) = \frac{Cz^{-Q_1}\prod_{i=1}^{Q_2}(1 - q_i(\mathbf{r}_j)z^{-1})}{\prod_{i=1}^{P}(1 - p_i z^{-1})} = \frac{\sum_{i=0}^{Q} b_i(\mathbf{r}_j)z^{-i}}{1 + \sum_{i=1}^{P} a_i z^{-i}}, \quad (4)$$

where

$\hat{H}(\mathbf{r}_j, z)$ proposed CAPZ model of an RTF corresponding to each $\mathbf{r}_j$,

$p_i$ common acoustical pole,

$a_i$ common AR coefficient.

The difference between (4) and (3) is that in (4), poles $p_i$ and AR coefficients $a_i$ do not depend on the $\mathbf{r}_j$.

With the proposed model, first the common acoustical poles are estimated. Then, only zeros are estimated for each RTF. With the conventional pole/zero model, on the other hand, both poles and zeros are estimated for each RTF. The block diagram of the proposed modeling of the RTF's is shown in Fig. 5. The proposed model represent RTF's with one recursive filter that has common AR coefficients $a_i$ and with nonrecursive filters that each have a unique set of MA coefficients $b_i(\mathbf{r}_j)$. Consequently, the proposed model requires $P + M \times (Q + 1)$ coefficients to represent $M$ RTF's $H(\mathbf{r}_j, z)$ $(j = 1, \cdots, M)$.

### B. Estimation of Common Acoustical Poles by Least Squares Method

All acoustical poles are not necessarily observed in a single RTF, although they are common to all the RTF's in a room. For example, in Fig. 4, the resonance (standing wave) cannot be observed at node point $C$. This phenomenon can be explained as the effect of zeros in the RTF's. Zeros are dependent on the source and receiver positions, and they strongly influence or cancel some of the poles [13]. Thus, zeros cause erroneous estimation of poles when poles are estimated from a single RTF. Common acoustical poles should therefore be estimated from many RTF's corresponding to different $\mathbf{r}_j$'s.

Common acoustical poles are estimated as common AR coefficients, which are equivalent to the poles, as shown in (4). According to (4), the impulse response of the CAPZ model $\hat{h}(\mathbf{r}_j, k)$ is expressed by

$$\hat{h}(\mathbf{r}_j, k) = -\sum_{i=1}^{P} a_i \hat{h}(\mathbf{r}_j, k - i) + \sum_{i=0}^{Q} b_i(\mathbf{r}_j)\delta(k - i) \quad (5)$$

where

$\hat{h}(\mathbf{r}_j, k)$ modeled impulse response for each $\mathbf{r}_j$,
$a_i$ estimated common AR coefficients,
$b_i(\mathbf{r}_j)$ estimated MA coefficients for each $\mathbf{r}_j$,
$k$ discrete time,
$\delta(k) = 1$ for $k = 0$, and $\delta(k) = 0$ for any other $k$ (unit pulse function).

The error (output error) $\varepsilon_{\text{out}}(\mathbf{r}_j, k)$ between the actual impulse response $h(\mathbf{r}_j, k)$ and the model impulse response $\hat{h}(\mathbf{r}_j, k)$ is defined by

$$\begin{aligned}\varepsilon_{\text{out}}(\mathbf{r}_j, k) &= h(\mathbf{r}_j, k) - \hat{h}(\mathbf{r}_j, k) \\ &= h(\mathbf{r}_j, k) + \sum_{i=1}^{P} a_i \hat{h}(\mathbf{r}_j, k-i) - \sum_{i=0}^{Q} b_i(\mathbf{r}_j)\delta(k-i).\end{aligned} \tag{6}$$

However, finding $a_i$'s and $b_i$'s that minimize the mean-square of the output error $\varepsilon_{\text{out}}(\mathbf{r}_j, k)$ is known to be difficult. In such a situation, the "equation error" $\varepsilon_{\text{eq}}(\mathbf{r}_j, k)$ is introduced [14]:

$$\varepsilon_{\text{eq}}(\mathbf{r}_j, k) = h(\mathbf{r}_j, k) + \sum_{i=1}^{P} a_i h(\mathbf{r}_j, k-i) - \sum_{i=0}^{Q} b_i(\mathbf{r}_j)\delta(k-i). \tag{7}$$

Note that $\hat{h}(\mathbf{r}_j, k-i)$ in (6) is replaced by $h(\mathbf{r}_j, k-i)$ in (7).

