Auditory space expansion via linear filtering

Barry D. Van Veen
Department of Electrical and Computer Engineering, University of Wisconsin, 1415 Johnson Drive, Madison, Wisconsin 53706-1691

Rick L. Jenison
Department of Psychology, University of Wisconsin, 1202 West Johnson Street, Madison, Wisconsin 53706-1691

A signal-processing algorithm that modifies the interaural time delays associated with directional sources is described. Signals received at two microphones are processed by four linear filters arranged in a lattice configuration to produce two outputs, one for each ear. Since the processing is linear, the method is equally applicable to single or multiple directional sources. The filters are designed to minimize the average squared error between a user specified desired space warping function and the actual warping function that they implement. Two classes of filters are considered: filters whose frequency response is unconstrained and filters constrained to be causal with finite impulse response. In both cases the solution to the least-squares problem is given and properties of the actual space warping function are examined. Perceptual experiments and analysis of acoustic waveforms are utilized to demonstrate the effectiveness of the algorithm. Extension of this method for utilizing more than two microphones is described.

PACS numbers: 43.66.Qp, 43.66.Pn, 43.66.Nm, 43.60.Gk [WAY]

INTRODUCTION

An individual's ability to organize auditory space is greatly enhanced by having two channels of acoustic input. Central processing for localization depends upon neural interaction between these channels in the auditory brain stem. Most theories of binaural processing originate from the neural network model of Jeffress (1948, 1972). In general, a bank of peripheral auditory bandpass filters partitions the incoming acoustic signal into channels of nerve fiber activity. Each channel is cross correlated with the contralateral channel having the same corresponding center frequency. Central processes subsequently map the degree of interaural coincidence to the perception of the sound source location in the lateral plane. It has been suggested that the consistency of coincidence patterns across different frequencies is also important in attributing the patterns to a unique, realizable sound source (Stern *et al.*, 1988).

This paper describes the development of a binaural signal-processing algorithm designed to effectively "expand" the lateral plane represented by the neural interaction. Signals received at a pair of microphones are processed with the goal of changing the apparent spatial separation between signals arriving from different directions. The processed signals are presented via right and left channels to the listener.

This work was motivated by the desire to assist the listener in environments where they have greatest difficulty in understanding speech. These environments are often characterized by multiple competing speakers. In contrast to many noise reduction approaches, this signal-processing algorithm does not attempt to improve the acoustic signal to noise ratio by removing interfering sources, but instead effectively moves them to a more peripheral region of auditory space. The listener is provided with an auditory space that may afford improved spatial segregation analogous to the phenomenon of auditory stream segregation (Bregman and Campbell, 1971; Bregman, 1990). The organization of the auditory space is performed by the listener; the algorithm only changes the auditory cues that describe source location.

Several multimicrophone signal-processing schemes have been proposed over the years for modifying the listner's perception of the auditory space. Mitchell *et al.* (1971) suggested a nonlinear signal-processing circuit that preserved the on-center speech while distorting off-center speech, thus rendering the on-center speech relatively more intelligible. While their technique exploits a multimicrophone input, the signal delivered to the listener is monaural. Durlach and Pang's (1986) philosophy is more closely aligned with our strategy. They propose a method for magnifying interaural differences by transforming the complex spectra (magnitude and phase) at each ear. The spectra are raised to powers in an attempt to increase the relative phase between the two ears. Although this operation effectively moves a single off-center source, it breaks down with multiple off-center sources. Increasing the relative phase between the signals received at the two ears does not guarantee that any of the sources will move further off-center when multiple sources are present.

The technique proposed here uses linear transformations of the signals at each ear to accomplish source movement. Given a desired space expansion function, a least-squares procedure is employed to design filters that

minimize the squared error between the actual and desired expansion functions. Note that, since the processing is linear, the superposition principle dictates that the output for multiple sources is the sum of the outputs due to each source alone. Thus this approach is equally effective with single or multiple sources. We derive the minimum mean-squared-error solution for unconstrained filters and for filters constrained to be finite impulse response (FIR). Perceptual studies and analysis of the binaural acoustic waveforms are utilized to demonstrate the effectiveness of this technique at moving the perceived location of sources. We only consider modification of interaural time delay; however, this approach is easily extended (in theory) to also include interaural amplitude differences. It is also straightforward to extend this technique to design filters when more than two microphones are available.

The paper is organized as follows. Section I formulates the least-squares filter design problem and discusses its solution in the frequency domain, corresponding to unconstrained filter design. This provides an upper bound on the mean-squared error performance of this method. The filters are constrained to be FIR and an algorithm is given for their design in Sec. II. Sections I and II discuss properties of the approximation error for the unconstrained and FIR cases and present an example. Perceptual experiments employed to measure the effectiveness of the algorithm on speech are described in Sec. III. Acoustic analyses of the experimental stimuli are presented in Sec. IV to support the results of the perceptual experiments. Conclusions are given in Sec. V.

I. LEAST-SQUARES FREQUENCY-DOMAIN INTERAURAL MAGNIFICATION

In the present work, we consider only manipulation of interaural time delays. It is well known that interaural time delay represents the primary localization clue for sound frequencies below 1 kHz (Kuhn, 1987). Interaural magnitude differences resulting from head shadowing and filtering by the pinna are the primary localization clue for frequencies above 1 kHz (Kuhn, 1987). Extension of this method to incorporate interaural magnitude differences is indicated.

Consider the situation depicted in Fig. 1. The sound source $s(t)$ is located at an angle of θ relative to the perpendicular between the two sensors. Assuming plane-wave propagation and omnidirectional sensors, the signal received at sensor 1 is $x_1(t) = s(t + \tau)$, while that at sensor 2 is $x_2(t) = s(t - \tau)$. Here, $\tau = d \sin \theta / 2c$ and we have refer-
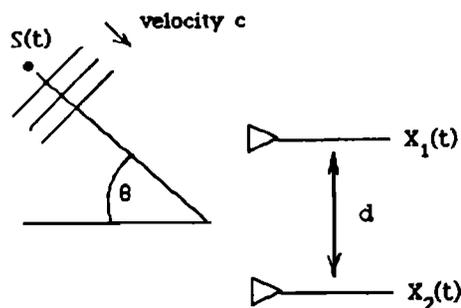
enced the time origin ($t = 0$) of $s(t)$ to the midpoint between the sensors. This implies that the interaural time delay T is $T = 2\tau$. In order to accomplish interaural magnification, we seek an algorithm that will produce output signals $y_1(t) = s[t + f(\tau)]$ and $y_2(t) = s[t - f(\tau)]$ in channels 1 and 2, respectively, where $f(\tau) > \tau$ is a function representing the interaural time delay expansion. A simple example is $f(\tau) = 2\tau$, which represents a doubling of interaural time delay. Note that it is possible to choose functions $f(\tau)$ which do not correspond to physically realizable angles. If we restrict $-90 < \theta < 90$, then restricting $-d/2c < f(\tau) < d/2c$ will ensure that the magnified time delay corresponds to a physical source location. Angles greater than 90 deg cannot be distinguished from those less than 90 deg using only time delay information.

