Common-Acoustical-Pole and Zero Modeling of Head-Related Transfer Functions

Yoichi Haneda, *Member, IEEE,* Shoji Makino, *Member, IEEE,*
Yutaka Kaneda, *Member, IEEE,* and Nobuhiko Kitawaki, *Member, IEEE*

***Abstract*—Use of a common-acoustical-pole and zero model is proposed for modeling head-related transfer functions (HRTF's) for various directions of sound incidence. The HRTF's are expressed using the common acoustical poles, which do not depend on the source directions, and the zeros, which do. The common acoustical poles are estimated as they are common to HRTF's for various source directions; the estimated values of the poles agree well with the resonance frequencies of the ear canal. Because this model uses only the zeros to express the HRTF variations due to changes in source direction, it requires fewer parameters (the order of the zeros) that depend on the source direction than do the conventional all-zero or pole/zero models. Furthermore, the proposed model can extract the zeros that are missed in the conventional models because of pole-zero cancellation. As a result, the directional dependence of the zeros can be traced well. Analysis of the zeros for HRTF's on the horizontal plane showed that the nonminimum-phase zero variation was well formulated using a simple pinna-reflection model. The *common-acoustical-pole and zero* (CAPZ) model is thus effective for modeling and analyzing HRTF's.**

***Index Terms*— Head-related transfer function, modeling, nonminimum-phase zeros, resonance frequency, poles and zeros.**

I. Introduction

A HEAD-RELATED transfer function (HRTF) describes the sound transmission characteristics from a source to a point in an ear canal in a free field. Because HRTF's contain the information a person uses to determine a source location [1], [2], modeling them is important in many applications. For example, models of HRTF's are applied to room acoustics simulation and to three-dimensional (3-D) sound image localization in virtual reality technologies [3]–[6].

An all-zero model is a typical HRTF model. (Strictly speaking, the nth order all-zero model has implicitly n poles at $z = 0$). It simply uses the impulse response of the measured HRTF as its coefficients and can be constructed using a finite impulse response (FIR) filter. When using the all-zero model, however, many parameters (filter coefficients) are needed to simulate HRTF's corresponding to the various source directions. This is because a low-order all-zero model (low-order FIR filter) is not able to represent steep frequency structures. Furthermore, the impulse response of an HRTF strongly depends on the source direction. So, many FIR filter coefficients for every source direction must be utilized.

To reduce the number of parameters, pole/zero modeling of an HRTF has been studied [7]–[9]. The pole/zero model can be constructed using an infinite impulse response (IIR) filter. An IIR filter uses more instructions when implemented on digital signal processors (DSP's) and is more sensitive to numerical errors in the fixed-point arithmetic than a same-order FIR filter. Nevertheless, the model is more efficient than the all-zero model because the poles can represent the long impulse response caused by resonances with fewer parameters. In conventional pole/zero modeling, however, because both the poles and the zeros are estimated for every source direction of the HRTF, the estimated poles depend on the source direction, even though the physical poles of the HRTF do not. Therefore, a different set of poles and zeros is used to represent the HRTF for each different source direction.

We recently proposed the *common-acoustical-pole and zero* (CAPZ) model for acoustic transfer functions [10]. This model expresses the acoustic transfer function by using the common acoustical poles, which are independent of the source and receiver positions, and the zeros, which depend on those positions. An HRTF represented by the CAPZ model has two parts: a direction-independent part (common acoustical poles) and a direction-dependent part (zeros) [11]. The common acoustical poles are estimated as values common to the measured HRTF's for the various source directions; they correspond to the physical resonance system of an ear canal.

Because the CAPZ model represents the directional dependence of the HRTF using only the zero variations, it requires fewer parameters than the conventional all-zero model or the pole/zero model. Furthermore, it can extract even the zeros that are missed in conventional models due to pole-zero cancellation. This means that the CAPZ model can trace well the zero variations due to changes in the source direction and thus could offer a new characterization method for HRTF variations due to change in the source direction.

This paper is organized as follows. In Section II, we describe the principle of the CAPZ model for HRTF's. In Section III, we demonstrate the efficiency of the proposed HRTF modeling compared with conventional modeling for HRTF's on the horizontal plane. In Section IV, we discuss an example of the directional dependence of the zeros in our proposed model.

Manuscript received June 11, 1997; revised March 20, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan.

Y. Haneda, S. Makino, and Y. Kaneda are with NTT Human Interface Laboratories, Tokyo 180-8585, Japan (e-mail: haneda@splab.hil.ntt.co.jp).

N. Kitawaki is with the Institute of Information Science and Electronics, University of Tsukuba, Ibaraki 305-0006, Japan.

Publisher Item Identifier S 1063-6676(99)01632-6.

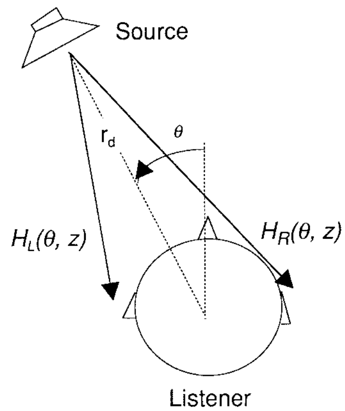

Fig. 1. Head-related transfer functions between a source and a listener's ears.