Now, the common AR coefficients are estimated as those that minimize the cost function $J_{\text{eq}}$.

$$J_{\text{eq}} = \sum_{j=1}^{M} \sum_{k=0}^{\infty} \varepsilon_{\text{eq}}^2(\mathbf{r}_j, k) \tag{8}$$

where $M$ is the number of impulse responses used for estimation. Let us assume the order of the impulse responses is $N$; in other words, $h(\mathbf{r}_j, k) = 0$ for $k > N$. Then, $\varepsilon_{\text{eq}}(\mathbf{r}_j, k) = 0$ for $k > N + P$ and the upper limit of summation parameter $k$ in (8) can be reduced to $N + P$.

For the cost function $J_{\text{eq}}$ to attain its minimum value, all the partial derivatives of (8) with respect to $a_i$ and $b_i$ $(\mathbf{r}_j)$ must be equal to zero, as shown by

$$\frac{\partial J_{\text{eq}}}{\partial a_i} = 0 \quad (i = 1, 2, \cdots, P), \tag{9a}$$

and

$$\frac{\partial J_{\text{eq}}}{\partial b_i(\mathbf{r}_j)} = 0 \quad (i = 0, 1, 2, \cdots, Q;\ j = 1, 2, \cdots, M). \tag{9b}$$

Here, substituting (7) into (8)

$$\begin{aligned}J_{\text{eq}} &= \sum_{j=1}^{M} \sum_{k=0}^{N+P} \varepsilon_{\text{eq}}^2(\mathbf{r}_j, k) \\ &= \sum_{j=1}^{M} \sum_{k=0}^{N+P} \left( h(\mathbf{r}_j, k) + \sum_{i=1}^{P} a_i h(\mathbf{r}_j, k-i) - \sum_{i=0}^{Q} b_i(\mathbf{r}_j)\delta(k-i) \right)^2.\end{aligned} \tag{10}$$

The partial derivatives of $J_{\text{eq}}$ with respect to $a_i$ and $b_i(\mathbf{r}_j)$ are

$$\begin{aligned}\frac{\partial J_{\text{eq}}}{\partial a_i} &= 2 \sum_{j=1}^{M} \sum_{k=0}^{N+P} h(\mathbf{r}_j, k-i) \\ &\quad \times \left( h(\mathbf{r}_j, k) + \sum_{m=1}^{P} a_m h(\mathbf{r}_j, k-m) - \sum_{m=0}^{Q} b_m(\mathbf{r}_j)\delta(k-m) \right) \\ &\quad (i = 1, 2, \cdots, P),\end{aligned} \tag{11a}$$

$$\begin{aligned}\frac{\partial J_{\text{eq}}}{\partial b_i(\mathbf{r}_j)} &= 2 \sum_{j=1}^{M} \sum_{k=0}^{N+P} \delta(k-i) \\ &\quad \times \left( h(\mathbf{r}_j, k) + \sum_{m=1}^{P} a_m h(\mathbf{r}_j, k-m) - \sum_{m=0}^{Q} b_m(\mathbf{r}_j)\delta(k-m) \right) \\ &\quad (i = 0, 1, 2, \cdots, Q;\ j = 1, 2, \cdots, M).\end{aligned} \tag{11b}$$

Finally, we get the following simultaneous equations:

$$\begin{aligned}&\sum_{j=1}^{M} \sum_{k=0}^{N+P} \left( h(\mathbf{r}_j, k-i) h(\mathbf{r}_j, k) + \sum_{m=1}^{P} a_m h(\mathbf{r}_j, k-i) \right. \\ &\qquad \times h(\mathbf{r}_j, k-m) - \sum_{m=0}^{Q} b_m(\mathbf{r}_j) \\ &\qquad \left. \times h(\mathbf{r}_j, m-i) \right) = 0 \\ &\quad (i = 1, 2, \cdots, P),\end{aligned} \tag{12a}$$

$$\begin{aligned}&h(\mathbf{r}_j, i) + \sum_{m=1}^{P} a_m h(\mathbf{r}_j, i-m) - b_i(\mathbf{r}_j) = 0 \\ &\quad (i = 0, 1, 2, \cdots, Q:\ j = 1, 2, \cdots, M).\end{aligned} \tag{12b}$$

Here, we use the relationship $h(\mathbf{r}_j, k-i)\delta(k-m) = h(\mathbf{r}_j, m-i)$.

We can obtain the common acoustical poles $a_i$ $(i = 1, 2, \cdots, P)$ with least-squares error by solving these simultaneous equations (12a) and (12b). The solution is given in a matrix formula in Appendix A.