Let the Fourier transform of $s(t)$ be given by $S(\omega)$. Time delays correspond to linear phase shifts in the frequency domain so $X_1(\omega) = S(\omega)e^{j\omega\tau}$, $X_2(\omega) = S(\omega)e^{-j\omega\tau}$, $Y_1(\omega) = S(\omega)e^{j\omega f(\tau)}$, $Y_2(\omega) = S(\omega)e^{-j\omega f(\tau)}$. Thus we desire an algorithm such that

$$S(\omega)\begin{bmatrix} e^{j\omega\tau} \\ e^{-j\omega\tau} \end{bmatrix} \rightarrow S(\omega)\begin{bmatrix} e^{j\omega f(\tau)} \\ e^{-j\omega f(\tau)} \end{bmatrix}. \qquad (1)$$

The class of algorithms that we shall consider consists of four linear filters $H_1(\omega), H_2(\omega), G_1(\omega)$, and $G_2(\omega)$, which are configured in a lattice section as illustrated in Fig. 2. Now,

$$\begin{bmatrix} Y_1(\omega) \\ Y_2(\omega) \end{bmatrix} = \begin{bmatrix} H_1(\omega) & G_1(\omega) \\ H_2(\omega) & G_2(\omega) \end{bmatrix}\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix}, \qquad (2)$$

so the single-input multiple-output transfer function resulting from propagation to the sensors and processing is

$$\mathbf{r}(\omega,\tau) = \begin{bmatrix} r_1(\omega,\tau) \\ r_2(\omega,\tau) \end{bmatrix} = \begin{bmatrix} H_1(\omega) & G_1(\omega) \\ H_2(\omega) & G_2(\omega) \end{bmatrix}\mathbf{d}(\omega,\tau), \qquad (3a)$$

where

$$\mathbf{d}(\omega,b) = \begin{bmatrix} e^{j\omega b} \\ e^{-j\omega b} \end{bmatrix}. \qquad (3b)$$

Notationally, boldface lower and upper case symbols represent vector and matrix quantities, respectively. The superscript -1 denotes matrix inverse, superscript H denotes complex conjugate transpose, and the superscript asterisk (*) denotes element by element complex conjugate. Note
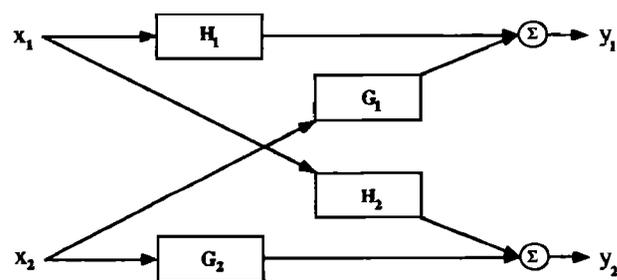

FIG. 1. Sound source relative to two sensors.

FIG. 2. Binaural filter network.

that $r(\omega,\tau)$ is termed the transfer function because the spectrum of the outputs is the product of $\mathbf{r}(\omega,\tau)$ and $S(\omega)$:

$$\begin{bmatrix} Y_1(\omega) \\ Y_2(\omega) \end{bmatrix} = S(\omega)\mathbf{r}(\omega,\tau).$$

The processing method proposed here is linear so the principle of superposition applies when multiple sources are present. Suppose there are two sources $S_1(\omega)$ and $S_2(\omega)$ located at angles that correspond to time delays τ_1 and τ_2 so

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = S_1(\omega)\mathbf{d}(\omega,\tau_1) + S_2(\omega)\mathbf{d}(\omega,\tau_2).$$

Using (2) and (3a) and (3b), we have

$$\begin{bmatrix} Y_1(\omega) \\ Y_2(\omega) \end{bmatrix} = S_1(\omega)\mathbf{r}(\omega,\tau_1) + S_2(\omega)\mathbf{r}(\omega,\tau_2).$$

Each source is processed as if it were the only source present; i.e., interaural magnification is applied independently to each source. This is a significant advantage over the nonlinear processing scheme proposed by Durlach and Pang (1986). Furthermore, the characteristics of the processing are independent of the source spectral content.

In general, we do not expect to find filters that implement the mapping (1) exactly over arbitrary ranges of ω, τ, with arbitrary $f(\tau)$. Thus we consider choosing H_1, G_1, H_2, and G_2 such that $\mathbf{r}(\omega,\tau) \approx \mathbf{d}[\omega, f(\tau)]$. This problem is solved here using a minimum mean-squared error criterion. Define the errors $e_1(\omega,\tau)$ and $e_2(\omega,\tau)$ as

$$\begin{bmatrix} e_1(\omega,\tau) \\ e_2(\omega,\tau) \end{bmatrix} = \mathbf{d}[\omega, f(\tau)] - \begin{bmatrix} H_1(\omega) & G_1(\omega) \\ H_2(\omega) & G_2(\omega) \end{bmatrix}\mathbf{d}(\omega,\tau). \tag{4}$$

Modification of (4) to design filters that effect interaural magnitude differences is accomplished by replacing $\mathbf{d}[\omega, f(\tau)]$ with

$$\begin{bmatrix} a_1[\omega, f(\tau)]e^{j\omega f(\tau)} \\ a_2[\omega, f(\tau)]e^{-j\omega f(\tau)} \end{bmatrix},$$

where $a_i[\omega, f(\tau)]$, $i = 1,2$, is the desired interaural magnitude in the ith channel as a function of frequency and warped direction. As stated previously, in the work presented here, we assume $a_i[\omega, f(\tau)] = 1$.

Let $\tau_a \leqslant \tau \leqslant \tau_b$ represent the sector of directions to be expanded. We seek filters that solve

$$\min_{H_i, G_i} \int_{\tau_a}^{\tau_b} |e_i(\omega,\tau)|^2 d\tau, \quad i = 1,2. \tag{5}$$

Note that a non-negative weighting function can be inserted into the integrand to emphasize some frequencies and/or directions over others. Equation (5) represents a standard least-squares problem with solution

$$[H_i(\omega) \quad G_i(\omega)] = \mathbf{p}_i(\omega)\mathbf{R}^{-1}(\omega), \quad i = 1,2, \tag{6a}$$

where the matrix $\mathbf{R}(\omega)$ and vectors $\mathbf{p}_i(\omega)$, $i = 1,2$ are given by

$$\mathbf{R}(\omega) = \int_{\tau_a}^{\tau_b} \mathbf{d}(\omega,\tau)\mathbf{d}^H(\omega,\tau)d\tau, \tag{6b}$$

$$\mathbf{p}_1(\omega) = \int_{\tau_a}^{\tau_b} e^{j\omega f(\tau)}\mathbf{d}^H(\omega,\tau)d\tau,$$

$$\mathbf{p}_2(\omega) = \int_{\tau_a}^{\tau_b} e^{-j\omega f(\tau)}\mathbf{d}^H(\omega,\tau)d\tau. \tag{6c}$$

