External ear transfer function modeling: A beamforming approach

Jiashu Chen
Department of Electrical and Computer Engineering and of Neurology, University of Wisconsin-Madison, Madison, Wisconsin 53706

Barry D. Van Veen
Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706

Kurt E. Hecox
Department of Neurology, University of Wisconsin-Madison, Madison, Wisconsin 53792

(Received 26 March 1991; accepted for publication 11 June 1992)

In this article, a beamformer is proposed as a functional model for the spatial and temporal filtering characteristics of the external ear. The output of a beamformer is a weighted combination of the data received at an array of spatially distributed sensors. The beamformer weights and array geometry determine its spatial and temporal filtering characteristics. A procedure is described for choosing the weights to minimize the mean-squared error between the beamformer response and the measured response of the external ear. The effectiveness of the model is demonstrated by designing a beamformer of several hundred weights that duplicates and interpolates the measured external ear response of a cat over broad ranges of frequency and direction. A limited investigation of modeling performance as a function of array geometry is reported.

PACS numbers: 43.64.Ha, 43.64.Bt, 43.66.Qp

INTRODUCTION

Extensive physical and behavioral studies have revealed that the external ear (including torso, head, pinna, and canal) plays an important role in spatial hearing. The external ear provides directional amplification of the incident sound-pressure level (Shaw, 1974), that is, sound from different directions undergoes different overall gain. The external ear modifies the spectrum of the incoming sound according to incidence angle of that sound (Shaw, 1974). In the context of binaural hearing, the spectral difference created by the external ears introduces important cues for localizing sounds in addition to interaural time and intensity differences (Oldfield, 1984). When the sound source is within the sagittal plane, or in the case of monaural hearing, the spectral cues provided by the external ear are utilized almost exclusively by the auditory system to identify the location of the sound source (Butler and Belendiuk, 1977). The external ears also externalize the sound image. Sounds presented binaurally with the original time and intensity differences but without the spectral cues introduced by the external ears are typically perceived as originating inside the listener's head (Wightman and Kistler, 1989a).

These observations have greatly stimulated interest in the physical and functional aspects of the external ear transformation characteristics. Physical models seek to represent the physical mechanism by which the external ear transforms the incident sound pressure, while functional models focus on directly representing the transformation characteristics of the external ear. Physically modeling the external ear is a challenging task since: (1) the size of the external ear (human's and large mammal's) is of the same magnitude as the sound wavelengths of interest; hence, the external ear should be modeled as a distributed rather than a lumped system; (2) the geometry of the external ear is very complex and an analytic solution to the boundary value problem for this system is probably impossible; (3) the diversity of the external ear geometry among subjects is large and a physical model would have to be individualized to account for the significant psychoacoustic consequences of subtle geometry change.

In spite of these difficulties, various physical models have been proposed and evaluated by different authors. Two models, described by Blauert as the frequency- and time-domain models (Blauert, 1983), typify these efforts. The time-domain model was proposed by Batteau (Batteau, 1967, 1968). He measured and studied the impulse response of an enlarged pinna replica and concluded that the physical structure of the external ear introduced two significant echoes in addition to the original sound. One echo varies with the azimuthal position of the sound source, having a latency in the 0- to 80-μs range, while the other varies with elevation in the 100- to 300-μs range. He conjectured that the external ear could be modeled as a three-hole coupler which is equivalent to a two-delay-and-add system (see Fig. 1). The output $y(t)$ is related to the input $x(t)$ as

$$y(t) = x(t) + a_1 x(t - \tau_1) + a_2 x(t - \tau_2), \quad (1)$$

where τ_1 refers to azimuthal echo latency and τ_2 refers to the elevation echo latency; a_1 and a_2 are the two constants representing the corresponding reflection coefficients. The frequency response of this model is

$$H(\omega) = 1 + a_1 \exp(-j\omega\tau_1) + a_2 \exp(-j\omega\tau_2). \quad (2)$$

Later studies showed that the frequency response of such a two-delay-and-add system roughly resembles the response of the external ear for angles of incidence in the lateral vertical plane (Watkins, 1982; Wright *et al.*, 1974).

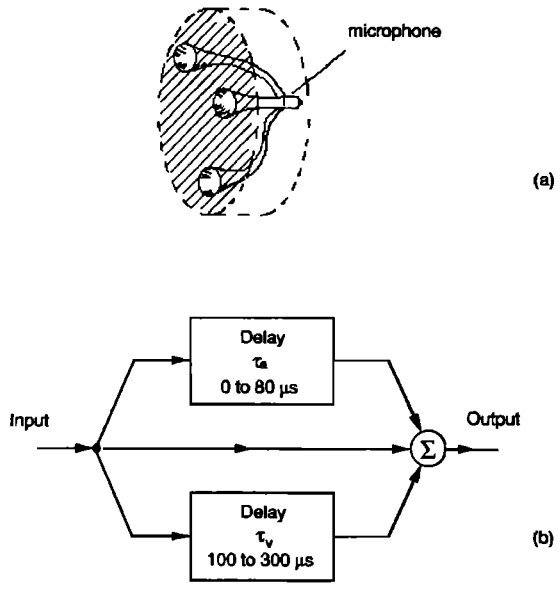

FIG. 1. (a) Batteau's three-hole coupler for the external ear (after Batteau, 1967); (b) Two-delay-and-add system representing Batteau's three-hole coupler. The upper and lower paths introduce variable delay "echoes" of (0–80 ms) and (100–300 ms) with respect to the central undelayed path (after Watkins, 1982).

The advantage of Batteau's model is its simplicity. It contains only four parameters: a_1, a_2, τ_1, and τ_2. The location of the source (determined by the azimuth angle and elevation angle) is explicitly coded in this model. Although Batteau's model is based on physical considerations, it can also be used to simulate the function of the external ear on a computer using Eq. (1). The limitation of Batteau's model is that it does not give an accurate representation of the external ear response. This is because the dimension of the external ear's reflecting surface is small in comparison with the wavelength for most audible frequencies so dispersion occurs instead of reflection. Thus there are actually many additional echoes besides the two Batteau described (Blauert, 1983).

The frequency-domain approach was taken by Shaw (1979). By measuring the frequency responses of a simplified replica of human pinna, he identified eight resonance frequencies between 2.6 and 15.7 kHz in the model and found that these resonance frequencies varied with the incident angle of the sound source. Shaw's experiments greatly expanded our understanding of the physics of the pinna. However, this work does not lead to an analytic expression for the external ear transformation characteristics. Hence, this type of model is purely physical and cannot be used to simulate the response of the external ear.

The primary contribution of Batteau's and Shaw's modeling work was to provide explanations for the acoustical mechanics of the external ear. In many applications, e.g., performing sound localization experiments via earphone presentation, the exact transformations of the sound source introduced by the external ear and other parts of the body as a function of source location are of great interest. This interest has motivated measurement of the so-called "head-related-transfer-functions" (HRTF). Due to its practical importance, a large number of investigators have studied this problem (Blauert, 1969; Blauert, 1972; Feddersen *et al.*, 1957; Harrison and Downey, 1970; Jahn and Vogelsang, 1959; Robinson and Whittle, 1960; Schirmer, 1963; Shaw, 1966; Sivian and White, 1933; Tröger, 1930; Wiener, 1947; Wiener and Ross, 1946; Wilkens, 1971, 1972; Blauert *et al.*, 1971; Laws, 1972; Mellert, 1972; Mehrgardt and Mellert, 1977). More recent work has been reported by Middlebrooks *et al.* (1989), Wightman and Kistler (1989a, b), Musicant *et al.* (1990), Asano *et al.* (1990), and Loomis *et al.* (1990).

The HRTF provides a subject specific observation of the spatial/temporal transformation characteristics of the external ear. These observations may be related to the behavioral differences among subjects. Also, the HRTF can be used to simulate the eardrum signal as a function of source location. However, the HRTF is based solely on acoustical measurements and thus results in vast quantities of data, especially if many source locations are of interest. It does not represent the external ear over the entire auditory space, but only at the measurement locations. It does not generalize the data in a parametric way so that intersubject or interspecies differences can be characterized.