Determining the model orders $P$ and $Q$ is not easy. Several methods have been proposed for this problem, such as the AIC criterion [15]. However, here, $P$ and $Q$ are determined as follows. First, the normalized mean-squared output error index $J_{\text{out}}$, which corresponds to the accuracy of the modeling, is defined:

$$J_{\text{out}} = \frac{1}{M} \sum_{j=1}^{M} \left( \frac{\sum_{k=0}^{\infty} \varepsilon_{\text{out}}^2(\mathbf{r}_j, k)}{\sum_{k=0}^{N} h^2(\mathbf{r}_j, k)} \right) \tag{13}$$

and the desired value of $J_{\text{out}}$ is predetermined. Next, a set of $P$ and $Q$ are determined to minimize the sum of $P$ and $Q$ that satisfies the predetermined value of $J_{\text{out}}$.

Notice that the MA order $Q$ could be reduced in (7) at the pole estimation process. In other words, there are cases where

common acoustical poles can be estimated satisfactorily with an MA order that is lower than the model order. Small $Q$ reduces the number of simultaneous equations in (12a) and (12b) and, thus, lightens the computational load for deriving common AR coefficients.

The stability of the estimated poles is an important issue. When poles are estimated with $Q = 0$ in (7), the stability is assured, as is proven in Appendix B. However, the proof of stability for $Q > 0$ is left for future study. As far as we searched, we could not find any unstable estimated poles.

### C. Estimation of Common Acoustical Poles by Averaging Method

When $P$ and $Q$ are large, a large amount of computation is needed to calculate the common AR coefficients that minimize the error index $J_{\text{eq}}$ of (8). In such a situation, a set of AR coefficients $a_i(\mathbf{r}_j)$ that minimizes $J_{\text{eq}}(\mathbf{r}_j)$ ,

$$J_{\text{eq}}(\mathbf{r}_j) = \sum_{k=0}^{\infty} \varepsilon_{\text{eq}}^2(\mathbf{r}_j, k) \tag{14}$$

for each $\mathbf{r}_j$ $(j = 1, \cdots, M)$ is first obtained. Common AR coefficients are then estimated as the averaged values of each set of AR coefficients:

$$a_i = \frac{1}{M} \sum_{j=1}^{M} a_i(\mathbf{r}_j) \qquad (i = 1,\ 2,\ 3, \cdots, P). \tag{15}$$

This estimation method requires far less computation than the method described in Section III-B. Although the modeled RTF's based on the averaged AR coefficients were stable as far as we tested, the theoretical background and stability conditions of this averaging method are left for future study.

## IV. Estimation of Acoustical Poles by the Proposed Methods

We evaluated the estimation methods by comparing estimated poles with the theoretical ones for two cases. In the first case, the model order is the same as the theoretical order of the acoustical poles. The second case is where the model order is much lower than the theoretical order. Because the acoustical poles could be obtained theoretically for a rectangular room [16], the estimation was conducted using the impulse responses of the RTF's simulated by assuming a rectangular perpendicular room (86 $m^3$: 6.7 × 4.3 × 3.0 $m^3$). These impulse responses were computed using the image method [17].

### A. Estimation with the Same Order as the Theoretical Pole Order

In the first case, the acoustical poles were estimated with the same order as the theoretical order of the poles. The impulse responses were simulated under the condition that all the wall reflection coefficients were 0.95. The sampling frequency was 250 Hz, and the frequency range was from 50 to 110 Hz. In such a low frequency range, the order of poles is small and the poles are located sparsely in the complex plane. Therefore, it is easy to compare the theoretical poles and the estimated common acoustical poles. Fig. 6 shows a typical simulated impulse response and its frequency response.

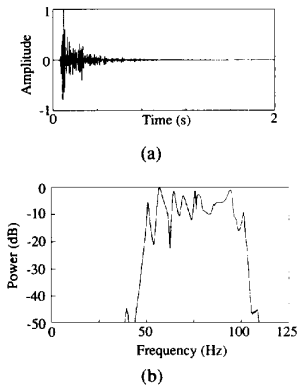


Fig. 6. Example of an impulse response and its frequency response simulated assuming a room size of 6.7 × 4.3 × 3.0 $m^3$ and wall reflection coefficients of 0.95. The frequency range is from 50 to 100 Hz. The sampling frequency is 250 Hz: (a) Simulated impulse response; (b) frequency response.

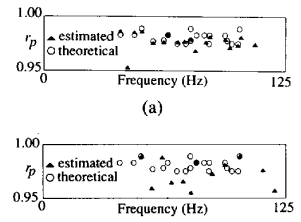


Fig. 7. Comparison of estimated ▲ and theoretical (o) acoustical poles. The order of the poles is the same as the theoretical order for the estimations: (a) Poles estimated from 20 RTF's with different source and receiver positions: (b) poles estimated from a single RTF.