The squared error for the solution in (6a) is given by

$$\int_{\tau_a}^{\tau_b} |e_i(\omega,\tau)|^2 d\tau = \tau_b - \tau_a - \mathbf{p}_i(\omega)\mathbf{R}^{-1}(\omega)\mathbf{p}_i^H(\omega). \tag{7}$$

It is straightforward to show that $e_1(\omega,\tau)$ and $e_2(\omega,\tau)$ are related and that $[H_1(\omega)G_1(\omega)]$ and $[H_2(\omega)G_2(\omega)]$ are related. Express $e_1(\omega,\tau)$ as

$$e_1(\omega,\tau) = e^{j\omega f(\tau)} - e^{j\omega\tau}H_1(\omega) - e^{-j\omega\tau}G_1(\omega). \tag{8}$$

Note that

$$e_1^*(\omega,\tau) = e^{-j\omega f(\tau)} - e^{j\omega\tau}G_1^*(\omega) - e^{-j\omega\tau}H_1^*(\omega) \tag{9}$$

is of the same form as

$$e_2(\omega,\tau) = e^{-j\omega f(\tau)} - e^{j\omega\tau}H_2(\omega) - e^{-j\omega\tau}G_2(\omega). \tag{10}$$

Thus, if $H_1(\omega)$ and $G_1(\omega)$ minimize (5) for $i = 1$, then (9) and (10) indicate that $H_2(\omega) = G_1^*(\omega)$ and $G_2(\omega) = H_1^*(\omega)$ minimize (5) for $i = 2$ since $|e_2(\omega,\tau)|^2 = |e_1^*(\omega,\tau)|^2 = |e_1(\omega,\tau)|^2$.

This also implies $r_2(\omega,\tau) = r_1^*(\omega,\tau)$; thus, the magnitude distortion is identical in each ear, while the phase distortion has the opposite sign in each ear. Equal transfer function magnitudes indicate that the signal processing does not artificially introduce a head shadow effect where the signal at one ear is louder than the other. Deviation of the desired transfer functions from unity simply appears as equal spectral distortion of the signal in each channel. The opposite sign relationship between the transfer function phases mimics the relationship between the desired transfer function phases, $\omega f(\tau)$ and $-\omega f(\tau)$. If the transfer function phases are nonlinear in ω for fixed τ, then the sound at one frequency appears to arrive from one direction, while the sound at another frequency appears to arrive from another direction. However, the opposite sign relationship ensures that the sound at each frequency appears to arrive from physically realizable directions.

It can be shown that $H_1(\omega)$, $H_2(\omega)$, $G_1(\omega)$, and $G_2(\omega)$ are real (zero phase) if $\tau_b = -\tau_a$ and $f(\tau)$ is an odd function. The following example illustrates this characteristic.

Assume that $c = 345$ m/s, $d = 17$ cm, $-90 < \theta < 90$, and $f(\tau) = 2\tau$. The maximum interaural delay in this case is $T_{\max} = 500$ μs. Thus set $\tau_b = -\tau_a = 125$ μs. This represents a linear expansion of a symmetric sector into the full range of physically possible τ. The solutions for $H_1(\omega)$ and $G_1(\omega)$ as given in (6a) are plotted in Fig. 3 for frequencies up to 1 kHz. From the previous results in this section, we have $H_2(\omega) = G_1(\omega)$ and $G_2(\omega) = H_1(\omega)$. Figure 4(a) and (b) depicts the magnitude and phase of $r_1(\omega,\tau)$ for several values of τ. Ideally, the magnitude of $r_1(\omega,\tau)$ would be unity and the phase $2\omega\tau$. Note that both the magnitude and the phase deviate from the ideal. Recall that $r_1(\omega,\tau) = r_2^*(\omega,\tau)$ so the magnitude distortion is identical in both ears, while the phase distortion is opposite in sign.
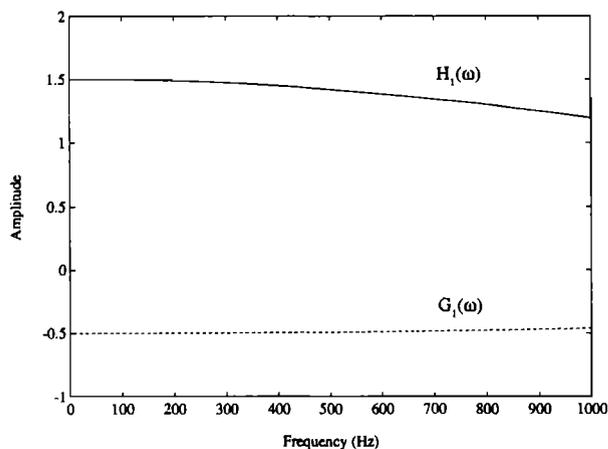
FIG. 3. Plot of $H_1(\omega)$ and $G_1(\omega)$ for linear expansion by a factor of 2.

Except for $\tau = 100\,\mu s$ and $\tau = 125\,\mu s$, the phase is approximately linear in frequency, which indicates that the transformed source does not undergo spreading in space; it remains compact. Nonlinear phase indicates the sound at one frequency appears to arrive from one direction while the
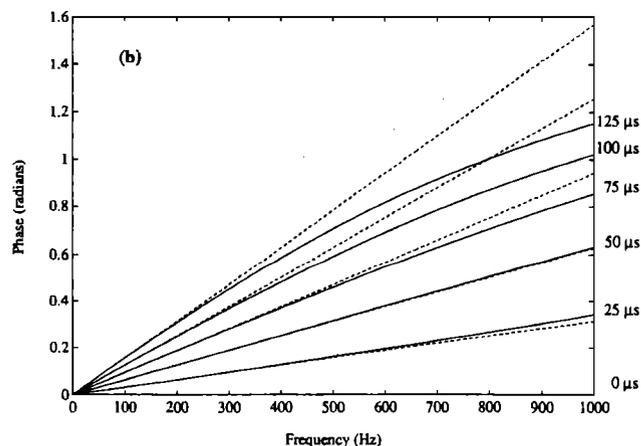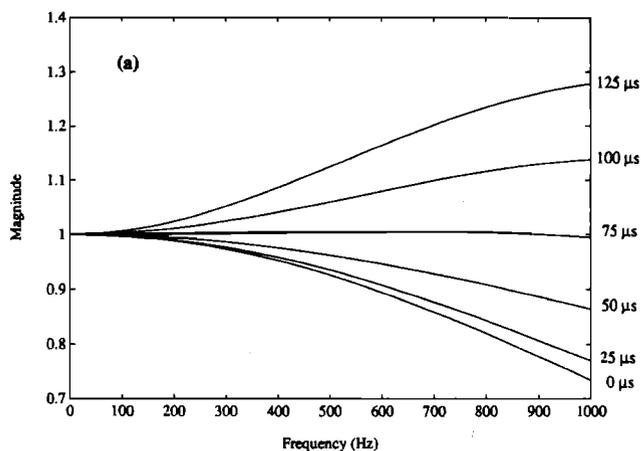

FIG. 4. (a) Magnitude of $r_1(\omega,\tau)$ as a function of frequency for several values of τ. (b) Ideal (- - -) and actual (—) phase of $r_1(\omega,\tau)$ as a function of frequency for several values of τ; $H_1(\omega)$ and $G_1(\omega)$ are designed for linear expansion by a factor of 2 and are depicted in Fig. 3.

sound at another frequency arrives from another direction. The phase response is converted to interaural time delay as a function of frequency in Fig. 5 for several values of τ. The desired interaural time delay is constant over frequency and is given by $T = 2f(\tau) = 4\tau$.