A parametric model was reported by Genuit (1986). He asserted that by using Kirchhoff's diffraction integrals it is possible to describe approximately the external ear transfer function influenced by the head, torso, shoulder, and pinna. In his model, 16 channels of filter-delay combinations represent the acoustical contributions of the head, torso, shoulder, and pinna to the overall responses of the external ear. All of these delays and filter coefficients are explicit mathematical functions of the ear geometry and source location. This is the first model that completely describes all the acoustical functions of the external ear and establishes the mathematical relationship between the ear geometry and the transfer function. He demonstrated that the predicted external ear amplitude spectrum was between the minimum and maximum amplitudes obtained over six measurements of a subject's external ear response in a single direction.

In this paper, we report efforts toward modeling the function of the external ear using beamforming concepts. A beamformer implements a joint spatial and temporal filter by forming a weighted combination of the data collected at an array of spatially distributed sensors. The weights and sensor geometry completely determine the spatial and temporal filtering characteristics of the beamformer. Historically, the weights in the earliest beamformers were chosen to replicate the spatial response of continuous aperture antennae. In this same spirit, we view the external ear as an acoustic antenna and attempt to duplicate its response using a beamformer. The sensor geometry is chosen based on simple physical considerations of the external ear acoustics. Given the sensor geometry, a least-squares procedure is used to determine the weights so that the mean-squared error between the spatial/temporal responses of the beamforming model and the measured response of the external ear is minimized. The model response is the inner product of the weight vector and an "array response vector." The array response vector explicitly represents the array geometry and the

source location. The effectiveness of this approach is demonstrated by modeling the measured temporal and spatial response of a cat's external ear over a limited spatial region. The results show that the model is very accurate over a broad range of frequency and spatial directions.

We emphasize that this approach leads to a functional model for the external ear. In contrast with a physical model (such as Shaw's), this model does not explicitly model the physics of the external ear; it models the transformation characteristics of the external ear in terms of an input–output relationship. The physics are only modeled in a general sense by the sensor geometry and the FIR filters in each sensor channel.

This functional approach has the following advantages: (a) It provides an explicit mathematical expression that accounts for the directional dependence of the acoustic wave at the eardrum; (b) it represents the external ear transformation characteristics as a continuous function of spatial direction even though discrete measurements are used to determine the model parameters; (c) it has computational efficiencies compared to alternative approaches making feasible its inclusion in other models (e.g., cochlear transmission models); (d) it has sufficient generality that it can be applied to a wide range of acoustic environments, acoustic signals and receiver classes (multiple species); (e) it is subject specific, that is, given subject specific measurements the model can be customized to individual subjects via intersubject differences in the respective weight vectors; (f) it results in substantial data compression when used to represent HRTF's measured at many directions and frequencies.

The paper is organized as follows. Section I mathematically defines beamforming and derives the expression for the temporal and spatial beamformer response. Batteau's model is also interpreted as a beamformer in this section. In Sec. II, we formulate and solve the least-squares criterion for choosing the beamformer weights to approximate a desired temporal and spatial response. Sections III and IV discuss the experimental aspects of this work. In Sec. III, we briefly describe the data used to verify the effectiveness of the beamforming model. Section IV describes the results of using the beamformer to model a cat's pinna. The paper concludes with a discussion in Sec. V. Notationally, we use boldface lower- and upper-case symbols to represent vectors and matrices, respectively. Superscripts T and H denote matrix transpose and complex conjugate transpose, respectively. The symbol * denotes complex conjugate.

I. PRINCIPLES OF BEAMFORMING

There exists a vast body of technical references in the antenna engineering and signal processing literatures that describe beamformer design, analysis, and implementation for a wide variety of applications. In this section, we introduce concepts and terminology that are relevant to our external ear modeling work. A recent tutorial by Van Veen and Buckley (1988) contains additional detail and numerous references.

A beamformer generates a single output signal as a weighted sum of the signals received at multiple sensors. Historically, the beamformer provided an alternative to continuous aperture dish antennae. The advantage of beamforming is that the spatial response is synthesized using the weights instead of aperture shape. Hence, the spatial response is not constrained by the physical limitations associated with construction of a continuous aperture antenna. The weights in the earliest beamformers were designed so that the spatial response approximated that of known continuous aperture antennae. This is analogous to the problem addressed here: We seek to design a beamformer whose response mimics that of the external ear, our equivalent of a continuous aperture antenna.

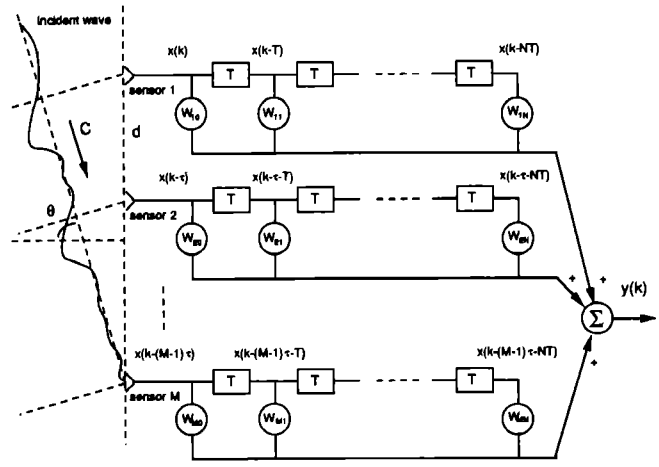

FIG. 2. Beamformer structure incorporating FIR filters in each sensor channel: C stands for the velocity of wave propagation; θ denotes the angle between the direction of propagation and a line perpendicular to the axis of the sensor array; d stands for the intersensor distance; and T stands for the sampling interval.

A beamforming structure commonly used when broadband signals are of interest is depicted in Fig. 2. The output of the beamformer is a weighted combination of delayed versions of the input signal. The response can be expressed in the same form

$$h(\omega,\theta) = \mathbf{w}^H \mathbf{d}(\omega,\theta), \tag{3}$$

where the array response vector $\mathbf{d}(\omega,\theta)$ is defined as

$$\begin{aligned}\mathbf{d}^H(\omega,\theta) = [\,&1,\exp(j\omega T)\ldots,\exp(j\omega NT),\exp(j\omega\tau_1),\\ &\exp(j\omega(\tau_1+T)),\ldots,\\ &\exp(j\omega(\tau_1+NT)),\ldots,\exp(j\omega\tau_{M-1}),\\ &\exp(j\omega(\tau_{M-1}+1)),\ldots,\exp(j\omega(\tau_{M-1}+NT))\,]\end{aligned} \tag{4}$$

and

$$\mathbf{w}^H = [\,w^*_{0;0},\ldots,w^*_{0;N},w^*_{1;0},\ldots,w^*_{1;N},w^*_{M-1;1},\ldots,w^*_{M-1;N}\,]. \tag{5}$$

In Eq. (4), N and T represent the FIR filter order and sampling interval, respectively. In the following discussion we assume $T=1$ for simplicity. The symbol τ_i denotes the time delay due to propagation from the first to the ith sensor. Hence τ_i encodes both the sensor array geometry and source location. In Eq. (5), the first subscript in the definition of $\mathbf{w}$ refers to the sensor channel index and the second subscript to the tap index.

Comparing Eqs. (2) and (3), Batteau's three hole

coupler depicted in Fig. 1 can be viewed as a simple beamformer since the output is a sum of delayed versions of the input where the delays are a function of source location.[1] However, this model has very limited capabilities because it only uses three sensors and does not include temporal samples in each sensor channel.

II. LEAST-SQUARES BEAMFORMER DESIGN

The beamformer response $h(\omega,\theta)$ is a function of the weight vector $\mathbf{w}$ and the manner in which the array structure samples the propagating signal. The sampling characteristics are a function of the array geometry and are reflected in the definition of the array response vector $\mathbf{d}(\omega,\theta)$. In this section, we determine $\mathbf{w}$ for a given array geometry so that the actual beamformer response approximates a desired response. Optimization of array geometry is not considered; it is a difficult problem due to the complicated manner in which the geometry enters into $\mathbf{d}(\omega,\theta)$.