The common acoustical poles were estimated from 20 RTF's with different source and receiver positions using by the least squares method described in Section III-B. Here, the order of poles $P$ was chosen to be 50, which is nearly the same as the theoretical order. The order of zeros was chosen to be 120 to achieve a $J_{\text{out}}$ of $-30$ dB. Note that the all-zero model needed 250 coefficients to achieve a $J_{\text{out}}$ of $-30$ dB. Fig. 7(a) shows the poles estimated from 20 RTF's, and Fig. 7(b) shows the poles estimated from a single RTF. The values of $r_p$ (vertical axis in Fig. 7) represent the absolute values of complex poles. The symbols (▲) indicate the estimated poles. The symbols (o) indicate the theoretical poles which were calculated based on the averaged reverberation decay curve [16]. Fig. 7 indicates that many of the poles estimated from the 20 RTF's fit the theoretical poles better than the poles estimated from a single RTF.

### B. Estimation with a Lower Order than the Theoretical Pole Order

The theoretical order of acoustical poles is proportional to the room volume and to the third power of the frequency. The order of poles is twice the order of modes. The approximate order of modes $n_p$ for a room with a volume $V$ up to the high

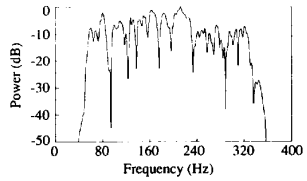


Fig. 8. Example of the frequency response of an impulse response simulated assuming a room size of 6.7 × 4.3 × 3.0 $m^3$ and wall reflection coefficients of 0.85. The frequency range is from 60 to 320 Hz. The sampling frequency is 800 Hz.

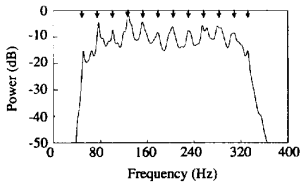


Fig. 9. Frequency response of the common acoustical poles that are estimated with a lower order of poles than the theoretical order. The arrows show the frequencies of the theoretical poles that have high $Q$ factors.

frequency $f_c$ is [12]

$$n_p = \frac{4\pi}{3} V \left(\frac{f_c}{c}\right)^3 \tag{16}$$

where $c$ is the velocity of sound. Therefore, the order of poles is too large for all of them to be estimated when the frequency is high, and the room is large. For example, when the room volume is 86 $m^3$ and the upper frequency is 320 Hz, the order of poles is 1200. For such a situation, we propose to estimate common acoustical poles with a lower order than the theoretical order of poles.

We used 10 impulse responses simulated for different source and receiver positions in a room whose volume was 86 $m^3$. The conditions of the simulation were that all the wall reflection coefficients were 0.85, the sampling frequency was 800 Hz, and the frequency range of the band-pass filter was from 50 to 320 Hz. Fig. 8 shows a typical frequency response of the simulated impulse response. The theoretical order of the poles is 1200 at frequencies below 320 Hz for this room. The estimation of the poles was conducted with pole order $P = 100$ and zero order $Q = 200$ using the least squares method in Section III-B.

The solid line in Fig. 9 shows the frequency response of the transfer function

$$A(z) = \frac{1}{1 + \sum_{i=1}^{P} a_i z^{-i}} \tag{17}$$

where $a_i$ $(i = 1, 2, \cdots, P)$ are estimated common AR coefficients. The peaks of this response correspond to the frequencies of the estimated poles. The arrows in Fig. 9 indicate the frequencies of the theoretical poles that have high $Q$ factors. Fig. 9 shows that the estimated poles correspond well to the theoretical poles that have high $Q$ factors. On the other hand, these theoretical poles are not clear in the single RTF in Fig. 8.

The solid line in Fig. 10 shows the frequency response derived by the averaging method in Section III-C, and again,

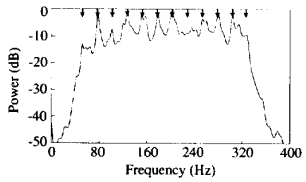


Fig. 10. Frequency response of the averaged AR coefficients. The arrows show the frequencies of the theoretical poles that have high $Q$ factors.

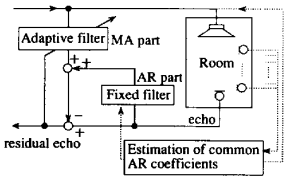


Fig. 11. Acoustic echo canceller that has a fixed filter with estimated common AR coefficients.

the arrows indicate the frequencies of the theoretical poles that have high $Q$ factors. It is easily understood that theoretical poles are also well estimated by the averaging method.