This approach is easily extended to utilize more than two microphones. For example, if there are four microphones available, then eight filters would be used to form a four-input two-output system. For the purposes of illustration, assume that four microphones are arranged in a linear equal-spaced geometry and that the midpoint of the array is the zero time reference for $s(t)$. The signals received at the four microphones are then given by $s(t + 3\tau)$, $s(t + \tau)$, $s(t - \tau)$, and $s(t - 3\tau)$. Following the reasoning used to obtain (3a), we have

$$\mathbf{r}(\omega,\tau) = \begin{bmatrix} H_1(\omega) & G_1(\omega) & F_1(\omega) & E_1(\omega) \\ H_2(\omega) & G_2(\omega) & F_2(\omega) & E_2(\omega) \end{bmatrix} \mathbf{d}_4(\omega,\tau),$$
(11a)

where

$$\mathbf{d}_4(\omega,\tau) = \begin{bmatrix} e^{j3\omega\tau} \\ e^{j\omega\tau} \\ e^{-j\omega\tau} \\ e^{-j3\omega\tau} \end{bmatrix}.$$
(11b)

Define errors

$$\begin{bmatrix} e_1(\omega,\tau) \\ e_2(\omega,\tau) \end{bmatrix}$$

$$= \mathbf{d}[\omega, f(\tau)]$$

$$- \begin{bmatrix} H_1(\omega) & G_1(\omega) & F_1(\omega) & E_1(\omega) \\ H_2(\omega) & G_2(\omega) & F_2(\omega) & E_2(\omega) \end{bmatrix} \mathbf{d}_4(\omega,\tau).$$
(12)

Minimizing the mean-squared errors as in (5) produces a solution of the same form as (6a):

$$[H_i(\omega) \quad G_i(\omega) \quad F_i(\omega) \quad E_i(\omega)]$$
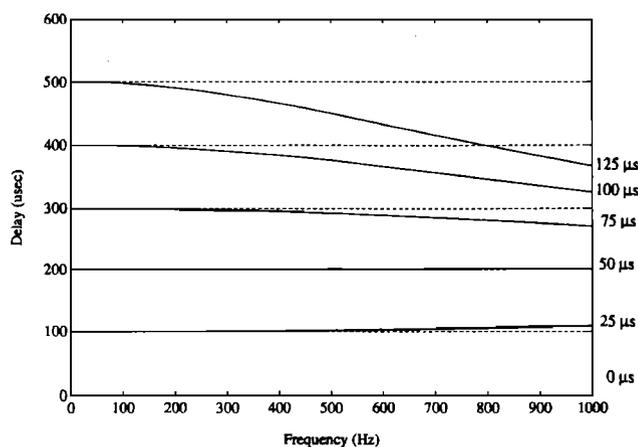$$= \mathbf{p}_i(\omega)\mathbf{R}^{-1}(\omega), \quad i = 1,2,$$
(13a)

FIG. 5. Ideal (- - -) and actual (—) interaural time delay as a function of frequency for several values of τ; $H_1(\omega)$ and $G_1(\omega)$ are designed for linear expansion by a factor of 2 and are depicted in Fig. 3.

where now the matrix $\mathbf{R}(\omega)$ and the vectors $\mathbf{p}_i(\omega)$, $i=1,2$ are given by

$$\mathbf{R}(\omega) = \int_{\tau_a}^{\tau_b} \mathbf{d}_4(\omega,\tau)\mathbf{d}_4^H(\omega,\tau)d\tau, \tag{13b}$$

$$\mathbf{p}_1(\omega) = \int_{\tau_a}^{\tau_b} e^{j\omega f(\tau)}\mathbf{d}_4^H(\omega,\tau)d\tau, \tag{13c}$$

$$\mathbf{p}_2(\omega) = \int_{\tau_a}^{\tau_b} e^{-j\omega f(\tau)}\mathbf{d}_4^H(\omega,\tau)d\tau.$$

The solution for an arbitrary geometry and number of microphones is of the same form as (13a) with the appropriately defined time delay vector replacing $\mathbf{d}_4(\omega,\tau)$. Use of additional microphones offers the potential for substantially more accurate warping of the auditory space at the expense of additional computational burden.

II. LEAST-SQUARES FIR FILTER SOLUTION

The frequency responses obtained in the previous section are not directly amenable to implementation with existing hardware. Thus, in this section, we constrain each of the filters to be FIR with $P+1$ coefficients. One could design FIR filters to approximate the frequency responses obtained in Sec. I using established FIR filter design techniques. However, instead, we reformulate the least-squares problem to optimize the FIR filters directly. This results in less error than a two-stage approximation approach.

The frequency response $H(\omega)$ of an FIR filter whose impulse response is represented by the $P+1$ vector $\mathbf{h}$ is given by

$$H(\omega) = \mathbf{v}^H(\omega)\mathbf{h}, \tag{14}$$

where

$$\mathbf{v}^H(\omega) = [1 \quad e^{-j\omega} \quad e^{-j2\omega} \ldots e^{-jP\omega}].$$

Let $\mathbf{h}_1, \mathbf{h}_2, \mathbf{g}_1$, and $\mathbf{g}_2$ be length $P+1$ vectors whose elements represent the impulse response sequences of filters $H_1(\omega)$, $H_2(\omega)$, $G_1(\omega)$, and $G_2(\omega)$ in Fig. 2. Analogous to (4), define errors $e_1(\omega,\tau)$ and $e_2(\omega,\tau)$ as

$$\begin{bmatrix} e_1(\omega,\tau) \\ e_2(\omega,\tau) \end{bmatrix} = e^{j\phi(\omega)}\mathbf{d}[\omega, f(\tau)]$$

$$- \begin{bmatrix} \mathbf{v}^H(\omega)\mathbf{h}_1 & \mathbf{v}^H(\omega)\mathbf{g}_1 \\ \mathbf{v}^H(\omega)\mathbf{h}_2 & \mathbf{v}^H(\omega)\mathbf{g}_2 \end{bmatrix}\mathbf{d}(\omega,\tau). \tag{15}$$

The phase factor $e^{j\phi(\omega)}$ is included in the desired response in (15) to reduce the approximation error when the filters are constrained to be causal. This is not necessary for the problem posed in the previous section because the frequency responses of the optimum filters are unconstrained. However, with causal FIR filters the phase must be negative, and a large error results when a causal FIR filter is used to approximate zero or positive phase. Recall from the previous section that an entire class of unconstrained filters have zero phase. The phase factor in (15) does not change the relative phase between the two channels and thus does not affect perceived direction. If $\phi(\omega)$ is a linear function of ω, then it corresponds to introducing an identical time delay in each

channel. Throughout this paper, we only consider linear $\phi(\omega)$.