II. Principle of Common-Acoustical-Pole and Zero (CAPZ) Model

A. Common-Acoustical-Pole and Zero Modeling of HRTF's

An HRTF contains a resonance system composed of an ear canal [12]. The resonance frequencies and Q factors of this system can be assumed to be independent of the source directions [11]. The poles of the HRTF represent these resonance frequencies and Q factors, and are therefore independent of the source directions. Based on this, the HRTF can be efficiently modeled by using the poles that are independent of the source directions. We call these directional-independent poles *common acoustical poles* because they are commonly included in the HRTF for any source direction, and we call the model the CAPZ model [10].

We shall consider the modeling of the HRTF's for various source directions in the horizontal plane, as shown in Fig. 1. Functions $H_L(\theta, z)$ and $H_R(\theta, z)$ represent the HRTF's of the left and right ears. Source direction θ is set to 0° for the front direction. In this paper, we consider only the left ear's HRTF on the horizontal plane, and represent it as $H(\theta, z)$. The same method is also applied to the right ear.

When the HRTF $H(\theta, z)$ is modeled using the CAPZ model, the poles are independent of the source direction θ, but the zeros depend on it. Therefore, the CAPZ model of HRTF $H(\theta, z)$ is expressed as

$$H_{\mathrm{C}}(\theta, z) = \frac{B_{\mathrm{C}}(\theta, z)}{A_{\mathrm{C}}(z)} = \frac{C z^{-Q_1} \prod_{i=1}^{Q_2} [1 - q_i(\theta) z^{-1}]}{\prod_{i=1}^{P} (1 - p_{\mathrm{C}i} z^{-1})} = \frac{\sum_{i=0}^{Q} b_i(\theta) z^{-i}}{1 - \sum_{i=1}^{P} a_{\mathrm{C}i} z^{-i}} \tag{1}$$

where $p_{\mathrm{C}i}$ denotes the common acoustical poles independent of source direction θ, and $q_i(\theta)$ denotes the zeros dependent on source direction θ. The P and Q $(Q = Q_1 + Q_2)$ are, respectively, the orders of the poles and of the zeros. The coefficients $a_{\mathrm{C}i}$ $(i = 1, \cdots, P)$ denote the common autoregressive (AR) coefficients corresponding to the common acoustical poles $p_{\mathrm{C}i}$, and the coefficients $b_i(\theta)$ $(i = 0, \cdots, Q)$ denote the moving-average (MA) coefficients corresponding to the zeros $q_i(\theta)$.

Fig. 2 compares the conventional all-zero model, the conventional pole/zero model, and the proposed CAPZ model, for modeling the multiple HRTF's corresponding to M source directions. Because all the parameters of the all-zero model depend on the source direction θ, a completely different set of parameters is required for each source direction. That is, $B_{\mathrm{Z}}(\theta_1, z), B_{\mathrm{Z}}(\theta_2, z), \cdots, B_{\mathrm{Z}}(\theta_M, z)$ in Fig. 2 are completely different. Therefore, the all-zero model requires many parameters to represent HRTF's for various source directions.

The conventional pole/zero model divides an HRTF $H(\theta, z)$ into its pole function $A_{\mathrm{P}}(\theta, z)$ and its zero function $B_{\mathrm{P}}(\theta, z)$. Both the pole $A_{\mathrm{P}}(\theta, z)$ and the zero functions $B_{\mathrm{P}}(\theta, z)$ are estimated for each source direction of the HRTF $H(\theta, z)$. As a result, although the poles in the HRTF are physically invariant for source direction θ, the estimated poles $A_{\mathrm{P}}(\theta_m, z)$ $(m = 1, 2, \cdots, M)$ vary with the source direction θ_m because of interference from the direction-dependent zeros. Therefore, both poles in $A_{\mathrm{P}}(\theta_m, z)$ and zeros in $B_{\mathrm{P}}(\theta_m, z)$ have to be kept for all source directions (θ_1 to θ_M) to represent M-source direction HRTF's.

In contrast, the proposed CAPZ model represents HRTF $H(\theta, z)$ by using the common-acoustical-pole function $A_{\mathrm{C}}(z)$, which is independent of the source direction, and by the zero function $B_{\mathrm{C}}(\theta, z)$, which depends on the source direction. Because the CAPZ model expresses the source directional dependence of the HRTF by using only the zero variations, the number of parameters that depend on the source direction is reduced. Another remarkable feature of the CAPZ model is that it can extract the zeros missing due to pole-zero cancellation.

B. Estimation Method of Model Parameters

The common acoustical poles are physically included in the HRTF for any source direction. They cannot, however, be estimated using a single HRTF measured for an arbitrary source direction because they are usually strongly affected or canceled by the direction-dependent zeros. That is, the poles estimated using a single measured HRTF cannot be regarded as common poles. Therefore, we estimate the common acoustical poles by using an entire set of HRTF's measured for different source directions.