The following summarizes the results for these two cases:

i) Common acoustical poles are better estimated from many RTF's than from a single RTF.
ii) Poles estimated with the same order as the theoretical pole order fit well with the theoretical poles.
iii) Poles estimated with a lower order than the theoretical pole order correspond to the major resonance frequencies that have high $Q$ factors.
iv) The averaging method in Section III-C also estimates theoretical poles well.

## V. Effectiveness of the Proposed Model

The effectiveness of the proposed CAPZ model was evaluated by applying it to an acoustic echo canceller (AEC). AEC's are widely used to cancel echo signals and to prevent acoustic feedback in teleconference systems or in active noise controllers. The AEC reduces the echo signal by subtracting the estimated echo signal. The estimated echo signal is produced by an adaptive filter that models the RTF between a loudspeaker and a microphone.

Fig. 11 shows a block diagram of an AEC based on the proposed CAPZ model. It has a series-parallel-type structure and it has a fixed filter with estimated common AR coefficients and an adaptive filter with variable MA coefficients. The AEC first measures multiple impulse responses for different microphone positions using the dotted right-hand outer loop in the figure. Then, the common AR coefficients are estimated and copied to the fixed filter.

This AEC was simulated in a computer using impulse responses measured with different source and receiver positions in a real room. The room volume was 80 $m^3$, and its reverberation time was 0.6 s. Because the proposed model is especially effective at low frequencies, the frequency range of the impulse responses was set to 60 to 800 Hz. The sampling

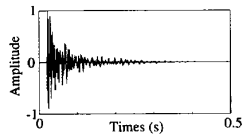


Fig. 12. Example of the measured impulse response in a real room. The sampling frequency is 2 kHz.

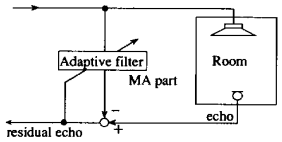


Fig. 13. Acoustic echo canceller based on the conventional all-zero model.

frequency was 2 kHz. Fig. 12 shows an example of the measured impulse response.

The number of coefficients was 250 for the fixed filter and 450 for the adaptive filter. These numbers were chosen to achieve 35 dB of stationary echo return loss enhancement (ERLE), where ERLE is defined as the ratio of echo power to residual echo power:

$$\text{ERLE} = 10\log_{10}\left(\frac{\text{echo power}}{\text{residual echo power}}\right)(\text{dB}). \quad (18)$$

Because the orders of the filters were so large, common AR coefficients were obtained using the averaging method in Section III-C with 10 measured impulse responses. The common AR coefficients were copied to the fixed filter, and the normalized LMS algorithm [18] was used for the adaptive MA filter.

The AEC was evaluated using an impulse response that was not used for estimating the common AR coefficients. The ERLE convergence of the AEC based on the proposed CAPZ model was compared with an AEC based on the conventional all-zero model, which is shown in Fig. 13. The conventional AEC had only an adaptive filter and required 800 coefficients to achieve 35 dB of stationary ERLE. The proposed AEC needed only about half as many adaptive filter coefficients as the conventional AEC because the proposed model reduced the number of the variable parameters (MA coefficients) by using common AR coefficients. Fig. 14 shows the ERLE convergence curve of the two AEC's. Because the smaller number of adaptive filter coefficients gives faster convergence, the proposed AEC converges about 1.5 times faster than the conventional AEC.

Thus, the proposed model reduces the number of variable parameters for RTF variations and makes the adaptive filter converge quickly. These results show the efficiency of the proposed model.

## VI. Conclusion

A new model has been proposed for a room transfer function by using common acoustical poles, which are invariant for the RTF variation due to changes in source or receiver position. We also presented two methods of estimating the common

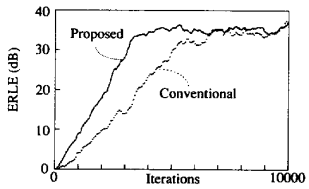


Fig. 14. ERLE of two AEC's. The solid line shows the convergence of an AEC with a fixed AR filter with 250 coefficients and an adaptive MA filter with 450 coefficients based on the proposed model. The dotted line shows the convergence of an AEC with an adaptive MA filter with 800 coefficients based on the conventional all-zero model. The frequencies were from 60 to 800 Hz, and the sampling frequency was 2 kHz.

acoustical poles as common AR coefficients: i) by minimizing the mean-squared equation error for multiple impulse responses and ii) by averaging the AR coefficients derived from each impulse response. The estimated common acoustical poles agree well with the theoretical poles when the estimation is conducted with the same order as the theoretical pole order. When a lower order than the theoretical one is used, the estimated common acoustical poles represent the major resonance properties of the room.