Let $h_d(\omega,\theta)$ represent the desired response. In the present application $h_d(\omega,\theta)$ is the external ear response. The error is

$$e(\omega,\theta) = h_d(\omega,\theta) - \mathbf{w}^H\mathbf{d}(\omega,\theta). \tag{6}$$

Minimization of the mean-squared error over a range of ω denoted by the set Ω and range of θ denoted by the set Θ is expressed as

$$\min_{\mathbf{w}} \int_{\theta\in\Theta}\int_{\omega\in\Omega} \gamma(\omega,\theta)|e(\omega,\theta)|^2\, d\omega\, d\theta. \tag{7}$$

Here, $\gamma(\omega,\theta)$ is a non-negative weighting function used to emphasize some frequencies and directions over others. In practice, the external ear response cannot be measured as a continuous function of ω and θ, but is measured at a set of distinct directions θ_i, $i = 1,2,...,Q$ and frequencies ω_j, $j = 1, 2,...,P$. Thus we consider the problem

$$\min_{\mathbf{w}} \sum_{i=1}^{Q}\sum_{j=1}^{P} \gamma(\omega_j,\theta_i)|e(\omega_j,\theta_i)|^2. \tag{8}$$

Problem (8) is equivalent to solution of a system of linear equations in a least-squares sense. Define an error vector

$$\begin{aligned}\mathbf{e}^H = [&e(\omega_1,\theta_1),e(\omega_2,\theta_1),...,e(\omega_P,\theta_1),e(\omega_1,\theta_2),\\ &e(\omega_2,\theta_2),...,e(\omega_P,\theta_2),...,e(\omega_1,\theta_Q),\\ &e(\omega_2,\theta_Q),...,e(\omega_P,\theta_Q)]\end{aligned} \tag{9}$$

and let

$$\mathbf{e}^H = \mathbf{h}_d^H - \mathbf{w}^H\mathbf{D}^H, \tag{10}$$

where

$$\begin{aligned}\mathbf{h}_d^H = [&h_d(\omega_1,\theta_1),h_d(\omega_2,\theta_1),...,h_d(\omega_P,\theta_1),h_d(\omega_1,\theta_2),\\ &h_d(\omega_2,\theta_2),...,h_d(\omega_P,\theta_2),...,h_d(\omega_1,\theta_Q),h_d(\omega_2,\theta_Q),...,\\ &h_d(\omega_P,\theta_Q)]\end{aligned} \tag{11}$$

and

$$\begin{aligned}\mathbf{D}^H = [&\mathbf{d}(\omega_1,\theta_1),\mathbf{d}(\omega_2,\theta_1),...,\mathbf{d}(\omega_P,\theta_1),\mathbf{d}(\omega_1,\theta_2),\\ &\mathbf{d}(\omega_2,\theta_2),...,\mathbf{d}(\omega_P,\theta_2),...,\mathbf{d}(\omega_1,\theta_Q),\mathbf{d}(\omega_2,\theta_Q),...,\\ &\mathbf{d}(\omega_P,\theta_Q)].\end{aligned} \tag{12}$$

Write $\gamma(\omega_j,\theta_i)$ as a diagonal weighting matrix of dimension PQ by PQ

$$\Gamma = \mathrm{diag}\{\gamma^{1/2}(\omega_1,\theta_1),\gamma^{1/2}(\omega_2,\theta_1),...,\gamma^{1/2}(\omega_P,\theta_Q)\}. \tag{13}$$

Equation (8) is rewritten as

$$\min_{\mathbf{w}}\|\Gamma(\mathbf{h}_d - \mathbf{D}\mathbf{w})\|_2^2 \tag{14}$$

which is equivalent to the least-squares solution of a system of $S = PQ$ linear equations in $L = M(N+1)$ unknowns

$$\Gamma\mathbf{D}\mathbf{w} = \Gamma\mathbf{h}_d. \tag{15}$$

In general, $S > L$.

Thus far we have assumed $\mathbf{w}$ is complex valued. The response of the physical systems that we are attempting to model are complex conjugate symmetric in frequency, i.e., $h_d(\omega,\theta) = h_d^*(-\omega,\theta)$. This implies we should constrain the modelled response to be complex conjugate symmetric in frequency, or equivalently, $h^*(\omega,\theta) - h(-\omega,\theta) = 0$. Now $h^*(\omega,\theta) = \mathbf{w}^T\mathbf{d}^*(\omega,\theta)$ and $h(-\omega,\theta) = \mathbf{w}^H\mathbf{d}(-\omega,\theta)$. Note that $\mathbf{d}^*(\omega,\theta) = \mathbf{d}(-\omega,\theta)$ so we must have $h_d(\omega,\theta) - h_d^*(-\omega,\theta) = (\mathbf{w}^T - \mathbf{w}^H)\mathbf{d}(-\omega,\theta) = 0$. This implies that $\mathbf{w}^T = \mathbf{w}^H$ and hence $\mathbf{w}$ must be real valued.

Let $\mathbf{e} = \mathbf{e}_R + j\mathbf{e}_I$, where $\mathbf{e}_R$ and $\mathbf{e}_I$ are the real and imaginary components of $\mathbf{e}$. Assuming $\mathbf{w}$ is real, we have

$$\mathbf{e}_R = \mathbf{h}_{dR} - \mathbf{D}_R\mathbf{w}, \tag{16a}$$

$$\mathbf{e}_I = \mathbf{h}_{dI} - \mathbf{D}_I\mathbf{w}, \tag{16b}$$

where $\mathbf{h}_{dR}$, $\mathbf{h}_{dI}$, $\mathbf{D}_R$, and $\mathbf{D}_I$ denote the real and imaginary components of $\mathbf{h}_d$ and $\mathbf{D}$, respectively. Using the fact that $\|\Gamma\mathbf{e}\|_2^2 = \|\Gamma\mathbf{e}_R\|_2^2 + \|\Gamma\mathbf{e}_I\|_2^2$, it is easy to show that (14) is equivalent to the real-valued problem

$$\min_{\mathbf{w}}\|\mathbf{g} - \mathbf{C}\mathbf{w}\|_2^2, \tag{17}$$

where $\mathbf{g}^T = [\mathbf{h}_{dR}^T\Gamma\ \ \mathbf{h}_{dI}^T\Gamma]$ and $\mathbf{C}^T = [\mathbf{D}_R^T\Gamma\ \ \mathbf{D}_I^T\Gamma]$.

There are several standard techniques for solving Eq. (17) (see, for example, Lawson and Hanson, 1974). Let $\mathbf{C}$ have a singular value decomposition (SVD)

$$\mathbf{C} = \sum_{i=1}^{r} \sigma_i\mathbf{u}_i\mathbf{v}_i^T. \tag{18}$$

r is the rank of $\mathbf{C}$. Here, $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right singular vectors and σ_i are the singular values. We assume $\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r \geqslant \sigma_{r+1} = ... = \sigma_L = 0$. The solution for $\mathbf{w}$ is expressed as

$$\mathbf{w} = \sum_{i=1}^{r} \sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i^T\mathbf{g}. \tag{19}$$

$\mathbf{w}$ in (19) depends on the inverse of the singular values. If $\mathbf{C}$ is rank deficient, numerical evaluation of the SVD will not usually give exactly zero singular values. These very small σ_i result in a $\mathbf{w}$ with very large norm. Thus, it is common to use only ρ "significant" singular values in solving for $\mathbf{w}$:

$$\mathbf{w} = \sum_{i=1}^{\rho} \sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i^T\mathbf{g}. \tag{20}$$

Here, ρ is usually chosen based on the number of singular values which exceed a specified threshold δ. The mean-squared error using (20) is

$$\|\mathbf{g} - \mathbf{C}\mathbf{w}\|_2^2 = \mathbf{g}^T \sum_{i=\rho+1}^{2S} \mathbf{u}_i \mathbf{u}_i^T \mathbf{g}. \quad (21)$$

while the norm of $\mathbf{w}$ is

$$\|\mathbf{w}\|_2^2 = \mathbf{g}^T \sum_{i=1}^{\rho} \sigma_i^{-2} \mathbf{u}_i \mathbf{u}_i^T \mathbf{g}. \quad (22)$$

Clearly, the error increases and norm of $\mathbf{w}$ decreases as ρ decreases.

III. EXPERIMENTAL DATA DESCRIPTION

The effectiveness of the beamforming model has been verified using the external ear transfer function of a cat. The external ear transfer function of a cat is of interest over a broader frequency range and generally has a similar spectral shape to that of a human. We believe that if the model is effective at duplicating the response of a cat, it will also be effective with human responses.

The data were acquired in the sound localization research lab in Department of Neurophysiology, University of Wisconsin-Madison and is fully described in Musicant *et al.* (1990). Ear canal recordings were made with a miniature probe microphone placed very close to the tympanic membrane in the cat's ear canal with a sound source located at 75 cm from the recording site. The sound source was located at an elevation of 18 deg above the recording microphone. Defining the front direction on the median plane as 0-deg azimuth, the sound source azimuth varied from 0 to 90 deg clockwise in increments of 4.5 deg. The signal applied to the speaker was a digitally generated rectangular pulse, 10 μs in duration. The signal received at the microphone was sampled by an A/D converter at a rate of 160 kHz and averaged over 300 trials to minimize the effect of noise. A total of 1024 sampling points, representing 6.4-ms duration, were recorded for each speaker location.