Practically, the common acoustical poles are estimated as the common AR coefficients $a_{\mathrm{C}i}$ using HRTF's for several source directions. Equation (1) can be modified as

$$H_{\mathrm{C}}(\theta, z) \left(1 - \sum_{i=1}^{P} a_{\mathrm{C}i} z^{-i}\right) = \sum_{i=0}^{Q} b_i(\theta) z^{-i}. \tag{2}$$

Taking the inverse z-transform of (2), the impulse response $h_{\mathrm{C}}(\theta, k)$ of the CAPZ model $H_{\mathrm{C}}(\theta, z)$ can be described in the

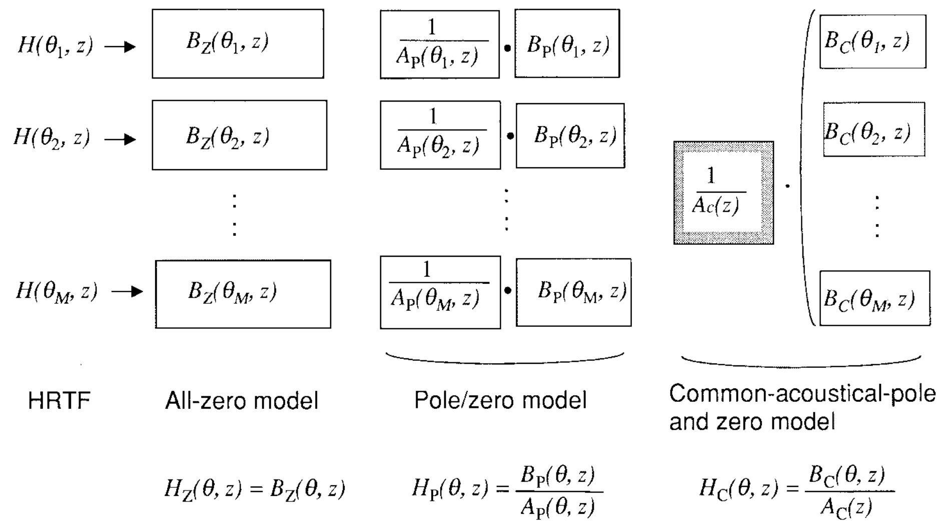

Fig. 2. Modeling of multiple HRTF's corresponding to different source directions, $\theta_1, \theta_2, \cdots, \theta_M$, using an all-zero model, a pole/zero model, and the proposed common-acoustical-pole and zero model.

time domain as

$$h_{\mathrm{C}}(\theta,k)=\sum_{i=1}^{P} a_{\mathrm{C}i}h_{\mathrm{C}}(\theta,k-i)+\sum_{i=0}^{Q} b_i(\theta)\delta(k-i) \quad (3)$$

where $\delta(k)$ is the unit pulse function.

The output error $e_{\mathrm{out}}(\theta,k)$ between the impulse response $h(\theta,k)$ of the original (measured) HRTF $H(\theta,z)$ and the impulse response $h_{\mathrm{C}}(\theta,k)$ of the CAPZ model is defined by

$$\begin{aligned} e(\theta,k) &= h(\theta,k)-h_{\mathrm{C}}(\theta,k) \\ &= h(\theta,k)-\sum_{i=1}^{P} a_{\mathrm{C}i}h_{\mathrm{C}}(\theta,k-i) \\ &\quad -\sum_{i=0}^{Q} b_i(\theta)\delta(k-i). \end{aligned} \quad (4)$$

However, finding values of $a_{\mathrm{C}i}$ and b_i that minimize the mean-square of the output error $e_{\mathrm{out}}(\theta,k)$ is known to be difficult [13], so we use equation error $e_{\mathrm{eq}}(\theta,k)$:

$$e_{\mathrm{eq}}(\theta,k)=h(\theta,k)-\sum_{i=1}^{P} a_{\mathrm{C}i}h(\theta,k-i)-\sum_{i=0}^{Q} b_i(\theta)\delta(k-i). \quad (5)$$

The common AR coefficients are determined so as to minimize cost function J_{eq}, which is defined as the squared sum for time index k, and source direction index m:

$$J_{\mathrm{eq}}=\sum_{m=0}^{M}\sum_{k=0}^{N+P} e_{\mathrm{eq}}^2(\theta_m,k), \quad (6)$$

where M is the number of HRTF's and N is the length of the original impulse response $h(\theta,k)$.

The coefficients that minimize the cost function J_{eq} using the least-squares method can be represented in vector form [10]:

$$\mathrm{x}=(\mathrm{A}^T\mathrm{A})^{-1}\boldsymbol{A}^T\mathrm{h}_a, \quad (7)$$

where we have the formulation shown at the bottom of the next page, and

$$\mathrm{D}=\left.\underbrace{\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \\ 0 & & 0 \\ & \ddots & \\ 0 & & 0 \end{bmatrix}}_{Q+1}\right\} N+P.$$

The orders P and Q in the CAPZ model are determined as follows. First, output error index J_{out}, which is the average output error energy normalized by the impulse response energy, and which corresponds to the accuracy of the modeling, is defined as

$$J_{\mathrm{out}}=\frac{1}{M}\sum_{m=1}^{M}\frac{\sum_{k=0}^{N} e_{\mathrm{out}}^2(\theta_m,k)}{\sum_{k=0}^{N} h^2(\theta_m,k)}. \quad (8)$$

Then, the desired value of J_{out} is predetermined. Next, a P and Q pair is determined so as to minimize the sum of P and Q that satisfies the predetermined value of J_{out} [10].

III. Performance of CAPZ Modeling of HRTF'S

A. Measurement of the HRTF's

We measured the impulse responses of the HRTF's corresponding to different source directions on the horizontal plane by using a dummy head (B&K: Type 4128) in an anechoic room. The distance between the source and the dummy head (r_{d} in Fig. 1) was 1.5 m. The frequency band was 100 Hz to 20 kHz, and the sampling frequency was 48 kHz. A small one-way loudspeaker was used to measure the HRTFs; the impulse response of the loudspeaker was measured using an omnidirectional microphone in the same anechoic room. The coefficients of the inverse filter of the loudspeaker response were calculated to compensate for the loudspeaker response. The measured HRTF's were equalized using this inverse filter. Although HRTF modeling for other planes, e.g., the vertical plane, is also interesting, we consider only the horizontal plane for reasons of brevity in this paper.

