MODELING OF A ROOM TRANSFER FUNCTION USING COMMON ACOUSTICAL POLES

Yoichi HANEDA, Shoji MAKINO, and Yutaka KANEDA

NTT Human Interface Laboratories,
3-9-11, Midoricho, Musashino-shi, Tokyo, 180 Japan

Abstract

A new method is proposed for modeling a room transfer function (RTF) by using estimated common acoustical poles that correspond to resonance properties of a room. These poles are estimated as common values of the multiple RTFs corresponding to different source and receiver positions. This 'common-acoustical-pole and zero' (CAPZ) model requires far fewer variable parameters to represent RTFs than conventional all-zero or pole/zero models. This model was applied to an acoustic echo canceller and to head-related transfer functions. At low frequencies, the acoustic echo canceller based on this model converges 1.5 times faster than the one based on the all-zero model. Head-related transfer functions that have resonance characteristics of the external ear are also successfully modeled by the proposed model.

1. Introduction

A room transfer function (RTF) expresses the transmission characteristics of a sound between a source and a receiver. Modeling an RTF is a key technique for many applications, such as acoustic echo cancellers (AECs) and sound field controllers.

An all-zero model (MA model) or pole/zero model (ARMA model) is usually used for modeling an RTF, but they require a large number of parameters which change in a complex manner when the position of the source or receiver changes. These shortcomings slow the convergence of an AEC based on the conventional models.

A possible solution to this problem would be to estimate parameters that remain constant despite RTF variations, and use them when modeling an RTF [1-3]. Figure 1 shows the idea of this solution. The conventional all-zero or pole/zero models require different sets of parameters D_i to represent different RTFs H_i (i=1,2,3,...,M). In the proposed model, common parameters $1/A$ are first estimated from multiple RTFs. Then the RTFs are represented by $1/A$ and by different sets of parameters B_i (i=1,2,3,...,M), which have far fewer parameters than the sets D_i of the conventional models.

This paper proposes a new modeling method that uses estimated acoustical poles of a room as common constant parameters. The proposed model is applied to two applications, an AEC and the modeling of a head-related transfer function. Computer simulations demonstrate the advantage of the proposed model.

2. Common-Acoustical-Pole and Zero Model

Acoustical poles are common to all RTFs, observed with different source and receiver positions in a room [4]. These poles correspond to the resonance frequencies (eigenfrequencies) and their Q-factors. They depend on the acoustical conditions of a room, such as its shape and absorption coefficients, but they do not depend on the positions of the source and receiver .

The proposed common-acoustical-pole and zero (CAPZ) model is thus represented in pole/zero form and

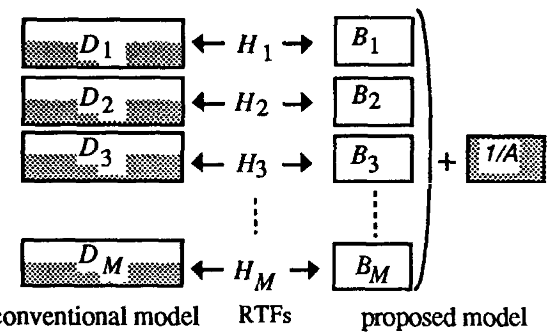


Fig. 1 Modeling of RTFs corresponding to different positions of the source and receiver. The conventional all-zero or pole/zero models require different sets of parameters D_i to represent different RTFs H_i (i=1,2,3,...,M). The proposed model reduces the number of parameters in the sets B_i by extracting parameters $1/A$ that are common to H_1-H_M.

ARMA form by z-transform:

$$G(\mathbf{R},z) = \frac{C\ z^{-Q_1}\prod_{i=1}^{Q_2}(1-q_i(\mathbf{R})\,z^{-1})}{\prod_{i=1}^{P}(1-p_i\,z^{-1})} = \frac{\sum_{i=0}^{Q} b_i(\mathbf{R})z^{-i}}{1+\sum_{i=1}^{P} a_i z^{-i}}, \quad (1)$$

$\boldsymbol{R}$: positions of a source and a receiver,
$G(\boldsymbol{R},z)$: CAPZ model of an RTF corresponding to each $\boldsymbol{R}$,
p_i: common acoustical pole,
$q_i(\boldsymbol{R})$: zero for each $\boldsymbol{R}$,
P: order of poles,
Q, Q_1, Q_2: order of zeros, ($Q = Q_1 + Q_2$),
C: constant,
a_i: common AR coefficient,
$b_i(\boldsymbol{R})$: MA coefficient for each $\boldsymbol{R}$.