For practical applications we require that $\mathbf{h}_i$ and $\mathbf{g}_i$, $i=1,2$, be real. Re-express $e_1(\omega,\tau)$ as

$$e_1(\omega,\tau) = e^{j[\phi(\omega) + \omega f(\tau)]} - \mathbf{d}_v^H(\omega,\tau)\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{g}_1 \end{bmatrix}, \tag{16}$$

where $\mathbf{d}_v^H(\omega,\tau) = [e^{j\omega\tau}\mathbf{v}^H(\omega) \quad e^{-j\omega\tau}\mathbf{v}^H(\omega)]$. Since the elements of $\mathbf{d}_v^H(\omega,\tau)$ are complex exponentials, $\mathbf{d}_v^H(\omega,\tau)$ can be expressed as the sum of a vector of cosines and a vector of sines:

$$\mathbf{d}_v^H(\omega,\tau) = \mathbf{c}_v^H(\omega,\tau) + j\mathbf{s}_v^H(\omega,\tau); \tag{17}$$

i.e., $\mathbf{c}_v^H(\omega,\tau) = \text{real}\{\mathbf{d}_v^H(\omega,\tau)\}$ and $j\mathbf{s}_v^H(\omega,\tau) = \text{imaginary}\{\mathbf{d}_v^H(\omega,\tau)\}$. Substituting into (16) and assuming $\mathbf{h}_1$ and $\mathbf{g}_1$ are real yield

$$e_1(\omega,\tau) = \left\{ \cos[\phi(\omega) + \omega f(\tau)] - \mathbf{c}_v^H(\omega,\tau)\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{g}_1 \end{bmatrix} \right\}$$

$$+ j\left\{ \sin[\phi(\omega) + \omega f(\tau)] - \mathbf{s}_v^H(\omega,\tau)\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{g}_1 \end{bmatrix} \right\}. \tag{18a}$$

Similarly, $e_2(\omega,\tau)$ is expressed as

$$e_2(\omega,\tau) = \left\{ \cos[\phi(\omega) - \omega f(\tau)] - \mathbf{c}_v^H(\omega,\tau)\begin{bmatrix} \mathbf{h}_2 \\ \mathbf{g}_2 \end{bmatrix} \right\}$$

$$+ j\left\{ \sin[\phi(\omega) - \omega f(\tau)] - \mathbf{s}_v^H(\omega,\tau)\begin{bmatrix} \mathbf{h}_2 \\ \mathbf{g}_2 \end{bmatrix} \right\}. \tag{18b}$$

Minimizing the average-squared error over a band of frequencies $\omega_a \leqslant \omega \leqslant \omega_b$ and sector of directions $\tau_a \leqslant \tau \leqslant \tau_b$, we seek filters that solve

$$\min_{\mathbf{h}_i,\mathbf{g}_i} \int_{\omega_a}^{\omega_b} \int_{\tau_a}^{\tau_b} |e_i(\omega,\tau)|^2\,d\tau\,d\omega, \quad i=1,2. \tag{19}$$

In general, the integrals in (19) cannot be evaluated in closed form. Thus we approximate (19) by minimizing the error over a discrete set of points ω_k, $k=1,2,\ldots,K$, τ_m, $m=1,2,\ldots,M$, where $\omega_a \leqslant \omega_k \leqslant \omega_b$ and $\tau_a \leqslant \tau_m \leqslant \tau_b$:

$$\min_{\mathbf{h}_i,\mathbf{g}_i} \sum_{k=1}^{K} \sum_{m=1}^{M} |e_i(\omega_k,\tau_m)|^2, \quad i=1,2. \tag{20}$$

The difference between the solutions to (19) and (20) can be made arbitrarily small by choosing K and M sufficiently large. In general, we assume $KM > 2P + 2$. Substitute (18a) and (18b) and rewrite (20) in matrix form as

$$\min_{\mathbf{h}_i,\mathbf{g}_i} \left\| \mathbf{a}_i - \mathbf{D}\begin{bmatrix} \mathbf{h}_i \\ \mathbf{g}_i \end{bmatrix} \right\|_2^2, \quad i=1,2. \tag{21}$$

Here,

$$\mathbf{a}_1 = \begin{bmatrix} \cos[\phi(\omega_1)+\omega_1 f(\tau_1)] \\ \sin[\phi(\omega_1)+\omega_1 f(\tau_1)] \\ \cos[\phi(\omega_2)+\omega_2 f(\tau_1)] \\ \sin[\phi(\omega_2)+\omega_2 f(\tau_1)] \\ \vdots \\ \cos[\phi(\omega_K)+\omega_K f(\tau_1)] \\ \sin[\phi(\omega_K)+\omega_K f(\tau_1)] \\ \cos[\phi(\omega_1)+\omega_1 f(\tau_2)] \\ \sin[\phi(\omega_1)+\omega_1 f(\tau_2)] \\ \vdots \\ \cos[\phi(\omega_K)+\omega_K f(\tau_2)] \\ \sin[\phi(\omega_K)+\omega_K f(\tau_2)] \\ \vdots \\ \cos[\phi(\omega_K)+\omega_K f(\tau_M)] \\ \sin[\phi(\omega_K)+\omega_K f(\tau_M)] \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} \cos[\phi(\omega_1)-\omega_1 f(\tau_1)] \\ \sin[\phi(\omega_1)-\omega_1 f(\tau_1)] \\ \cos[\phi(\omega_2)-\omega_2 f(\tau_1)] \\ \sin[\phi(\omega_2)-\omega_2 f(\tau_1)] \\ \vdots \\ \cos[\phi(\omega_K)-\omega_K f(\tau_1)] \\ \sin[\phi(\omega_K)-\omega_K f(\tau_1)] \\ \cos[\phi(\omega_1)-\omega_1 f(\tau_2)] \\ \sin[\phi(\omega_1)-\omega_1 f(\tau_2)] \\ \vdots \\ \cos[\phi(\omega_K)-\omega_K f(\tau_2)] \\ \sin[\phi(\omega_K)-\omega_K f(\tau_2)] \\ \vdots \\ \cos[\phi(\omega_K)-\omega_K f(\tau_M)] \\ \sin[\phi(\omega_K)-\omega_K f(\tau_M)] \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{c}_v^H(\omega_1,\tau_1) \\ \mathbf{s}_v^H(\omega_1,\tau_1) \\ \mathbf{c}_v^H(\omega_2,\tau_1) \\ \mathbf{s}_v^H(\omega_2,\tau_1) \\ \vdots \\ \mathbf{c}_v^H(\omega_K,\tau_1) \\ \mathbf{s}_v^H(\omega_K,\tau_1) \\ \mathbf{c}_v^H(\omega_1,\tau_2) \\ \mathbf{s}_v^H(\omega_1,\tau_2) \\ \vdots \\ \mathbf{c}_v^H(\omega_K,\tau_2) \\ \mathbf{s}_v^H(\omega_K,\tau_2) \\ \vdots \\ \mathbf{c}_v^H(\omega_K,\tau_M) \\ \mathbf{s}_v^H(\omega_K,\tau_M) \end{bmatrix}. \quad (22)$$