Fig. 3 shows the frequency responses of the measured HRTF's of the left ear from 0–180° at every 10°. The peaks at 2.8, 9.0, and 12.2 kHz, for example, appear to be common to all HRTF's, so they can be estimated as common acoustical poles.

B. Estimation of Common Acoustical Poles

The pole and zero orders $P = 20$ and $Q = 40$ in the CAPZ model were chosen to achieve a J_{out} of −20 dB. Before modeling the HRTF's, the propagation delay of each HRTF was removed. The length N of the impulse responses of the original HRTF's were set to 128. The common acoustical poles were estimated as the common AR coefficients a_{Ci} by using the 12 HRTF's corresponding to the source directions at every 30° from 0° to 330°. The common acoustical poles can be estimated from such a relatively small set of the HRTF's because the poles canceled by the zeros are different when the source directions are largely different.

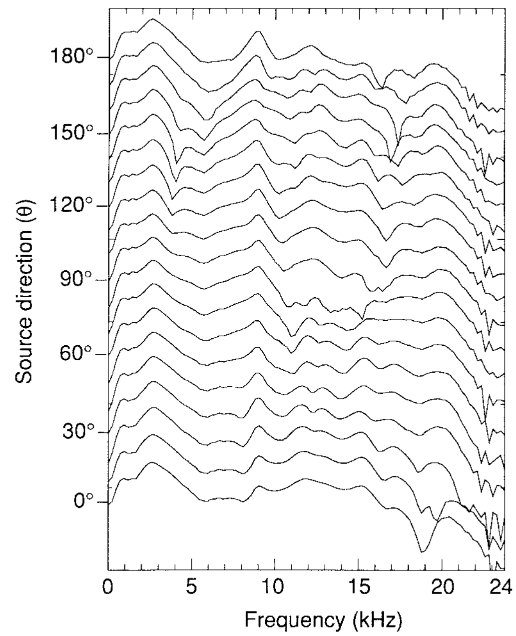

Fig. 3. Frequency responses of the measured HRTF's of the left ear.

Fig. 4 shows the frequency response of the transfer function

$$A_{\mathrm{C}}^{-1}(z) = \frac{1}{1 - \sum_{i=1}^{P} a_{\mathrm{C}i} z^{-i}} \tag{9}$$

$$\begin{aligned}
\mathbf{x} &= [\mathbf{a}^T, \mathbf{b}^T(\theta_1), \mathbf{b}^T(\theta_2), \cdots, \mathbf{b}^T(\theta_M)]^T, \\
\mathbf{a} &= [a_1, a_2, \cdots, a_p]^T, \\
\mathbf{b}(\theta_m) &= [b_0(\theta_m), b_1(\theta_m), \cdots, b_Q(\theta_m)]^T, \\
\mathbf{h}_a &= [\mathbf{h}^T(\theta_1), \mathbf{h}^T(\theta_2), \cdots, \mathbf{h}^T(\theta_M)]^T, \\
\mathbf{h}(\theta_m) &= [h_0(\theta_m), h_1(\theta_m), \cdots, h_{N-1}(\theta_m), 0, \cdots, 0]^T,
\end{aligned}$$

$$\mathbf{A} = \left.\underbrace{\begin{bmatrix} \mathbf{H}(\theta_1) & \mathbf{D} & 0 & & 0 \\ \mathbf{H}(\theta_2) & 0 & \mathbf{D} & & \\ \vdots & \vdots & & \ddots & \\ \mathbf{H}(\theta_M) & 0 & 0 & & \mathbf{D} \end{bmatrix}}_{P+M(Q+1)}\right\} M(N+P),$$

$$\mathbf{H}(\theta_m) = \left.\underbrace{\begin{bmatrix} 0 & 0 & & 0 \\ h_0(\theta_m) & 0 & & 0 \\ h_1(\theta_m) & h_0(\theta_m) & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{P-1}(\theta_m) & h_{P-2}(\theta_m) & \cdots & h_0(\theta_m) \\ \vdots & \vdots & & \vdots \\ h_{N-1}(\theta_m) & h_{N-2}(\theta_m) & \cdots & h_{N-P}(\theta_m) \\ 0 & h_{N-1}(\theta_m) & \cdots & h_{N-P-1}(\theta_m) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & h_{N-1}(\theta_m) \end{bmatrix}}_{P}\right\} N+P$$

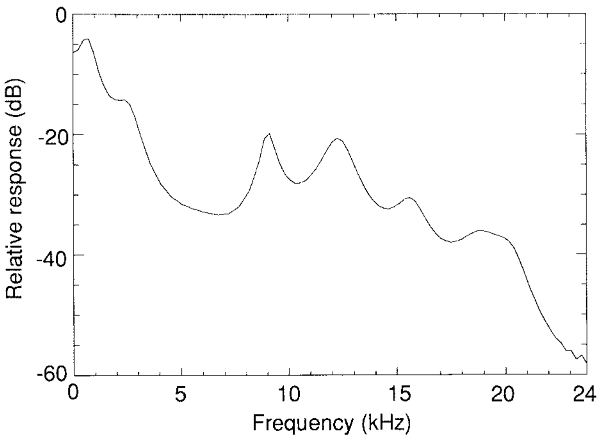

Fig. 4. Frequency response of transfer function with the estimated common AR coefficients.

with common AR coefficients a_{Ci}. This figure shows the common peaks at 2.8, 9.0, and 12.2 kHz in Fig. 3 have been extracted as the common acoustical poles. These frequencies almost match the resonance frequencies of the ear canal shown by Shaw [12]. The slight difference between the data is probably due to the differences in the ear used in the measurements. Because the common acoustical poles corresponding to the resonance frequencies of the ear canal must be invariant for the HRTF's on the vertical plane, this model would also be useful for the vertical plane.