With the proposed model, once the common acoustical poles are estimated, only zeros are estimated for different RTFs. With the conventional pole/zero model, on the other hand, both poles and zeros are always estimated for each RTF. Estimated poles in the conventional pole/zero model are usually different for different RTFs for the following reason.

3. Estimation of common acoustical poles

All acoustical poles cannot necessarily be observed in a single RTF, even though all the acoustical poles are common to all the RTFs in a room. This is because zeros that depend on the source and receiver positions influence or cancel some poles [3], thus causing erroneous estimation of poles. Common acoustical poles should therefore be estimated from multiple RTFs corresponding to different $\boldsymbol{R}$s.

Common acoustical poles are estimated as common AR coefficients, which are equivalent to the poles as shown in Eq. (1). According to Eq. (1), the impulse response of the CAPZ model for $\boldsymbol{R}_L$, $g(\boldsymbol{R}_L, k)$, is expressed as

$$g(\mathbf{R}_L, k) = -\sum_{i=1}^{P} a_i\, g(\mathbf{R}_L, k-i) + \sum_{i=0}^{Q} b_i(\mathbf{R}_L)\, \delta(k-i), \quad (2)$$

L: index for the variation of $\boldsymbol{R}$,
k: discrete time,
δ: unit pulse function.

The error (output error) between the actual impulse response $h(\boldsymbol{R}_L, k)$ and the model impulse response $g(\boldsymbol{R}_L, k)$ is represented as

$$\begin{aligned}\varepsilon_{out}(\mathbf{R}_L,k) &= h(\mathbf{R}_L, k) - g(\mathbf{R}_L, k)\\ &= h(\mathbf{R}_L, k) + \sum_{i=1}^{P} a_i\, g(\mathbf{R}_L, k-i) - \sum_{i=0}^{Q} b_i(\mathbf{R}_L)\, \delta(k-i)\ . \quad (3)\end{aligned}$$

Finding a_is and b_is that minimize output error $\varepsilon_{out}(\boldsymbol{R}_L, k)$ is, however, known to be difficult [5]. Therefore, the "equation error" $\varepsilon_{eq}(\boldsymbol{R}_L, k)$ is introduced:

$$\varepsilon_{eq}(\mathbf{R}_L,k) = h(\mathbf{R}_L, k) + \sum_{i=1}^{P} a_i\, h(\mathbf{R}_L, k-i) - \sum_{i=0}^{Q} b_i(\mathbf{R}_L)\, \delta(k-i). \quad (4)$$

Note that the difference between Eq. (3) and Eq. (4) is in the second term of the right-hand side, where $g(\boldsymbol{R}_L, k)$ is replaced by $h(\mathbf{R}_L, k)$.

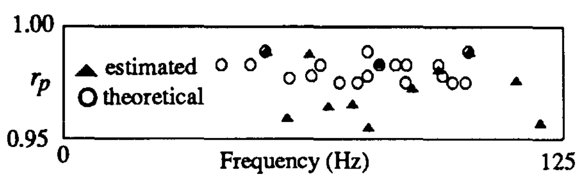
Now, the common AR coefficients are estimated as those that minimize the following error index I.

$$I = \sum_{k=0}^{\infty} \sum_{L=1}^{M} [\,\varepsilon_{eq}(\mathbf{R}_L,k)\,]^2\ , \quad (5)$$

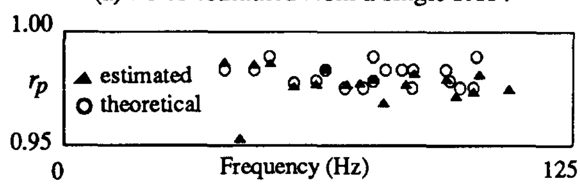
M : number of impulse responses.

The orders of P and Q in Eq. (4) are chosen to minimize the sum of P and Q for the predetermined values of sufficiently small I.

Figures 2(a) and 2(b) show one example of the poles estimated from a single RTF and from 30 RTFs. Because acoustical poles could be obtained theoretically for a rectangular room, RTFs were simulated by assuming a rectangular room (6.7 × 4.3 × 3.0 m, wall reflection coefficient 95 %) and using the image method. The symbols (▲) indicate estimated poles and the symbols (o) indicate the theoretical acoustical poles. The r_p



(a) Poles estimated from a single RTF.



(b) Poles estimated from 30 RTFs with different positions of the source and receiver.

Fig. 2 Comparison of estimated (▲) and theoretical (o) acoustical poles.

(vertical axis in Fig.2) represents the absolute values of complex poles. The poles estimated from a single RTF do not fit the theoretical poles (Fig.2(a)). On the other hand, many acoustical poles are well estimated from multiple RTFs (Fig.2(b)). Summarizing this result: 1) common acoustical poles are better estimated from multiple RTFs than from a single RTF, 2) the proposed method is effective for estimating common acoustical poles.