All of the stimulus generating and data recording devices were band limited to less than 40 kHz, so the 160-kHz sampling rate of the original data was reduced to 80 kHz using a decimation algorithm and then the transfer functions $h_d(\omega,\theta)$ were calculated as the ratios of the spectra of the ear canal recordings to the free-field recording. The 21 transfer functions, representing $h_d(\omega,\theta)$ at 21 different directions from 0 to 90 deg, are used in determining the model weights. Each consists of 257 frequency samples.

It has been discovered that the data contains small artifacts resulting from undesired acoustic reflections introduced in the data collection process (Chan, 1990). Fortunately, these reflections arrive well after the external ear impulse response decays and thus are easily removed with a time domain window. Figure 3(a) depicts a typical impulse response with the contaminating echoes marked in the plot. Figure 3(b) illustrates the amplitude spectrum of the contaminated impulse response. The echoes produce wiggles in the amplitude spectrum, particularly at low frequencies. Figure 3(c) and (d) depicts the impulse response and amplitude spectrum after removing the echoes with a 625-μs width time window.

The differential delay between the stimulus arrival times under free field and ear canal conditions is an uncontrolled

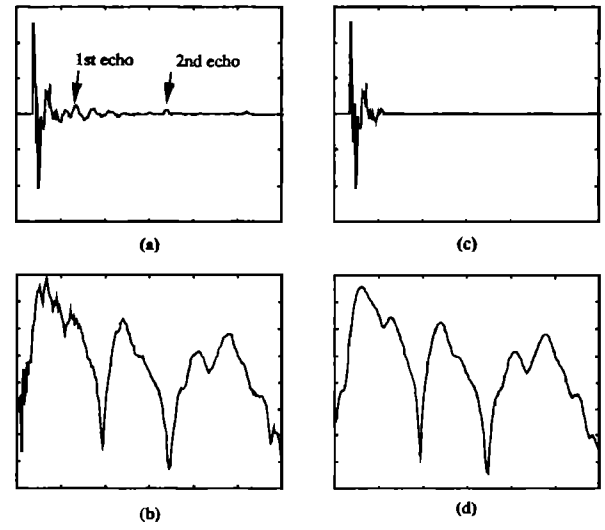

FIG. 3. Impulse response and amplitude spectra before and after removal of echo artifacts. (a) Contaminated impulse response; (b) contaminated amplitude spectrum; (c) windowed impulse response; (d) windowed frequency response. Impulse response x axes represent time from 0 to 4.8 ms; y axes represent amplitude from -3 to 3 V. Frequency response x axes represent frequency from 0 to 40 kHz; y axes represent amplitude from -15 to 25 dB.

factor in the data recording process. While this delay does not affect the amplitude spectrum, it does introduce ambiguity into the phase spectrum of $h_d(\omega,\theta)$. Therefore, the phase of $h_d(\omega,\theta)$ is not well defined. The importance of unambiguous representations for the external ear transfer function has been addressed by several authors (Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989a) in the context of psychoacoustical research.

In modeling external ear transfer functions with a beamformer, uniquely defined transfer functions are also needed to specify the desired response $\mathbf{h}_d$. We have used three different approaches to remove phase ambiguities: (1) a linear-phase method in which we replace the original phase by linear phase with group delay proportional to the beamformer's temporal aperture center (see Buckley, 1987); (2) a minimum-phase method (Oppenheim and Shafer, 1975) in which the original phase is replaced by its minimum phase component. This guarantees a unique representation of the original transfer function with minimum time delay; (3) the all-phase processing method suggested by Mehrgardt and Mellert (1977). It decomposes the original phase into a minimum phase component and an all-pass phase component. The all-pass phase is smoothed and the linear component representing pure time delay is removed. The nonlinear phase component that remains is added to the minimum-phase component.

For a given array structure, we obtain the best approximation to the desired amplitude response using method (1) to process the raw transfer functions. However, this method completely ignores the measured phase response; as a result the modeled external ear impulse response exhibits substantial differences in morphology when compared to the measured impulse response. At the other extreme, method (3) results in the least morphological distortion between the modeled and the measured impulse response, but the largest error between modeled and measured frequency response.

Minimum phase processing results in a compromise.

Mehrgardt and Mellert (1977) report that their human data shows minimum phase characteristics up to 10 kHz. This has proven true for data acquired by other authors (Wightman and Kistler, 1989a). Our cat's data exhibit minimum phase characteristics up to 25 kHz. Considering the fact that the higher frequency portion of the transfer functions represents a relatively small portion of the overall energy due to its lower amplitudes, we employ the minimum phase approach [method (2)] in the results presented in the following section. Comparison between the processed and unprocessed impulse responses shows that this method introduces a slight shift in energy distribution.

IV. RESULTS

A. Preliminary discussion

In this model, there are three sets of parameters. One is the weight vector which is determined by solving the least-squares problem. Another set of parameters are related to the physical structure of the beamformer. They are the array geometry, the number of sensors M, number of taps $N+1$ in each sensor channel, intersensor distance d (assuming uniform spacing), and the aperture of the beamformer. For a linear array with uniform sensor spacing, the aperture is product of $M-1$ and d. The aperture for an "L" shaped array is defined later. This set of parameters determine the array vector equation (4). The model is thus represented by the inner product of the two vectors. The third set of parameters that influence the model performance are the weighting function $\gamma(\omega,\theta)$ and the threshold value δ.

The model performance is evaluated both qualitatively and quantitatively. Plots of the measured and modeled amplitude and phase spectra and impulse responses are provided for qualitative evaluation. For quantitative performance evaluation, the average approximation error is utilized:

$$\|\mathbf{e}\|_2^2 = \frac{1}{Q}\sum_{i=1}^{Q}\frac{\Sigma_{j=1}^{P}|h_d(\omega_j,\theta_i)-\mathbf{w}^H\mathbf{d}(\omega_j,\theta_i)|^2}{\Sigma_{j=1}^{P}|h_d(\omega_j,\theta_i)|^2} = \frac{1}{Q}\sum_{i=1}^{Q}\|\mathbf{e}_i\|_2^2, \tag{23}$$

where

$$\|\mathbf{e}_i\|_2^2 = \frac{\Sigma_{j=1}^{P}|h_d(\omega_j,\theta_i)-\mathbf{w}^H\mathbf{d}(\omega_j,\theta_i)|^2}{\Sigma_{j=1}^{P}|h_d(\omega_j,\theta_i)|^2}. \tag{24}$$

Equation (24) is the cumulative mean-squared error over all frequencies normalized by the total energy of the measured response at one direction and Eq. (23), the average approximation error, is the average of $\|\mathbf{e}_i\|_2^2$ over all directions. The average approximation error, converted to percentile form, is used to evaluate the model performance in the following discussion.

For simplicity, we assume $\gamma(\omega,\theta)$ is unity for all ω and θ unless otherwise indicated, the threshold value δ used to determine ρ, the number of significant singular values, was set to $\alpha\times\sigma_1$, where σ_1 is the maximum singular value and $\alpha = 10^{-16}$. As is shown later, large weight magnitudes and poor interpolation performance can result when α is this small. However, a very small value for α was chosen to facilitate comparison of different sensor/tap configurations. This choice for α always results in $\rho = r$ [see Eqs. (19) and (20)] and yields the smallest possible mean-squared error for a given configuration. In all simulations, the sampling interval is kept as a constant of 12.5 μs, which corresponds to a sampling rate of 80 kHz.