C. Performance Evaluation of the CAPZ Model

We evaluated the performance of the CAPZ modeling of HRTF's by comparing the modeling error to that of a conventional all-zero model having the same number of coefficients. The coefficients of the all-zero model were obtained by using a rectangular window to truncate the original impulse response of the HRTF's. The modeling error $E_{\text{out}}(\theta)$ is defined as

$$E_{\text{out}}(\theta) = 10 \log_{10} \frac{\sum_{k=0}^{N} e_{\text{out}}^2(\theta, k)}{\sum_{k=0}^{N} h^2(\theta, k)}. \tag{10}$$

The CAPZ model had 20 common AR coefficients and 40 MA coefficients, and the conventional all-zero model had 60 MA coefficients, that is, both models used 60 coefficients to represent an HRTF. The HRTF's for source directions 20, 50, 80, 160, 280, and 340°, which were not used for estimating the common AR coefficients, were used for the evaluation.

Fig. 5 shows the evaluation results. The modeling errors of the proposed CAPZ model were smaller than those of the conventional all-zero model, although the CAPZ model used a smaller number of the variable coefficients that depend on the source direction. This indicates that the CAPZ model is suitable for HRTF representation. Fig. 6 shows the measured impulse response and the CAPZ-modeled impulse response at 20°. The two responses were equivalent within a modeling error of -24 dB.

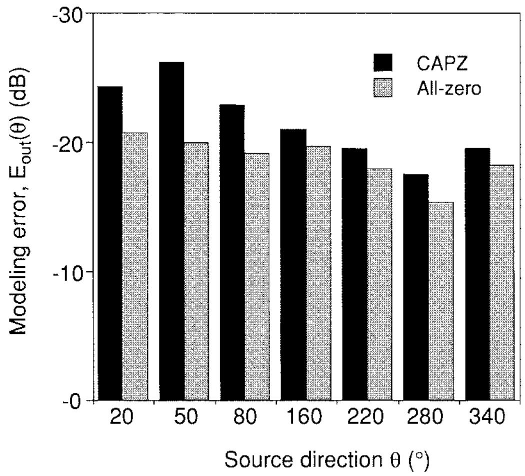

Fig. 5. Modeling errors of common-acoustical-pole and zero model (20 common AR coefficients + 40 variable MA coefficients, black bars) and the conventional all-zero model (60 MA coefficients, gray bars) for HRTF's.

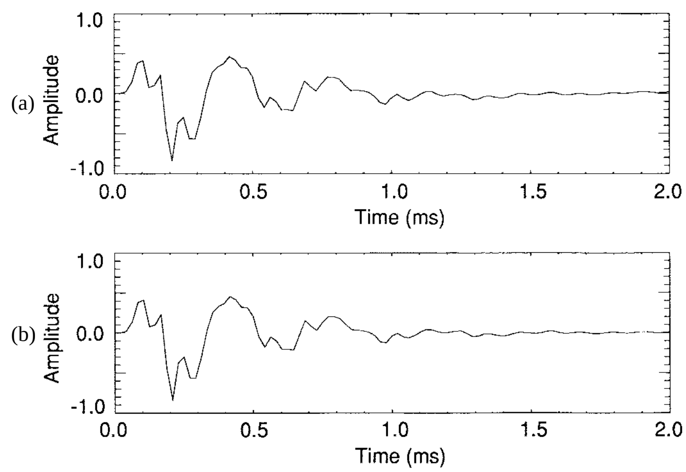

Fig. 6. Impulse responses at 20° of (a) actual HRTF and (b) common-acoustical-pole and zero modeled HRTF.

Let's consider the number of coefficients required to model multiple HRTF's, for example, 36 HRTF's (at every 10° from 0–360°). Because all the coefficients of an all-zero modeled HRTF depend on the source direction, the all-zero model requires $36 \times 60 = 2160$ coefficients to represent 36 HRTF's. In the CAPZ model, the 20 common-AR coefficients do not depend on the source direction, while the 40 MA coefficients do. Therefore, the CAPZ model requires $20 + (36 \times 40) = 1460$ coefficients to represent 36 HRTF's. That is, the CAPZ model requires 32% fewer coefficients.