When P and Q are large, however, a lot of computational power is needed to calculate the AR coefficients that minimize error index I of Eq. (5). In such a situation, a set of AR coefficients that minimize $\varepsilon_{eq}(R_L, k)$ of Eq. (4) is first obtained for each RTF. Common AR coefficients are then obtained by averaging each set of AR coefficients:

$$a_i = \frac{1}{M}\sum_{L=0}^{M} a_i(\mathbf{R}_L) \quad (i = 1,2,3,..,P) . \qquad (6)$$

The theoretical background of these averaged AR coefficients is left for future study, but the estimated poles fit the theoretical poles as well as they do when estimated by minimizing the error index I.

4. Performance of the proposed model

The performance of the proposed model was evaluated by using it in two applications, the modeling of an RTF for an acoustic echo canceller and the modeling of a head-related transfer function. Both transfer functions can be modeled by the CAPZ model.

4.1. Acoustic echo canceller

We used the proposed model in a series-parallel-type acoustic echo canceller (AEC) that had a fixed filter with estimated common AR coefficients and an adaptive filter with variable MA coefficients (Fig.3). The block diagram in Fig.3 was simulated in a computer by using measured impulse responses.

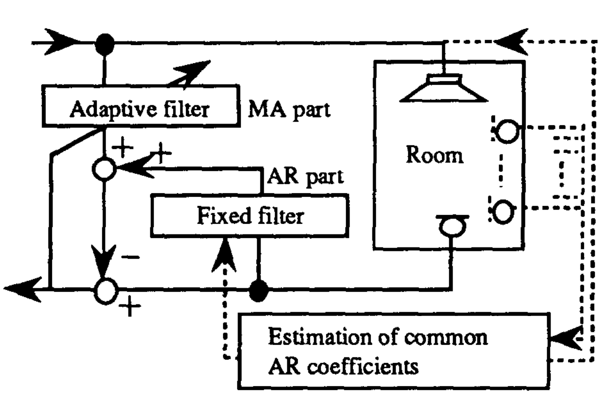


Fig. 3 Acoustic echo canceller that has a fixed filter with estimated common AR coefficients.

Because the proposed model for RTFs is especially effective at low frequencies, its performance was evaluated in the frequency band from 60 to 800 Hz. The sampling frequency was 2 kHz. Impulse responses of different source and receiver positions were measured in a room. The room volume was 80 m^3 and its reverberation time was 0.6 s.

The number of coefficients was 250 for the fixed filter, and 450 for an adaptive filter. These numbers were chosen to provide 35 dB of stationary echo return loss enhancement (ERLE), where ERLE is defined as the ratio of echo power to residual echo power. Because the orders of the filters are so large, the common AR coefficients were obtained using Eq.(6) with ten measured impulse responses. The common AR coefficients were copied to the fixed filter and the normalized LMS algorithm was used for the adaptive filter.

An AEC based on the conventional all-zero model with only an adaptive filter was also evaluated and found to require 800 coefficients to achieve 35 dB of stationary ERLE. The proposed AEC needs only about half as many adaptive filter coefficients as the conventional AEC needs.

The performances of both AECs were evaluated by simulating the convergence speed of the ERLE. Because the convergence becomes faster as the number of adaptive filter coefficients becomes smaller, the proposed AEC converges about 1.5 times faster than the conventional AEC (Fig. 4).

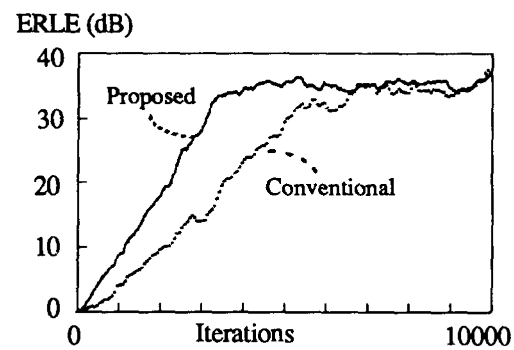


Fig. 4 ERLE (echo return loss enhancement) of two AECs. Solid line: proposed AEC with a fixed filter (AR part: 250 coefficients) and an adaptive filter (MA part: 450 coefficients). Dashed line: conventional all-zero AEC with an adaptive filter (MA part: 800 coefficients).

4.2. Head-related transfer function

A head-related transfer function (HRTF) represents the transmission characteristics from a sound-source to the ear of a listener. HRTF changes when the source direction changes. Humans localize a sound-source based on this difference in HRTFs for different directions. The modeling of an HRTF is important for such applications as the reproduction of a virtual sound image. The model with fewer parameters have been investigated [6].