Equation (21) represents a standard least squares problem involving $2KM$ equations and $2P+2$ unknowns (Lawson and Hanson, 1974). Let the matrix $\mathbf{D}$ have singular value decomposition (SVD)

$$\mathbf{D} = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^H, \qquad (23)$$

where r is the rank of $\mathbf{D}$, $\mathbf{u}_k$ and $\mathbf{v}_k$ are the left and right kth singular vectors, and σ_k is the kth singular value of $\mathbf{D}$. The solution to (21) is expressed in terms of the SVD as

$$\begin{bmatrix} \mathbf{h}_i \\ \mathbf{g}_i \end{bmatrix} = \sum_{k=1}^{r} \sigma_k^{-1} \mathbf{v}_k \mathbf{u}_k^H \mathbf{a}_i, \quad i=1,2. \qquad (24)$$

In general, r depends on ω_k, $k=1,2,...,K$, and τ_m, $j=1,2,...,M$; however, it is always less than or equal to $\min\{2P+2, 2KM\}$. The mean-squared error using the solution (24) is given by

$$\min_{\mathbf{h}_i,\mathbf{g}_i} \left\| \mathbf{a}_i - \mathbf{D}\begin{bmatrix} \mathbf{h}_i \\ \mathbf{g}_i \end{bmatrix} \right\|_2^2$$
$$= \left\| \mathbf{a}_i\left(\mathbf{I} - \sum_{k=1}^{r} \mathbf{u}_k \mathbf{u}_k^H \right) \right\|_2^2, \quad i=1,2. \qquad (25)$$

This is the projection of $\mathbf{a}_i$ onto the $2KM - r$ dimensional null space of $\mathbf{D}^T$.

In practice, numerical evaluation of the SVD will not yield exactly zero singular values, so we choose r based on the number of singular values that exceed a specified threshold. The solution in (24) is dependent on the inverse of the singular values. Thus very small singular values tend to produce numerically large coefficients in $\mathbf{h}_i$ and $\mathbf{g}_i$. Note that (25) indicates that reducing r increases the error, so there is a trade-off between coefficient norm and error.

There is a relationship between the four FIR filters similar to the relationship between the unconstrained filters in Sec. I. Assume that P is even and let $\phi(\omega) = -P\omega/2$. Express $e_1(\omega,\tau)$ as

$$e_1(\omega,\tau) = e^{-jP\omega/2}[e^{j\omega f(\tau)} - e^{j\omega\tau}\widetilde{H}_1(\omega)$$
$$- e^{-j\omega\tau}\widetilde{G}_1(\omega)], \qquad (26)$$

where we define $\widetilde{H}_i(\omega) = e^{jP\omega/2}H_i(\omega)$ and $\widetilde{G}_i(\omega) = e^{jP\omega/2}G_i(\omega)$. Now,

$$e_1^*(\omega,\tau) = e^{-jP\omega/2}[e^{-j\omega f(\tau)} - e^{j\omega\tau}\widetilde{G}_1^*(\omega)$$
$$- e^{-j\omega\tau}\widetilde{H}_1^*(\omega)], \qquad (27)$$

which is of the same form as

$$e_2(\omega,\tau) = e^{-jP\omega/2}[e^{-j\omega f(\tau)} - e^{j\omega\tau}\widetilde{H}_2(\omega)$$
$$- e^{-j\omega\tau}\widetilde{G}_2(\omega)]. \qquad (28)$$

Thus, if $H_1(\omega)$ and $G_1(\omega)$ minimize the mean-squared value of $e_1(\omega,\tau)$, then

$$\widetilde{H}_2(\omega) = \widetilde{G}_1^*(\omega) \qquad (29a)$$

and

$$\widetilde{G}_2(\omega) = \widetilde{H}_1^*(\omega) \qquad (29b)$$

minimize the mean-squared value of $e_2(\omega,\tau)$. Equations (29a) and (29b) imply

$$H_2(\omega) = e^{-j\omega P}G_1^*(\omega) \qquad (30a)$$

and

$$G_2(\omega) = e^{-j\omega P}H_1^*(\omega). \qquad (30b)$$

These solutions are only valid if $H_2(\omega)$ and $G_2(\omega)$ correspond to length P FIR filters. Let the inverse discrete Fourier transforms of $H_1(\omega)$ and $G_1(\omega)$ be sequences $h_1(n)$ and $g_1(n)$. Note that the values $h_1(n)$ and $g_1(n)$, $0 \leqslant n \leqslant P$, are represented by the elements of $\mathbf{h}_1$ and $\mathbf{g}_1$; all the other values in the sequence are zero; $h_1(n)$ and $g_1(n)$ are real-valued sequences so complex conjugation in the frequency domain corresponds to reflection about zero in the sequence domain. Multiplication by $e^{-j\omega P}$ corresponds to a delay of P samples in the sequence domain so (30a) and (30b) imply

$$h_2(n) = g_1(P-n) \qquad (31a)$$

and

$$g_2(n) = h_1(P-n). \qquad (31b)$$

Thus $h_2(n)$ and $g_2(n)$ are nonzero only on $0 \leqslant n \leqslant P$ so $H_2(\omega)$ and $G_2(\omega)$ as given by (30a) and (30b) correspond to valid FIR filters. In matrix notation,

$$\mathbf{h}_2 = \mathbf{J}\mathbf{g}_1, \tag{32a}$$

and

$$\mathbf{g}_2 = \mathbf{J}\mathbf{h}_1, \tag{32b}$$

where $\mathbf{J}$ is the matrix that turns vectors "upside down":

$$\mathbf{J} = \begin{bmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{bmatrix} \tag{33}$$

(ones on the cross diagonal and zeros elsewhere).

As with the unconstrained filters in the previous section, the relationships between $\mathbf{h}_1$, $\mathbf{g}_1$ and $\mathbf{h}_2$, $\mathbf{g}_2$ imply that $r_1(\omega,\tau) = r_2^*(\omega,\tau)$; i.e., the transfer functions to each ear have the same magnitude and their phases have opposite sign. This indicates that the signal processing does not artificially introduce an interaural magnitude difference; deviation of the desired transfer functions from unity results in equal spectral distortion of the signal in each ear.

The example from the previous section is repeated here to illustrate design of filters that are constrained to be FIR. Thus $c = 345$ m/s, $d = 17$ cm, $-90 \leqslant \theta \leqslant 90$, $f(\tau) = 2\tau$, and $\tau_b = -\tau_a = 125\,\mu$s. Recall that this represents a linear expansion by a factor of 2 of a symmetric sector into the full range of physically possible τ. For this example, we have chosen $\omega_a = 0$, $\omega_b = 2000\pi$, $P = 32$ (impulse response length 33), $\phi(\omega) = -16\omega$, $M = 51$ (samples in ω), and $K = 51$ (samples in τ). The rank of $\mathbf{D}$ is determined by the number of singular values that exceed 2.2204×10^{-16}. Figure 6(a) and (b) depicts the magnitude and phase responses of $r_1(\omega,\tau)$ [see (3a)] as a function of frequency for several values of τ. Comparison with the magnitude and phase response obtained using the unconstrained frequency domain solution illustrated in Fig. 4(a) and (b) indicates almost negligible performance loss associated with constraining the filters to be FIR. Note that the optimal filter frequency responses depicted in Fig. 3 are very smooth and thus are well approximated with an FIR filter.