When modeling the HRTF's using the conventional pole/zero model with 20 poles and 40 zeros, the average of the modeling errors was about 4 dB better than with the proposed CAPZ model. However, the pole/zero model required $36 \times (20 + 40) = 2160$ coefficients to represent 36 HRTF's. When the total number of coefficients was set equal to that of the CAPZ, that is, when the HRTF's were modeled with ten poles and 30 zeros, the average of the modeling errors of the conventional pole/zero model was 8 dB worse than with the CAPZ model. The CAPZ model was thus more efficient in reducing the total number of parameters while maintaining

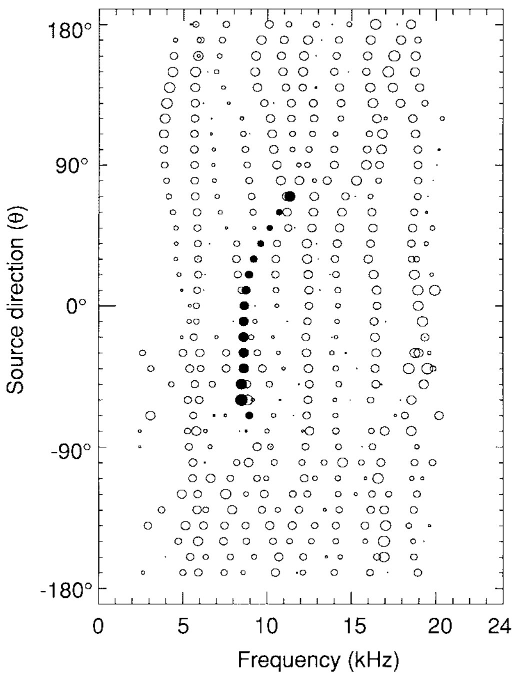

Fig. 7. Zeros for various sound directions. Open circles indicate minimum-phase zeros, and filled circles indicate nonminimum-phase zeros. Circle size corresponds to the distance between the zero and the unit circle on the z-plane (the bigger the circle, the smaller the distance).

good modeling accuracy for wave reconstruction, compared to the conventional all-zero model or the pole/zero model. The perceptual validity of the model is best addressed using listening tests which will be the subject of future studies.

IV. Analysis of Zeros in HRTF'S Using CAPZ Model

A. Directional Dependence of Zero Variations

The directional dependence of the HRTF's is an important cue for perceiving the source direction. Although the spectral cues of the HRTF's on the vertical plane are even more important than those on the horizontal plane [14], we analyzed the HRTF variations on the horizontal plane, which suffers the problem of front-back confusion. Fig. 3 shows the directional dependence of the HRTF's on the horizontal plane as frequency responses. Some regular variations of dips, or notches, can be seen. Actually, the dips are considered to be important for perceiving the source direction, and the dip variations have been analyzed in [15]–[20]. However, not all of the dips are well traced in frequency responses. For example, the 4-kHz dips between 100° and 150° cannot be traced in other angles. Because zeros create dips (notches) in the frequency response, the analysis of the zero variations must be useful.

When the HRTF's are modeled using the CAPZ model, the directional dependence of the HRTF's can be expressed using only the variations of the zeros. Furthermore, because the CAPZ modeling removes the effect of the common acoustical poles from the zeros, it can precisely estimate the zeros, which, in the conventional pole/zero model, are often influenced or canceled by the poles.

Fig. 7 shows the zeros' variations based on the CAPZ model for 36 source directions ($\theta = -180°$ to $180°$). The zero $q_i(\theta)$ of the CAPZ-modeled HRTF in (1) is a complex value, so is represented in a polar form as $q_i(\theta) = r_i(\theta)\ \exp(-2\pi j f_i(\theta)/f_s)$, where $r_i(\theta)$ denotes the magnitude of the zero, $f_i(\theta)$ denotes the frequency of the zero, and f_s denotes the sampling frequency. In Fig. 7, the horizontal axis denotes the frequency, the vertical axis is the source direction θ, and the circle position represents the frequency $f_i(\theta)$ of the zero. The open circles denote minimum-phase zeros whose magnitude $r_i(\theta)$ is smaller than the unity magnitude $r = 1$; that is, the zeros are inside the unit circle on the z-plane. The filled circles denote the nonminimum-phase zeros whose magnitude $r_i(\theta)$ is over one; that is, the zeros are outside the unit circle [21].

The size of the circles indicates the nearness of the zero's magnitude $r_i(\theta)$ to unity. The bigger the circle, the closer the zero's magnitude $r_i(\theta)$ is to unity. Zeros whose magnitude $r_i(\theta)$ is far from the unit circle on the z-plane are omitted. A source direction of $-90°$ indicates that the source was placed on the opposite side from the left ear (that is, near the right ear). Therefore, the signal-to-noise ratio (SNR) is low near $-90°$ and some estimation errors were found there.

Fig. 7 shows that the directional dependence of the HRTF's is characterized better by the zero variations than by the dip variations in the frequency responses shown in Fig. 3. For example, the dip around 4 kHz can be traced only between 100° and 150° in Fig. 3, but the zero variations can be well traced from 0–180° in Fig. 7.

B. Nonminimum-Phase Zero Variations

A physical model that could explain the directional dependence of the zero variations is attractive because it would provide rules for the zero variations, and if the zero variations due to change in source direction are known, the HRTF variations can be controlled. Let's consider a physical model for the nonminimum-phase zero variations indicated as filled circles in Fig. 7.

A nonminimum-phase zero is generally generated when indirect sound, or reflected sound, which arrives after the direct sound, has greater energy than the direct sound. When the several reflected sounds arrive at a receiver in a room at the same time, the indirect sound has greater energy than the direct sound. Let's consider an impulsive direct and a reflected sound, where the amplitudes of direct and reflected sound are, respectively, 1 and $-a$ $(a > 1)$. When the delay in arrival between direct and reflected sound is τ, a nonminimum-phase zero occurs at the frequency

$$f = 1/\tau. \tag{11}$$

The longer the arrival delay, the lower the frequency of the zero.