B. Array structure variations

1. Fixed number of weights—vary number of sensors and number of taps

The fact that the external ear is a joint spatial and temporal filter is demonstrated by Figs. 4 and 5. These figures represent results for seven models based on about 400 weights arranged in different sensor/tap configurations. The sensors are arranged in a linear equally spaced geometry. We note that in all configurations, the intersensor distance is a constant of 0.0043 m. Therefore, the spatial aperture is proportional to the number of sensors. Due to the constant sampling rate the time window width is approximately proportional to the number of taps in each sensor channel. The 1-sensor-400-tap configuration is an ordinary FIR filter. It does not have the ability to differentiate the transfer functions corresponding to different directions. As a result of this deficiency, this model only gives an averaged transfer function for every direction it is supposed to describe. The leftmost column in Fig. 5 depicts this case. The average approximation error is about 19%. The 2-sensor-200-tap configuration gives a much better average approximation error of 8.66% because spatial discrimination is now possible. At the other extreme, the 64-sensor-7-tap structure performs poorly because it does not have sufficient temporal aperture. It cannot track the frequency variation. The 4-sensor, 8-sensor, and 16-sensor structures perform well and

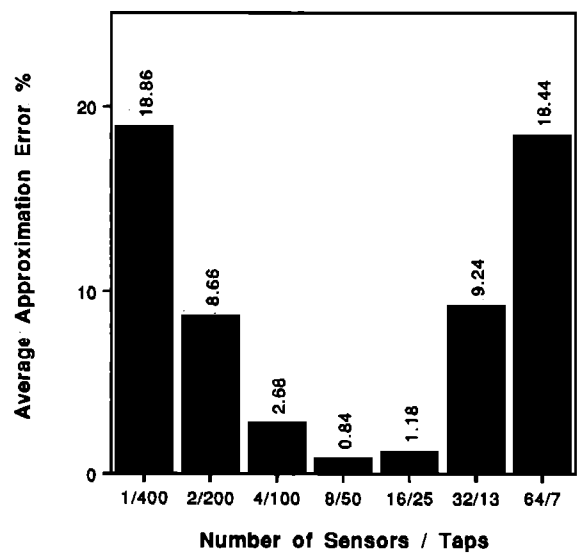

FIG. 4. Approximation error as a function of the number of sensors and taps. The total number of weights in the model is fixed at approximately 400. A linear array structure is used.

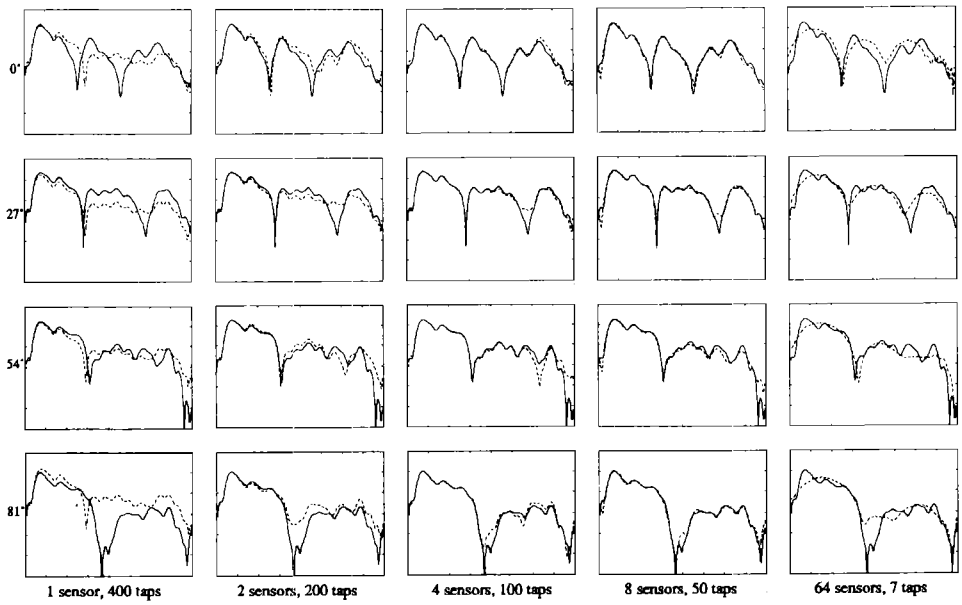

FIG. 5. Measured (solid line) and modeled (dashed line) amplitude responses in four directions for different combinations of numbers of sensors and numbers of FIR filter taps in each channel. The vertical scale for amplitude responses is − 30 to 30 dB. The horizontal axis spans from 0 to 40 kHz.

with the 8-sensor-50-tap resulting the best approximation. It has a sufficient temporal aperture width (50 taps at a sampling interval of 12.5 μs results in a time window of 625 μs) and a sufficient spatial aperture (intersensor distance of 0.0043 m timed by seven results in a spatial aperture of 30 mm which is roughly equal to the pinna dimension).

2. Vary array geometry

The relationship between beamformer response and array geometry is complex due to the manner in which the geometry enters into the response equation. Here, we discuss one particular variation in geometry suggested by the physical configuration of the cat's pinna and observations of linear array performance. The linear array geometry has difficulty in accurately duplicating the response for large values of azimuth θ (see, for example, 81° vs 0° in Fig. 5). In a linear array, the response depends on θ through the intersensor time delay which is a function of sin θ. For small θ the beamformer is sensitive to differences in direction because $\sin\theta \approx \theta$. However, the beamformer is insensitive to differences in direction for θ near $\pi/2$ since then $\sin\theta \approx 1$. The pinna itself is two dimensional when viewed in the azimuthal plane and suggests an L-shaped array geometry. The L-shaped array has one sensor always located at the intersection of the x and y axes and has other sensors lined up on both

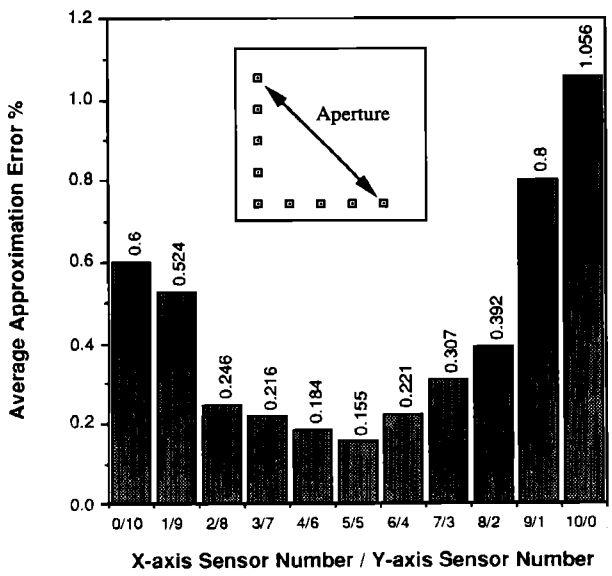

FIG. 6. Approximation error as a function of the number of sensors along each axis of an L-shaped array geometry. The number of sensors is 11 and one sensor is always located at the intersection of the array's x and y axes. Forty-six FIR filter taps are used in each sensor channel. The insert depicts the L-shaped array structure and its aperture. The intersensor distance is 8 mm for all cases.

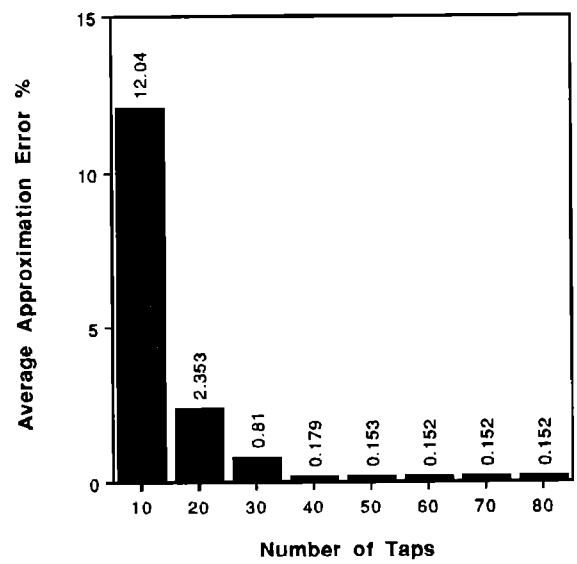

FIG. 7. Approximation error versus the number of FIR filter taps following each sensor for a 5-1-5 L-shaped array configuration. The intersensor distance is 8 mm for all cases.