The proposed common-acoustical-pole and zero model requires far fewer parameters that depend on the RTF variation than the conventional models require. Its effectiveness was confirmed in a simulation of an acoustic echo canceller. It required only half as many adaptive filter coefficients up to 800 Hz and converged about 1.5 times faster than an acoustic echo canceller based on the conventional all-zero model.

## Appendix A
## Calculation Method of Common AR Coefficients in a Matrix Formula

Equation (7) for $j = 1, 2, \cdots, M$ and $k = 0, 1, 2, \cdots, N+P$ is rewritten in matrix form as

$$\mathbf{e} = \mathbf{h} - \mathbf{A}\mathbf{x}, \quad (A.1)$$

where $\mathbf{e}$, $\mathbf{h}$, and $\mathbf{x}$ are vectors defined as

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{bmatrix} \quad \mathbf{e}_j = \begin{bmatrix} \varepsilon_{\text{eq}}(\mathbf{r}_j, 0) \\ \varepsilon_{\text{eq}}(\mathbf{r}_j, 1) \\ \vdots \\ \varepsilon_{\text{eq}}(\mathbf{r}_j, N+P) \end{bmatrix} \Bigg\} N+P+1 \quad (A.2)$$

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_M \end{bmatrix} \quad \mathbf{h}_j = \begin{bmatrix} h(\mathbf{r}_j, 0) \\ h(\mathbf{r}_j, 1) \\ \vdots \\ h(\mathbf{r}_j, N) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Bigg\} N+P+1 \quad (A.3)$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b}(\mathbf{r}_1) \\ \mathbf{b}(\mathbf{r}_2) \\ \vdots \\ \mathbf{b}(\mathbf{r}_M) \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_p \end{bmatrix} \quad \mathbf{b}(\mathbf{r}_j) = \begin{bmatrix} b_0(\mathbf{r}_j) \\ b_1(\mathbf{r}_j) \\ \vdots \\ b_Q(\mathbf{r}_j) \end{bmatrix} \quad (A.4)$$

where $\mathbf{A}$ in (A.1) is a $\{M(N+P+1)\} \times \{P+M(Q+1)\}$ matrix defined as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{D} & & \\ \mathbf{A}_2 & & \mathbf{D} & \mathbf{0} \\ \vdots & & & \\ & & \mathbf{0} & \ddots \\ \mathbf{A}_M & & & \mathbf{D} \end{bmatrix} \tag{A.5}$$

and (A6), which appears at the bottom of this page, and

$$\mathbf{D} = \overbrace{\begin{bmatrix} 1 & & & \\ & 1 & \mathbf{0} & \\ & & \cdot & \\ & \mathbf{0} & & 1 \\ 0 & \cdot & \cdot & 0 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & \cdot & \cdot & 0 \end{bmatrix}}^{Q+1} \Bigg\} N+P+1. \tag{A.7}$$

The cost function $J_{\text{eq}}$ of (8) is equal to $\mathbf{e}^T\mathbf{e}$, where $^T$ denotes the transpose, and the vector $\mathbf{x}$ that minimizes $\mathbf{e}^T\mathbf{e}$ is given by

$$x = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{h}. \tag{A.8}$$

This gives the solution of simultaneous equations (12a) and (12b).

## Appendix B
## Stability of Common Poles for $Q = 0$

Equation (A.8) is rewritten as

$$\mathbf{x} = \begin{bmatrix} \sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{A}_j & \mathbf{A}_1^T\mathbf{D} & \mathbf{A}_2^T\mathbf{D} & \cdots & \mathbf{A}_M^T\mathbf{D} \\ \mathbf{D}^T\mathbf{A}_1 & \mathbf{D}^T\mathbf{D} & 0 & \cdots & 0 \\ \mathbf{D}^T\mathbf{A}_2 & 0 & \mathbf{D}^T\mathbf{D} & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{D}^T\mathbf{A}_M & 0 & & & \mathbf{D}^T\mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{h}_j \\ \mathbf{D}^T\mathbf{h}_1 \\ \mathbf{D}^T\mathbf{h}_2 \\ \vdots \\ \mathbf{D}^T\mathbf{h}_M \end{bmatrix} \tag{A.9}$$