The frequency of the nonminimum-phase zero in Fig. 7 does not change for the source directions $-70°$ to $0°$, but it increases as the source direction increases from $0°$ to $70°$. This implies that the delay τ between a direct and a strong reflected sound decreases as the source direction increases from $0°$ to $70°$.

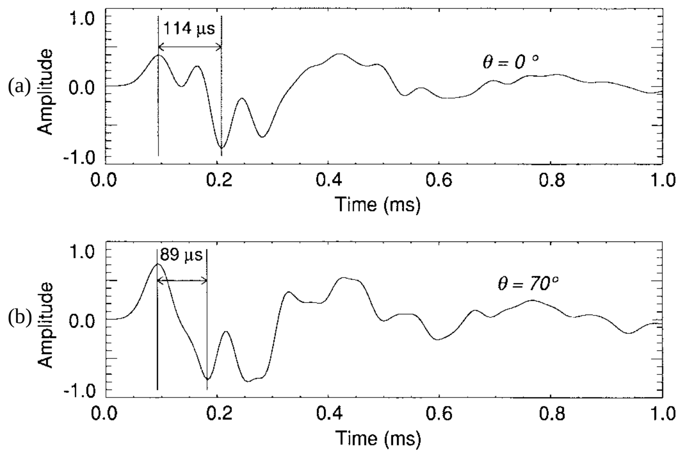

Fig. 8. Impulse responses of (a) 0° and (b) 70° HRTF's. The delay between the direct and reflected sounds decreases as the source-direction angle increases.

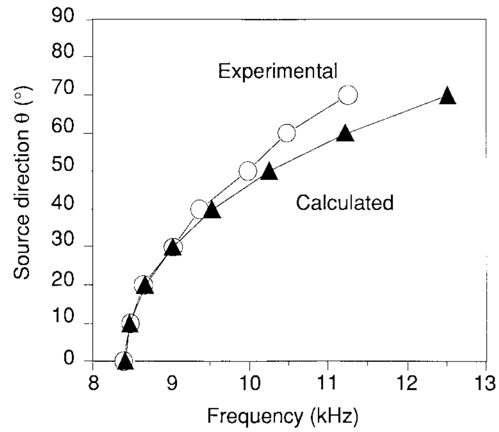

Fig. 10. Frequency variation of nonminimum-phase zero depending on source direction. Circles indicate the zeros extracted using common-acoustical-pole and zero modeling of HRTF's. Triangles indicate the zeros calculated using the simple reflection model.

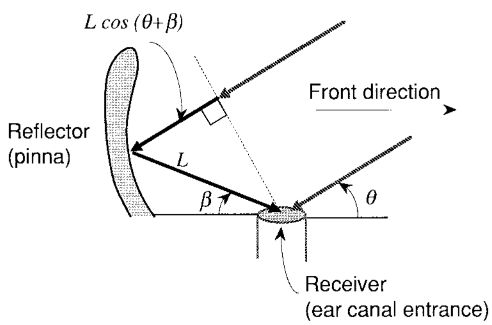

Fig. 9. Simple ear model in which sound is reflected by the pinna.

Fig. 8 shows the impulse responses of the HRTF's at source directions 0° and 70°. The delay τ in the impulse response at source direction 70° is shorter than that at 0°. The amplitude of the largest reflected sound is bigger than that of the direct sound in both responses. The nonminimum-phase zero in Fig. 7 can be considered to be generated by these direct and reflected sounds.

Here, we assume that the largest reflected sound is from the pinna, which collects the sound and reflects it toward the ear canal. Based on this assumption, the nonminimum-phase zero is analyzed as the characteristics of the sound reflected by the pinna. Moreover, that the pinna works well as a sound collector for frontal source directions explains the appearance of a nonminimum-phase zero between $-70°$ and $+70°$. Note that the characteristics of the sound reflected by the pinna have been analyzed as a notch in the frequency response of the HRTF [15], [17].

We now consider a simple model of an ear (Fig. 9) that has a reflector corresponding to the pinna and a receiver corresponding to the entrance of the ear canal. In this figure, the several sounds reflected from the pinna are assumed to be one strong reflected sound. Angle θ denotes the incident angle of the direct sound, β denotes the incident angle of the reflected sound, and L denotes the average distance between the reflector and the receiver (ear canal entrance). In this model, the path difference S between the direct and reflected sounds is expressed as

$$S = L[1 + \cos(\theta + \beta)]. \quad (12)$$

Because it is difficult to determine angle β, we assumed it to be simply 0°. As a result, the path difference S can be approximated as

$$S = L(1 + \cos\ \theta). \quad (13)$$

Because delay τ is the path difference S divided by sound velocity c, the frequency of a nonminimum-phase zero is approximately

$$f = \frac{1}{\tau} = \frac{c}{S} = \frac{c}{L(1 + \cos\ \theta)}. \quad (14)$$

Here, when incident angle θ is less than 0°, the denominator in (14) can be assumed to be a constant $2L$ from Fig. 9. Therefore, the frequency of a nonminimum-phase zero does not change for incident angles under 0°, as found from Fig. 7.