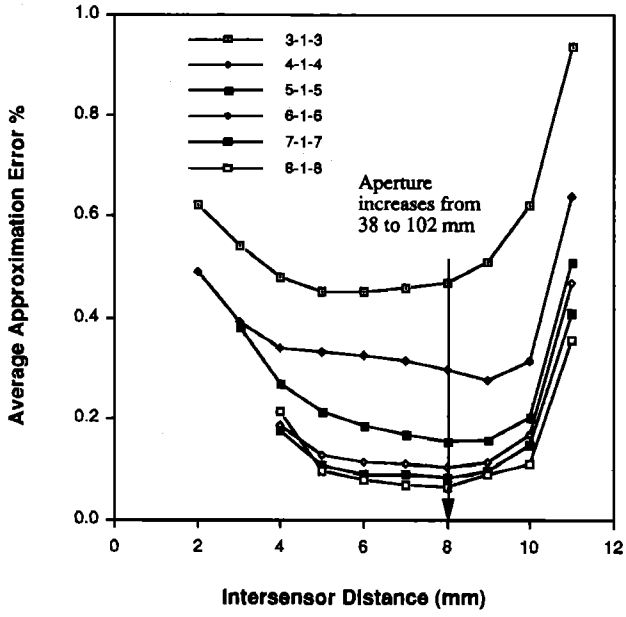

FIG. 8. The average approximation error versus intersensor distance for six different sensor configurations. All the configurations have 46 taps in each sensor channel.

x and y axes. The y axis corresponds to the axis of the linear sensor array used previously and the x axis is perpendicular to the y axis. In other words, the x axis is aligned with the angle of $\theta = 0°$.

Figure 6 depicts the average approximation error as a function of the number of sensors along each axis for an L-shaped geometry. Here, the total number of sensors is chosen as eleven. Forty-six taps are used in each sensor channel. The size of the aperture, in the case of L-shaped array, is defined as the distance between the two ends of the arms of the L-shaped array (see insert in Fig. 6). For the case of an equal arm 5-1-5 configuration, the aperture is 1.414 times of the length of one arm. Using an equal number of sensors on each axis (five on the x axis and five on the y axis with one at the intersection) the average approximation error is 0.155% relative to errors of 0.6% and of 1.06% for linear array geometries with all sensors on either x axis and y axis, respectively. This result is consistent with Shaw's finding that the resonators on the pinna have almost perpendicular maximum response directions (Shaw, 1979).

3. Fixed number of sensors—vary number of taps

Figure 7 shows the average approximation error versus the number of taps in each sensor channel with the number of sensors held constant at 11 arranged in an L-shaped geometry. Increasing the number of taps reduces the average approximation error. However, no additional reductions are obtained by using more than 50 taps. This is because the measured impulse responses have been previously time limited to 625 μs, the temporal aperture spanned by 50 taps. In general, an excessive number of taps is undesirable because it leads to increased computational complexity. Furthermore, excess degrees of freedom will be used to model artifacts and noise if they are present. Figure 7 indicates that excellent performance is obtained by using between 40 and 50 taps.

4. Vary array aperture and intersensor distance

The choice of intersensor distance (ISD) is a simple matter if the sensor array has a linear equally spaced geometry and the signal is narrow band. According to spatial sampling theory, the optimal ISD is $\lambda_0/2$, where λ_0 is the wavelength of the waveform. This ISD is optimal in the sense of obtaining the largest possible aperture while preventing spatial aliasing. Unfortunately, there are no rules for choosing the ISD with sensor array having L-shaped geometries, time

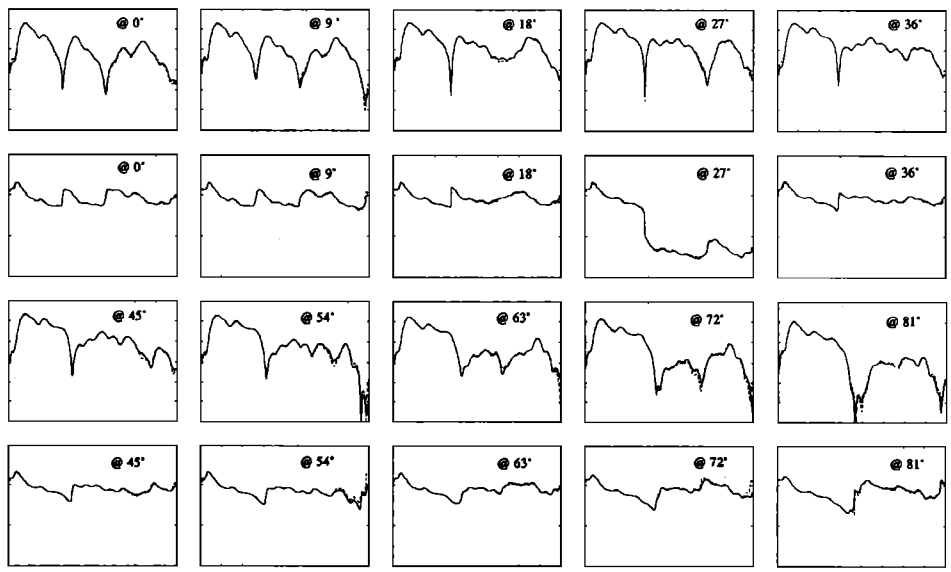

FIG. 9. Comparison of measured (solid line) and modeled (dashed line) amplitude and phase responses in ten directions. The vertical scale for amplitude responses (first and third rows) is − 30 to 30 dB while the scale for the phase responses (second and fourth rows) ranges from − 10 to 5 radians. The horizontal axis spans from 0 to 40 kHz. The elevation is 18 deg and the azimuth is denoted in each panel. The model has 11 sensors and 46 FIR filter taps in each sensor channel with a L-shaped sensor geometry. The intersensor distance is 8 mm.

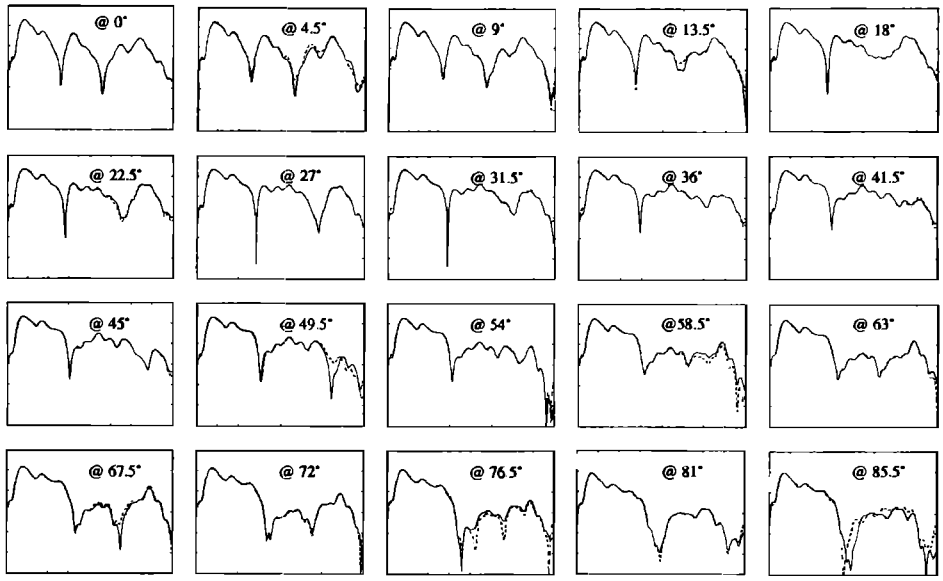

FIG. 10. Measured (solid line) and modeled (dashed line) external ear frequency responses illustrating interpolation capability. The model is determined using the measured data at 0,9,18,...,90 deg. The responses at 4.5,13.5,22.5,...,85.5 deg are interpolated by the model. The vertical scales span from − 30 to 30 dB and the horizontal from 0 to 40 kHz. This model has a structure of 11 sensors and 46 taps in each sensor channel with a L-shaped geometry. The intersensor distance is 8 mm.

taps in the sensor channels, and that process broadband signals.

Figure 8 depicts the average approximation error as a function of both the number of sensors and the ISD. In this figure, the six curves represent the average approximation error as a function of ISD for arrays based on different numbers of sensors (ranging from 7 to 17). For each sensor configuration there is an optimal ISD for which the average approximation error is minimum. Note that there is a range of ISD, from 5 to 9 mm, for which all arrays have small average approximation error. For 11 or more sensors an ISD of 8 mm gives the smallest average approximation error.

For a fixed ISD, the average approximation error decreases as the number sensor increases. This is because of degrees of freedom and the array aperture increase with increasing number of sensors. Note however, that the computational complexity and capability of modeling noise and artifacts increase as the number of degrees of freedom increase. The 5-1-5 configuration with ISD of 8 mm results in a reasonable compromise between average approximation error and complexity. Figure 9 depicts the measured and modeled amplitude and phase spectra at 9° degree increments in azimuth from 0 to 81 deg. There is an excellent agreement between measured and modeled response.