The interaural delay as a function of frequency computed from the phase response is shown in Fig. 7 for several values of τ. Several values of τ are chosen to illustrate the effect of the expansion algorithm on a source that arrives outside the design region ($\tau \leqslant 125\,\mu$s). While the deviation from ideal increases for $\tau > 125\,\mu$s, the phase distortion degrades gracefully and monotonically. Note that interaural delays greater than $500\,\mu$s do not correspond to physically possible source locations.

As a second example, we design FIR filters to effect a factor of 2 expansion over the one-sided sector defined by $\tau_a = 0$, $\tau_b = 125\,\mu$s. All the other parameters are the same as in the previous example. The magnitude and phase responses of $r_1(\omega,\tau)$ [see (3a)] are compared to the ideal responses in Fig. 8(a) and (b) as a function of frequency for several values of τ. The deviation from ideal is considerably smaller at larger τ than that of the previous example, especially for frequencies below 900 Hz. This is a result in part of performing the expansion over one-half the angular region
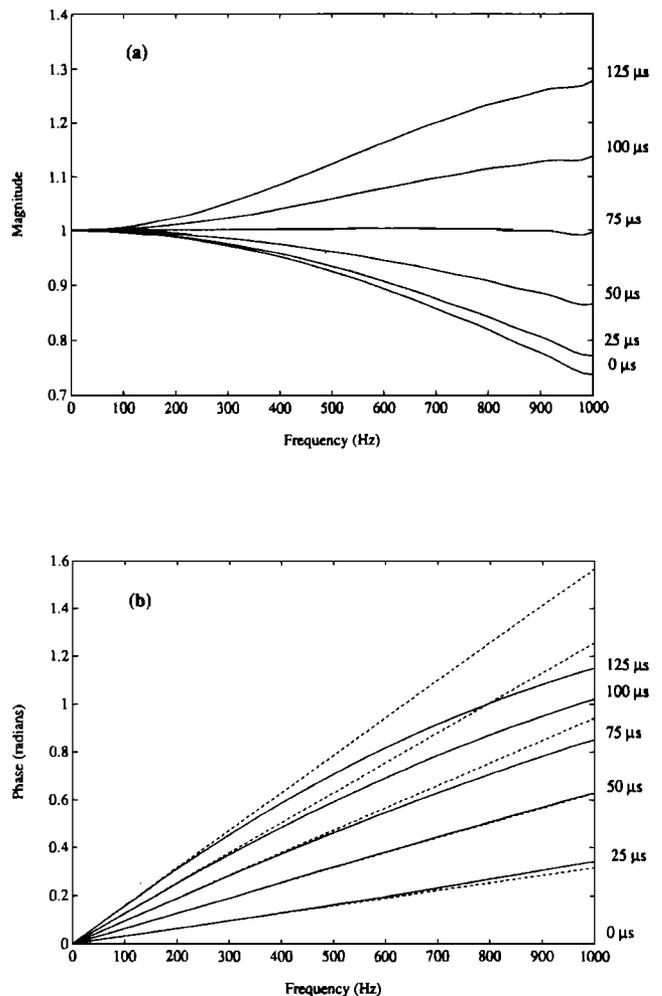

FIG. 6. Magnitude and phase responses for several values of τ within the region used to design FIR filters. The filters are designed to effect a factor of 2 expansion for $|\tau| \leqslant 125\,\mu$s. (a) Magnitude of $r_1(\omega,\tau)$ as a function of frequency for several values of τ. (b) Ideal (- - -) and actual (—) phase of $r_1(\omega,\tau)$ as a function of frequency for several values of τ.
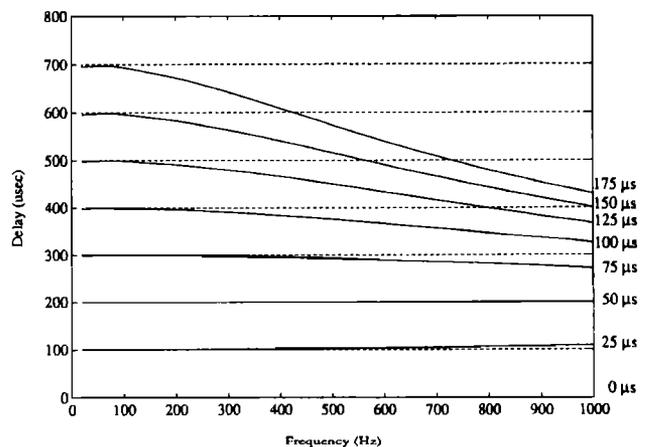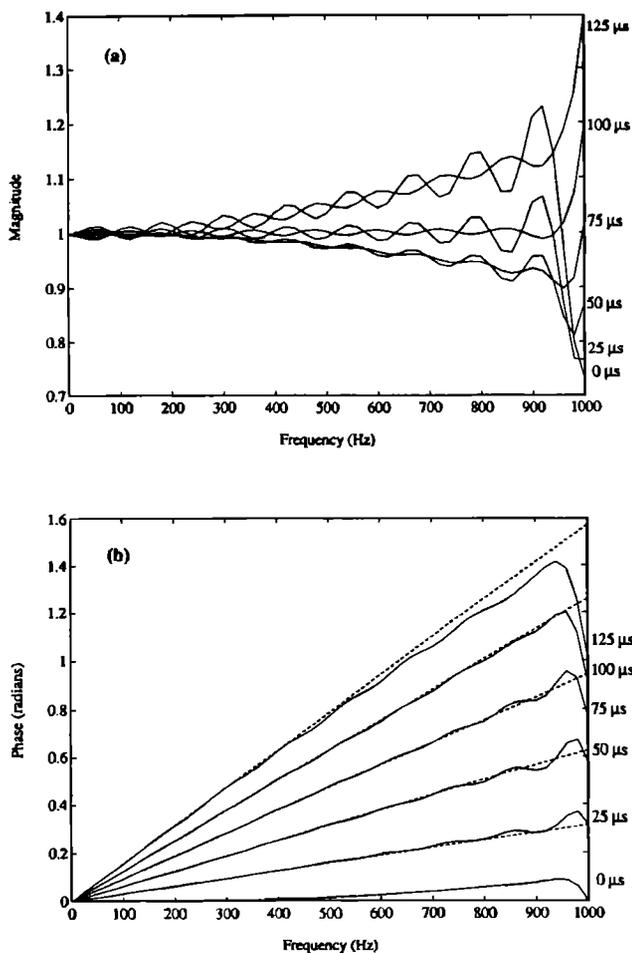

FIG. 7. Ideal (- - -) and actual (—) interaural time delay as a function of frequency for several values of τ computed from the phase response in Fig. 6, including τ's sampled outside of the region ($\tau > 125\,\mu$s) used to design the FIR filters.