Fig. 10 shows the frequency variations of a nonminimum-phase zero as calculated using (14) and the variations extracted from the measured HRTF's. The distance L was set to 20 mm, which is almost equal to the distance between the surface of the pinna and the ear canal of the dummy head used for the measurement. This result shows that the simple pinna-reflection model in Fig. 9 represents well the nonminimum-phase zero variations extracted from the measured HRTF's. Thus, the nonminimum-phase zeros in Fig. 7 can be considered to be generated by the direct sound and the sound reflected from the pinna. Although this physical model seems to be very simple, the rule of the nonminimum-phase zero variations is well formulated by (14).

V. Conclusion

We have proposed the use of the common acoustical poles to model HRTF's. These poles are invariant with HRTF variations that result from changes in the source direction. They are estimated as values common to the HRTF's, corresponding to the various source directions. The estimated common acoustical poles agree well with the resonance frequencies of the ear canal. Because the proposed common-acoustical-pole and zero model expresses the directional dependence of the HRTF's by using only the zeros, it requires fewer parameters that depend on the source direction than do the conventional models.

Furthermore, because the CAPZ model can extract the zeros that are missed due to pole-zero cancellation in the conventional pole/zero model, the directional dependence of the zeros is well traced. In addition, good agreement was found between the nonminimum-phase zero variations on the horizontal plane extracted by the CAPZ model and those predicted by a simple pinna-reflection model as a function of direction.

To date, we have analyzed the nonminimum-phase zero variations. Examination of the relationship between the directional dependence of the zero variations and human perception remains as future work. We expect that our CAPZ model will become a powerful technique for analyzing the directional dependence of HRTF's and for studying the spectral cues in HRTF's.

Acknowledgment

The authors are grateful to J. Kojima for his support and suggestions.

References

[1] J. Blauert, *Spatial Hearing*. Cambridge, MA: MIT Press, 1983.
[2] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321, May 1995.
[3] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, pp. 167–174, July 1980.
[4] H. Hamada, N. Ikeshoji, Y. Ogura, and T. Miura, "Relation between physical characteristics of orthostereophonic system and horizontal plane localization," *J. Acoust. Soc. Jpn. (E)*, vol. 6, pp. 143–154, July 1985.
[5] F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoust. Soc. Amer.*, vol. 85, pp. 868–878, Feb. 1989.
[6] S. Shimada and S. Hayashi, "Stereophonic sound image localization system using inner-earphones," *Acustica*, vol. 81, pp. 264–271, May/June 1995.
[7] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Amer.*, vol. 88, pp. 159–168, July 1990.
[8] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale, "Low-order modeling of head-related transfer functions using balanced model truncation," *IEEE Signal Processing Lett.*, vol. 4, pp. 39–41, Feb. 1997.
[9] M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 278–287, May 1997.
[10] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 320–328, Apr. 1994.
[11] ——, "Modeling a room transfer function using common acoustical poles," in *Proc. 1992 Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 2, pp. 289–292, Mar. 1992.
[12] E. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Amer.*, vol. 44, pp. 240–249, July 1968.
[13] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA: MIT Press, 1983.
[14] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *J. Acoust. Soc. Amer.*, vol. 92, pp. 2607–2624, Nov. 1992.
[15] R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," *J. Acoust. Soc. Amer.*, vol. 61, pp. 1264–1269, May 1977.
[16] H. L. Han, "Measuring a dummy head in search of pinna cues," *J. Audio Eng. Soc.*, vol. 42, pp. 15–37, Jan./Feb. 1994.
[17] C. A. P. Rodgers, "Pinna transformations and sound reproductions," *J. Audio Eng. Soc.*, vol. 29, pp. 226–234, Apr. 1981.
[18] J. Herbrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Amer.*, vol. 56, pp. 1829–1834, Dec. 1974.
[19] D. Wright, J. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," *J. Acoust. Soc. Amer.*, vol. 56. pp. 957–962, Sept. 1974.
[20] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Amer.*, vol. 63, pp. 1152–1165, Apr. 1978.
[21] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

Yoichi Haneda (A'92–M'97) was born in Sendai, Japan, on June 17, 1964. He received the B.S. and M.S. degrees in physics from Tohoku University, Sendai, in 1987 and 1989, respectively.

Since joining Nippon Telegraph and Telephone Corporation (NTT), Tokyo, Japan, in 1989, he has been investigating acoustic signal processing and acoustic echo cancellers. He is currently a Research Engineer at the Speech and Acoustics Laboratory, 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), Tokyo, Japan, 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, 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 Dr.Eng. 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), 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, 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, and 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.

Nobuhiko Kitawaki (M'87) was born in Aichi, Japan, on September 27, 1946. He received the B.E.E., M.E.E., and Dr.Eng. degrees from Tohoku University, Sendai, Japan, in 1969, 1971, and 1981, respectively.

From 1971 to 1997, he was engaged in research on speech and acoustics information processing at Nippon Telegraph and Telephone Corporation. From 1993 to 1997, he has been Executive Manager of the Speech and Acoustics Laboratory. He currently serves as Professor of the Institute of Information Sciences and Electronics, University of Tsukuba, Ibaraki, Japan. He also serves as Rapporteur of ITU-T Study Group 12.

Dr. Kitawaki is a member of the Institute of Electronics, Information, and Communication Engineers of Japan (IEICEJ) and the Acoustical Society of Japan. He received Paper Awards from IEICEJ in 1979 and 1984.