C. Interpolation error

One of the desirable characteristics of this model is its ability to represent the external ear response at directions other than those at which the response was measured. In effect, the model interpolates the response between measurement directions. Figure 10 depicts the measured and model-interpolated responses in several directions. The model response is determined using 11 sensors in a 5-1-5 L-shaped geometry with 46 taps per sensor. The threshold value α is set to be 10^{-7}. Here, the weights are determined based on measurements at 0,9,18,...,90 deg. The responses at 4.5, 13.5,22.5,...,85.5 degrees are calculated using these weights from Eq. (3). The average approximation error is about 0.3%, compared to an average approximation error of 0.16% for noninterpolated case. Comparison with Fig. 9 indicates that the errors for the interpolated responses are of the same order as the error obtained by including the interpolated directions in the weight vector determination process.

The interpolation performance of the model is sensitive to the threshold parameter δ used to determine the number of significant singular values ρ in Eq. (20). As ρ increases, the mean-squared error decreases, but the norm of the weight vector dramatically increases because the inverse of very small singular values are used to calculate **w** [see Eq.

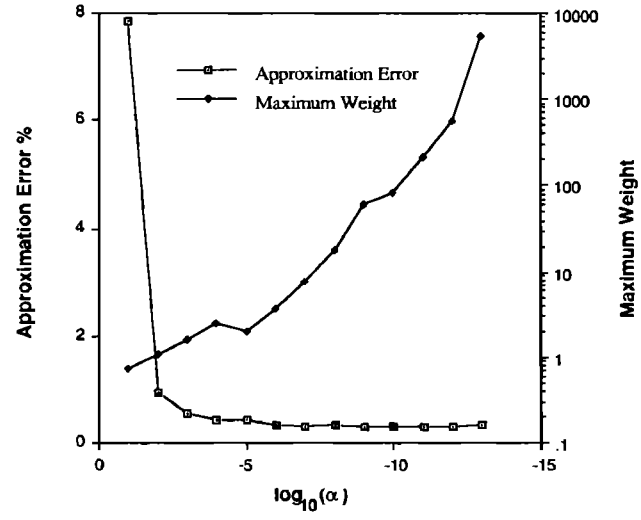

FIG. 11. Average approximation error and maximum weight magnitude versus singular value threshold α.

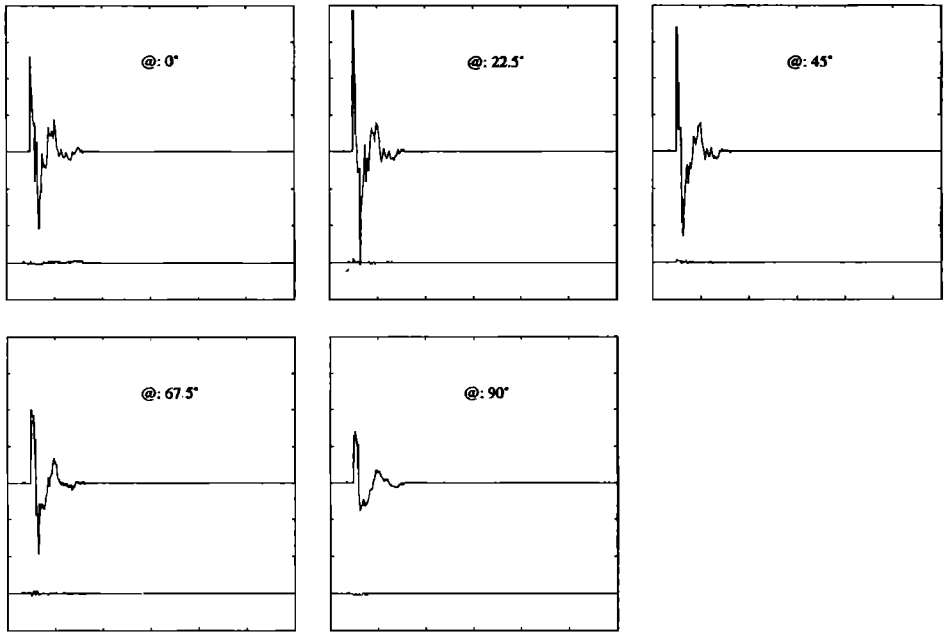

FIG. 12. Comparison between the measured (solid line) and modelled (dashed line) impulse responses at five different azimuthal locations: 0, 22.5, 45, 67.5, and 90 deg. The lower curve is the difference between measured and modeled responses. Each major division on the time axis represents 0.5 ms. The vertical scales span from −4 to 4 V.

(20)]. This leads to erratic response behavior between the spatial sample points used to determine $\mathbf{w}$. As ρ decreases, the mean-squared error increases and the response becomes smooth in between sample points.

A similar effect occurs in fitting polynomial curves to data. Use of a high-order polynomial will produce small error at the data points but can result in erratic behavior in between data points; a low-order polynomial will give a smoother overall fit at the expense of increased error at the data points.

Figure 11 illustrates the average approximation error and maximum weight as a function of $\log_{10}(\alpha)$. Recall $\alpha = \delta/\sigma_1$. The error is approximately constant for $\alpha < 10^{-5}$ and tends to increase rapidly for $\alpha > 10^{-5}$. In contrast, the magnitude of the largest element of $\mathbf{w}$ becomes large for $\alpha < 10^{-8}$. In these cases, the interpolation performance is unstable.

D. Time-domain verification

Thus far we have evaluated the model performance entirely in the frequency domain. An alternative perspective is obtained by evaluating the model performance in the time domain. We now compare the measured and modeled external ear impulse responses.

The frequency response of this model is defined as

$$h(\omega,\theta) = \mathbf{w}^{\mathrm{H}}\mathbf{d}(\omega,\theta),$$

so the model impulse response is the inverse Fourier transform of the frequency response

$$h(t,\theta) = \mathrm{F}^{-1}\{h(\omega,\theta)\}.$$

Similarly, the measured impulse response of the external ear is

$$h_d(t,\theta) = \mathrm{F}^{-1}\{h_d(\omega,\theta)\}.$$

Define the error between measured and modeled impulse responses as

$$e(t,\theta) = h_d(t,\theta) - h(t,\theta).$$

Figure 12 depicts the modeled and measured impulse responses in several directions. The model response is determined using 11 sensors in a 5-1-5 L-shaped geometry with 46 taps per sensor. The majority of the energy in the measured impulse response is located in the first 0.5 ms; these features are faithfully duplicated by the model. The bottom trace in each panel of Fig. 12 shows the difference between the two responses. We observe that the error has a noiselike behavior.

V. SUMMARY AND DISCUSSION

We have proposed modeling the spatial and temporal response characteristics of the external ear with a beamformer. The output of a beamformer is a weighted combination of the data received at an array of spatially distributed sensors. The array geometry and beamformer weights determine its spatial and temporal response characteristics. The weights in the beamforming model are chosen to minimize the mean-squared error between measured external ear and beamformer responses. This model represents the response as a continuous function of frequency and direction.

The effectiveness of the beamforming model is demonstrated using the measured external ear response of a cat. The model is able to actually duplicate and interpolate the measured response over 90° of azimuth and a 40-kHz bandwidth using several hundred weights. A limited investigation of model performance as a function of array geometry is also reported. We show that L-shaped array geometries result in substantially less average approximation error than linear array geometries. The L-shaped geometry is consis-

tent with the physical configuration of the external ear.

The mean-squared error is always reduced by increasing the number of parameters in the model, e.g., number of sensors or taps. However, it is generally desirable to follow the principle of parsimony and use a model with smallest number of parameters needed to model the desired features. Parsimonious models are computationally less complex and less likely to model artifacts and noise.

Batteau's model can be interpreted as a three-sensor-two-weight beamformer. Our beamforming model has far more sensors and weights. This increase in complexity accounts for the dispersion that occurs and permits explicit modeling of the external ear as a temporal/spatial filter. Our work is similar to Genuit's work in the sense that both use a multichannel filter structure. The primary difference is the methods by which the model parameters are determined. In Genuit's model the parameters are determined by an averaged geometry of the external ear. The spatial and temporal response is then predicted based on these parameters. The advantage of this approach is that the variations of the response due to geometry change can be predicted (or simulated) to some extent. However, if this approach is applied to an individual geometry, the predicted response perhaps needs to be verified by acoustical measurements unless an extensive verification of the dynamics of the parameters is performed with respect to all of the possible parameter variations. In our model, the parameters are partially determined by the external ear geometry (i.e., the array geometry) and partially determined through acoustical measurements. It can predict (or simulate) the exact response of the external ear it models. Thus it can be used to represent either average or individual external ear transfer functions, depending upon whether average or individual measurements are used in determining the model weights.