Assuming $Q = 0$

$$\mathbf{D} = [1, 0, 0, \cdots, 0]^T. \tag{A.10}$$

Therefore

$$\mathbf{D}^T\mathbf{D} = [1] \tag{A.11}$$

and

$$\mathbf{D}^T\mathbf{A} = [0, 0, \cdots, 0]^T \tag{A.12}$$

(A.9) becomes

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b}(\mathbf{r}_1) \\ \mathbf{b}(\mathbf{r}_2) \\ \vdots \\ \mathbf{b}(\mathbf{r}_M) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{A}_j & & & & \mathbf{0} \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ \mathbf{0} & & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{h}_j \\ \mathbf{D}^T\mathbf{h}_1 \\ \mathbf{D}^T\mathbf{h}_2 \\ \vdots \\ \mathbf{D}^T\mathbf{h}_M \end{bmatrix}$$

$$= \begin{bmatrix} \left(\sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{A}_j\right)^{-1} & & & & \mathbf{0} \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ \mathbf{0} & & & & 1 \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{h}_j \\ \mathbf{D}^T\mathbf{h}_1 \\ \mathbf{D}^T\mathbf{h}_2 \\ \vdots \\ \mathbf{D}^T\mathbf{h}_M \end{bmatrix}. \tag{A.13}$$

Equation (A.13) gives

$$\mathbf{a} = \left(\sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{A}_j\right)^{-1}\left(\sum_{j=1}^{M}\mathbf{A}_j^T\mathbf{h}_j\right). \tag{A.14}$$

Here, $\phi_j(k)$, $k = 1, 2, \cdots, N$ is defined as

$$\phi_j(k) = \sum_{i=0}^{N-k} h(\mathbf{r}_j, i)h(\mathbf{r}_j, i+k). \tag{A.15}$$

---

$$\mathbf{A}_j = \overbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ h(\mathbf{r}_j,0) & 0 & \cdots & 0 \\ h(\mathbf{r}_j,1) & h(\mathbf{r}_j,0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(\mathbf{r}_j,P-1) & h(\mathbf{r}_j,P-2) & \cdots & h(\mathbf{r}_j,0) \\ \vdots & \vdots & \ddots & \vdots \\ h(\mathbf{r}_j,N) & h(\mathbf{r}_j,N-1) & \cdots & h(\mathbf{r}_j,N-P+1) \\ 0 & h(\mathbf{r}_j,N) & \cdots & h(\mathbf{r}_j,N-P+2) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h(\mathbf{r}_j,N) \end{bmatrix}}^{P} \Bigg\} N+P+1 \tag{A.6}$$

Then, the matrix $\mathbf{A}_j^T\mathbf{A}_j$ and the vector $\mathbf{A}_j^T\mathbf{h}_j$ are represented as

$$\mathbf{A}_j^{\mathrm{T}}\mathbf{A}_j = \begin{bmatrix} \phi_j(0) & \phi_j(1) & \cdot & \cdot & \cdot & \phi_j(P-1) \\ \phi_j(1) & \phi_j(0) & & & & \phi_j(P-2) \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \vdots \\ \cdot & & & & \cdot & \phi_j(1) \\ \phi_j(P-1) & \cdot & \cdot & \cdot & \phi_j(1) & \phi_j(0) \end{bmatrix}, \tag{A.16}$$

$$\mathbf{A}_j^{\mathrm{T}}\mathbf{h}_j = \begin{bmatrix} \phi_j(1) \\ \phi_j(2) \\ \cdot \\ \cdot \\ \cdot \\ \phi_j(P) \end{bmatrix}. \tag{A.17}$$

Thus, the matrix $\left(\sum_{j=1}^{M} \mathbf{A}_j^T\mathbf{A}_j\right)$ is a positive definite matrix of Toepliz form. This ensures the stability of the AR model with the coefficients derived by (A.14).

## ACKNOWLEDGMENT

The authors would like to thank to Dr. M. Tohyama and Dr. N. Koizumi for their support and suggestions in this research. The authors also would like to thank Dr. T. Moriya for his valuable comments and Dr. S. Furui for his continuous encouragement.