HRTFs have a resonance system composed of the pinna and ear canal [7]. This resonance system can be considered as the common acoustical poles in all the HRTFs corresponding to different sound directions. The proposed model can thus be applied, by considering that the parameter $\boldsymbol{R}$ in Eqs. (1)-(6) corresponds to the source direction θ.

Modeling of HRTFs was evaluated by calculating the following modeling error $E(\theta)$:

$$E(\theta) = \sum_{k=0}^{N} \left(\varepsilon_{out}(\theta,k)\right)^2 / \sum_{k=0}^{N} \left(h(\theta,k)\right)^2, \qquad (7)$$

$\varepsilon_{out}(\theta, k)$: output error,
$h(\theta, k)$: measured impulse response for sound direction θ,
N: impulse response length.

The output error $\varepsilon_{out}(\theta,k)$ here is defined by Eq. (3), where θ is replaced by R_L.

Eighteen HRTFs with different source directions in the horizontal plane were measured in an anechoic room. The frequency band was from 100 Hz to 18 kHz. The sampling frequency was 45 kHz. The number of common AR coefficients, 20, and the number of different MA coefficients, 40, were predetermined. The common AR coefficients were then estimated with twelve of the eighteen measured HRTFs by minimizing error index I of Eq. (5).

Another six HRTFs were modeled using the CAPZ model with the 20 common AR coefficients estimated above, and 40 variable MA coefficients estimated for each HRTF. They were compared to the conventional all-zero model with 60 variable MA coefficients. To represent HRTFs with different source directions, the proposed model of an HRTF was given 2/3 as many variable parameters that depend on source directions as the all-zero model had.

The results of modeling errors of both models are shown in Fig. 5. Although using a small number of variable parameters, the modeling errors of the proposed model are as good as those of the conventional all-zero model. These results confirm that the proposed model with common acoustical poles successfully represents the HRTFs.

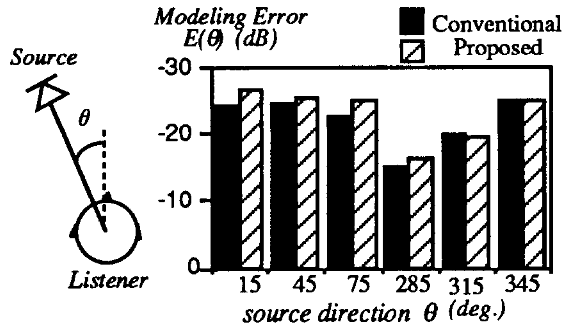


Fig. 5 The modeling errors of the conventional all-zero model (60 MA coefficients)(black bars) and the proposed common acoustical pole and zero model (20 common AR coefficients; 40 variable MA coefficients)(shaded bars) for head-related transfer functions. The common AR coefficients are independent of the source direction θ.

5. Conclusions

We have modeled a room transfer function by using common acoustical poles, which are constant when the position of source or receiver changes. These poles are estimated as common AR coefficients either by minimizing an equation error for multiple impulse responses or by averaging the AR coefficients derived from each impulse response.

The proposed common-acoustical-pole and zero model requires fewer parameters that depend on the source and receiver positions than the conventional models require. An acoustic echo canceller based on the proposed model requires half as many adaptive filter coefficients and converges about 1.5 times faster than an acoustic echo canceller based on the conventional all-zero model. For modeling head-related transfer functions, which are important for binaural cues of sound localization, the proposed model reduces by 1/3 the number of parameters that depend on the source directions.

References

[1] Y. Haneda, S. Makino, Y. Kaneda, "Estimation of Poles in Room Acoustic Transfer Functions," Proc. Spring meeting of Acoust. Soc. Jpn., p. 393, March 1991.
[2] J. Mourjopoulos, A. E. Tsopanoglou, N. D. Fakotakis, "A Vector Quantization Approach for Room Transfer Function Classification," ICASSP91, p.3593, May 1991.
[3] N. Koizumi, R. H. Lyon, "On the model order for the identification of acoustic systems," ASA 118th meeting, J.A.S.A, 86, Suppl. 1, S3, 1989.
[4] For example, H. Kuttruf, " Room Acoustics," Elsevier Applied Science, London, 1991.
[5] L. Ljung and T. Söderström, "Theory and Practice of Recursive Identification," MIT Press, Cambridge, 1983.
[6] F. Asano, Y. Suzuki, S. Kono, T. Sone, "Simulation and Modification of Head-related Transfer Function Using ARMA model," Proc. Autumn meeting of Acoust. Soc. Jpn., p.367, October 1987.
[7] J. Blauert, "Spatial Hearing," MIT Press, Cambridge, 1983.