The experimental results presented in this paper only tested the model as a function of azimuth for a fixed elevation angle. There are a number of directions in which the model can be extended, including variations in azimuth and elevation, use of a three-dimensional array geometry proportionally dimensioned to the physical size of the external ear and incorporation of an algorithm which accounts for interaural time differences, thus establishing a binaural external ear transfer function model. Ultimately, the advantages of this model mentioned in the introduction should become most apparent in complex and computationally demanding applications such as synthesizing a moving sound stimulus in space or a continuously changing stimulus in time. By controlling the model order and the weighting function, the synthesized external ear transfer function can be smoothed and/or emphasized both spatially and temporally, which will extend the experimenter's flexibility in spatial hearing research.

Several other issues remain for future research. We have demonstrated the effectiveness of the beamforming model at the acoustical level. Psychoacoustic validation will be an important step in fully establishing the effectiveness of this model. The present model has limited capability of representing the nonminimum phase components of the external ear response although the significance of this limitation is largely unknown because behavioral studies have not yet definitively established the importance of the nonminimum phase components of the response. Optimization of array geometry is very difficult due to the complicated manner in which the sensor locations enter into the expression for the beamformer's response. Yet progress in more closely linking details of the physical model with the functional model will not only increase the power of the model but decrease the computational burden by decreasing either the number of sensors or the length of FIR filters required to achieve a given level of accuracy. Despite the many remaining challenges, we are encouraged by the accuracy of model's performance and the "tractable" nature of the model computationally.

ACKNOWLEDGMENTS

We express our deep appreciation to our colleagues in the Department of Neurophysiology, University of Wisconsin-Madison. Their generosity in sharing their data and experience made this work possible. We also appreciate very helpful discussions with Dr. Joseph C. Chan, Dr. Alan D. Musicant, Professor J. Hind, and Professor W. Tompkins. We thank one of the reviewers for suggestions that substantially improved the paper. This research was supported by the National Institutes of Health NINCDS Grant No. NSR 0116436.

[1] In general the response is a function of range, azimuth, and elevation. For notational simplicity we use the single variable θ to describe location.

Asano, F., Suzuki, Y., and Sone, T. (1990). "Role of spectral cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Batteau, D. W. (1967). "The role of the pinna in human localization." Proc. R. Soc. London, Ser. B **168**, 158–180.

Batteau, D. W. (1968). *Listening with the Naked Ear. The Neuropsychology of Spatially Oriented Behavior* (Dorsey, Homewood, II).

Blauert, J. (1969). "Investigations of directional hearing in the median plane with the head immobilized," Dissertation, Technische Hochschule, Aachen.

Blauert, J., Hartmann, R., and Laws, P. (1971). "Entfernungs-und Richtungsabhängigkeit des Übertragungsfaktors des äusseren Ohres [Distance and direction dependence of the transfer factor of the external ear]," in Proceedings of the 7th International Congress on Acoustics, Budapest, 25 H 5 (as cited in Blauert, 1983).

Blauert, J. (1972). "Die Schallausbreitung im äusseren Ohr and Konswquenzen für das räumliche Hören [Sound propagation in the external ear and its consequences for spatial hearing]," Convention, Audio Engineering Society, München (as cited in Blauert, 1983).

Blauert, J. (1983). *Spatial Hearing* (MIT, Cambridge, MA).

Buckley, K. M. (1987). "Spatial/spectral filtering with linearly constrained minimum variance beamformers," IEEE Trans. ASSP **ASSP-35**, 249–266.

Butler, R. A., and Belendiuk, K. (1977). "Spectral cues utilized in the localization of sound in the median sagittal plane," J. Acoust. Soc. Am. **61**, 1264–1269.

Chan, J. C. (1990). Personal communication.

Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (1957). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.

Genuit, K. (1986). "A description of the human outer ear transfer function by elements of communication theory," Proceedings of 12th International Congress on Acoustics, Toronto, Canada. ADSTR. B6-8

Harrison, J. M., and Downey, P. (1970). "Intensity changes at the ear as a function of the azimuth of a tone source: A comparative study," J. Acoust. Soc. Am. **47**, 1509–1518.

Jahn, G., and Vogelsang, S. (1959). "Die einohrige Richtcharakteristik des

menschlichen Gehörs [The directional characteristics of human hearing with one ear]," Hochfreqenztech. u. Eletroakustik. **68**, 50–56 (as cited in Blauert, 1983).

Laws, P. (**1972**). "On the problem of distance hearing and the localization of auditory events inside the head," Dissertation, Technische Hochschule Aachen (as cited in Blauert, 1983).

Loomis, J. M., Hebert, C., and Cicinelli, J. G. (**1990**). "Active localization of virtual sounds," J. Acoust. Soc. Am. **88**, 1757–1764.

Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.

Mellert, V. (**1972**). "Construction of a dummy head after new measurements of the threshold of hearing," J. Acoust. Soc. Am. **51**, 1359–1361.

Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal." J. Acoust. Soc. Am. **86**, 89–108.

Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). "Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons," J. Acoust. Soc. Am. **87**, 757–781.

Oldfield, S. a. P. S. (**1984**). "Acuity of sound localization: a topograph of auditory space. II. Pinna cues absent," Perception **13**, 601–607.

Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-hall, Englewood Cliffs, NJ).

Robinson, D. E., and Whittle, L. S., (1960). "The loudness of directional sound fields," Acustica **10**, 74–80.

Schirmer, W. (**1963**). "Die Richtcharakeristik des Ohres [The directional characteristics of the ear]," Hochfrequenztech. u. Elektroakustik. **72**, 39–48 (as cited in Blauert, 1983).

Shaw, E. A. G. (**1966**). "Ear canal pressure generated by a free sound field," J. Acoust. Soc. Am. **39**, 745.

Shaw, E. A. G. (**1974**). "Physical models of the external ear," Proc. 8th Int. Congr. Acoust. **1**, 206.

Shaw, E. A. G. (**1979**). "The elusive connection: 1979 Rayleigh Medal Lecture," Ann. Mtg. Institute of Acoustics (U.K.), Southampton, England. Printed in "Acoustics and Noice Control in Canada."

Sivian, L. J., and White, S. D. (**1933**). "On minimum audible sound fields," J. Acoust. Soc. Am. **5**, 288–231.

Tröger, J. (**1930**). "Die Schallaufnahme durch das äussere Ohr (Collection of sound by the external ear)," Phys. Z. **31**, 26–47 (as cited in Blauert, 1983).

Van Veen, B. D., and Buckley, K. M. (**1988**). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**(2), 4–24.

Watkins, A. J. (**1982**). "The monaural perception of azimuth: A synthesis approach," in *Localization of Sound: Theory and Application* (Amphora, Groton, CT).

Wiener, F. M. (**1947**). "On the diffraction of a progressive sound wave by the human head," J. Acoust. Soc. Am. **19**, 143–146.

Wiener, F. M., and Ross, D. A. (**1946**). "The pressure distribution in the auditory canal in a progressive sound field," J. Acoust. Soc. Am. **18**, 401–408.

Wightman, F. L., and Kistler, D. J. (**1989a**). "Headphone simulation of free-field listening: I: Stimulus synthesis," J. Acoust. Soc. Am. **85**(2), 858–867.

Wightman, F. L. and Kistler, D. J. (**1989b**). "Headphone simulation of free-field listening: II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878.

Wilkens, H. (**1971**). "Subjektive Ermittlung der Richtcharakteristik des Kopfes und einer kopfbezogenen Aufnahmne und Wiedergabeanordnung [Evaluation of the directional characteristics of the head and of a head-related sound-collection and reproduction apparatus by means of auditory experiments]," in *Gemeinschaftstagung für Akustik und Schwingungstechnik* (VDI-Verlag, Düsseldorf, Berlin 1970) (as cited in Blauert, 1983)].

Wilkens, H. (**1972**). "Kopfbezügliche stereophone, ein Hilfsmittel für Vergleich und Beurteilung verschiedener Raumeindrüche [Head-related stereophony, and aid in comparing and judging various spatial impressions," Acustica **26**, 213–221 (as cited in Blauert, 1983)].

Wright, D., Hebbrank, J. H., and Wilson, B. (**1974**). "Pinna reflections as cues for localization," J. Acoust. Soc. Am. **56**, 957–962.