## REFERENCES

[1] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound &Vib.*, **146**, pp. 281–302, 1991.
[2] N. Koizumi and R. H. Lyon, "On the model order for the identification of acoustic systems," in*Proc. ASA 118th Mtg. J. Acoust. Soc. Amer.*, 86, Suppl. 1, S3, 1989.
[3] Y. Haneda, S. Makino, and N. Koizumi, "ARMA modeling for a room acoustic transfer function at low frequency bands," in *Proc. Spring Mtg. Acoust. Soc. Japan*, Mar. 1990, pp. 439–440 (in Japanese).
[4] G. Long, D. Shwed, and D. D. Falconer, "Study on a pole-zero adaptive echo canceller," *IEEE Trans Circuits Syst.*, vol. CAS-34, no. 7, pp. 765–769, 1987.
[5] O. Muron and J. Sikorav, "Modeling of reverberators and audioconference rooms," in *Proc. ICASSP86*, 1986, pp. 921–924.
[6] S. J. Flockton and S. Gudrangen, "Pole-zero modeling of room transfer functions," presented at the *2nd Int. Workshop Acoust. Echo Canceller*, L'Aquila, Italy, Sept. 1991.
[7] M. Schönle, U. Zölzer, and N. Fliege, "Modeling of room impulse responses by multirate systems," in *Proc. 93rd AES Convention*, Oct. 1992, p. 3447.
[8] Y. Haneda, S. Makino, and Y. Kaneda, "Estimation of poles in room acoustic transfer functions," in *Proc. Spring Mtg. Acoust. Soc. Japan*, Mar. 1991, pp. 393–394 (in Japanese).
[9] J. Mourjopoulos, A. E. Tsopanoglou, and N. D. Fakotakis, "A vector quatizantion approach for room transfer function classification," in *Proc. ICASSP91*, May 1991, pp. 3593–3597.
[10] Y. Haneda, S. Makino, and Y. Kaneda, "Modeling a room transfer function using common acoustical poles," in *Proc. ICASSP92*, Mar. 1992, pp. 289–292, vol. 2.
[11] I. Takumi, M. Hata, Y. Itoh and M. Kobayashi, "Identification with acoustic unknown system using fast Kalman algorithm," IEICE Techn. Rep., CAS85-187, pp. 53–60, 1985 (in Japanese).
[12] H. Kuttruf, *Room Acoustics*. London: Elsevier, 1991.
[13] R. H. Lyon, *Machinery Noise and Diagnostics*. Stoneham, MA: Butterworth, 1987.
[14] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: M.I.T. Press, 1983.
[15] H. Akaike, "A new look on the statical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, 1974.
[16] M. Tohyama and S. Yoshikawa, "Approximate formula of the averaged sound energy decay curve in a rectangular reverberant room," *J. Acoust. Soc. Amer.*, vol. 70, pp. 1674–1678, 1981.
[17] J. B. Allen and D. A. Berkly, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.
[18] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1991, 2nd ed.

**Yoichi Haneda** (A'92) was born in Sendai, Japan, on June 17, 1964. He received the B.S. and M.S. degrees in physics from Tohoku University, Sendai, Japan, in 1987 and 1989, respectively.

Since joining Nippon Telegraph and Telephone Corporation (NTT) in 1989, he has been investigating acoustic signal processing and acoustic echo cancellers. He is now a Research Engineer at the Speech and Acoustics Laboratory of the NTT Human Interface Laboratories.

Mr. Haneda is a member of the Acoustical Society of Japan and the Institute of Electronics, Information, and Communication Engineers of Japan.

**Shoji Makino** (A'89-M'90) was born in Nikko, Japan, on June 4, 1956. He received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He joined the Electrical Communication Laboratory of Nippon Telegraph and Telephone Corporation (NTT) in 1981. Since then, he has been engaged in research on electroacoustic transducers and acoustic echo cancellers. He is now a Senior Research Engineer at the Speech and Acoustics Laboratory of the NTT Human Interface Laboratories. His research interests include acoustic signal processing and adaptive filtering and its applications.

Dr. Makino is a member of the Acoustical Society of Japan and the Institute of Electronics, Information, and Communication Engineers of Japan.

**Yutaka Kaneda** (M'80) was born in Osaka, Japan, on February 20, 1951. He received the B.E., M.E. and Doctor of Engineering degrees from Nagoya University, Nagoya, Japan, in 1975, 1977 and 1990, respectively.

In 1977, he joined the Electrical Communication Laboratory of Nippon Telegraph and Telephone Corporation (NTT), Musashino, Tokyo, Japan. He has since been engaged in research on acoustic signal processing. He is now a Senior Research Engineer at the Speech and Acoustics Laboratory of the NTT Human Interface Laboratories. His recent research interests include microphone array processing, adaptive filtering, and sound field control.

Dr. Kaneda received the IEEE ASSP Senior Award in 1990 for an article on inverse filtering of room acoustics; he also received paper awards from the Acoustical Society of Japan in 1990 and 1992 for articles on adaptive microphone arrays and active noise control. He is a member of the Acoustical Society of Japan, the Acoustical Society of America, and the Institute of Electronics, Information, and Communication Engineers of Japan.