FIG. 8. Magnitude and phase responses for several values of τ within the region used to design FIR filters. The filters are designed to effect a factor of 2 expansion for $0 < \tau < 125$ μs. (a) Magnitude of $r_1(\omega,\tau)$ as a function of frequency for several values of τ. (b) Ideal (- - -) and actual (—) phase of $r_1(\omega,\tau)$ as a function of frequency for several values of τ.

used in the previous example. In general, deviation between actual and desired response in both the unconstrained and FIR designs is dependent on $f(\tau)$ and the region that is to be expanded. Determining functions $f(\tau)$ that are easy and difficult to approximate is of interest; however, it is unlikely that any such determination can be done analytically due to the manner in which $f(\tau)$ enters into the expression for the error.

The FIR filter solutions for more than two microphones are obtained using the same approach as that given in Sec. I.

III. PERCEPTUAL EVALUATION OF AUDITORY EXPANSION ON SPEECH

This section describes a preliminary perceptual experiment used to evaluate the effect of the space expansion algorithm. Listeners were asked to perform a lateralization task with and without the auditory expansion algorithm in effect. Below is an explanation of the experimental setup, followed by presentation of results.

A. Stimuli

The stimuli were constructed by mixing digitized natural speech recorded from a male and female speaker with 16-

bit resolution at a 20-kHz sampling rate. Each speaker produced /pa/ in an IAC sound-attenuated chamber with microphone. The tokens were equalized in loudness by normalizing the rms to 64 dB re: 1 least significant bit. The male speaker served as the reference sound source "located" at center with an interaural delay of 0 μs. Manipulation of the interaural delays shifted the female toward the observer's right periphery in steps of 50-μs delays for the first five steps and 100 μs for the remaining five steps to a maximum delay of 750 μs. The stimuli were constructed off-line and loaded into a 512 KB RAM buffer on a Ariel DSP-16 signal-processing board for delivery to the subject. The DSP-16 was under the control of an 80386-based computer that coordinated stimulus output and response collection. The tokens were low-pass filtered using a cutoff frequency of 1 kHz (98 dB/oct rolloff) in order to evaluate the processing of only interaural time cues. Despite this rather severe filtering, the tokens were still perceived as the utterance /pa/.

B. Subjects

Five female and seven male normal-hearing volunteers served as subjects. None of the listeners had prior experience performing psychoacoustic localization tasks. Prior to presentation of the experiment, subjects were allowed to practice the task for five trials.

C. Procedure

As noted above, subjects were asked to listen to monosyllable speech presented at a comfortable listening level over Beyer DT100 headphones. The stimuli were constructed to simulate a male and female speaking the utterance /pa/ simultaneously from various locations in the horizontal plane. The subjects were instructed to manipulate the position of the female until the two speakers appeared to separate spatially. This was accomplished by depressing one of two buttons on a response device (PC mouse) that "moved" the female in the horizontal plane. The subjects were instructed to "vote" when they had manipulated the speakers' relative position such that their auditory spatial location in the horizontal plane appeared to be just noticeably separate. This method of adjustment criterion served to evaluate the degree to which the space expansion algorithm afforded better discrimination of two sound sources overlapping in frequency and phonetic content. The filters were designed to effect a factor of 2 expansion [$f(\tau) = 2\tau$, $|\tau| \leqslant 125$ μs]. The signal-processing algorithm, based on the binaural filter network (Fig. 2), was programmed in M56001 assembly code for execution on a Spectrum M56001-based signal-processing board. The subjects received two signal processing conditions, either the algorithm present or not. Each condition was replicated eight times for a total of 16 trials.

D. Results

Mean minimal time delay separations for the 12 subjects are shown in Fig. 9 for phase expanded and unprocessed conditions, as labeled on the abscissa. There was a factor of 2 reduction in the minimal delay required for spatial separation. This indicates that the algorithm is effective at moving
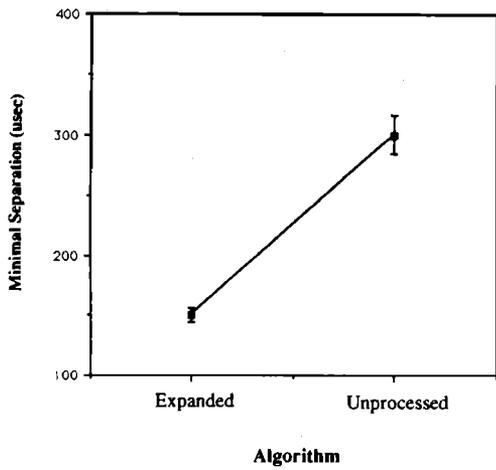
FIG. 9. Minimal discernable separation between the male and female speakers. Mean and standard error bars are plotted for $N = 12$ subjects.

the perceived location of female speaker by doubling the interaural delay. This conclusion is supported by the analysis of the speech waveforms given in the following section.

IV. ACOUSTIC ANALYSIS OF AUDITORY EXPANSION ON SPEECH

In order to assess the acoustic effects of the algorithm, the frequency responses of the transfer functions between the input signals x_1 and x_2 and the processed output signals y_1 and y_2 were computed (see Bendat and Piersol, 1971). Suppose $X_2(\omega) = X_1(\omega)H_x(\omega)$ and $Y_2(\omega) = Y_1(\omega)H_y(\omega)$. Note that, for a single source, $H_x(\omega)$ and $H_y(\omega)$ ideally correspond to the frequency response of a time delay. However, in the general case of multiple sources, they represent the combined effects of multiple time delays; $H_x(\omega)$ and $H_y(\omega)$ are evaluated as

$$H_x(\omega) = \frac{E\{X_1(\omega)X_2(\omega)\}}{E\{X_1(\omega)X_1(\omega)\}},$$

$$H_y(\omega) = \frac{E\{Y_1(\omega)Y_2(\omega)\}}{E\{Y_1(\omega)Y_1(\omega)\}}. \tag{34}$$

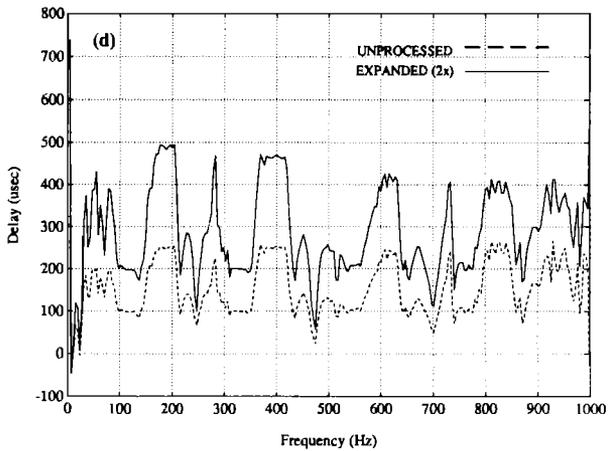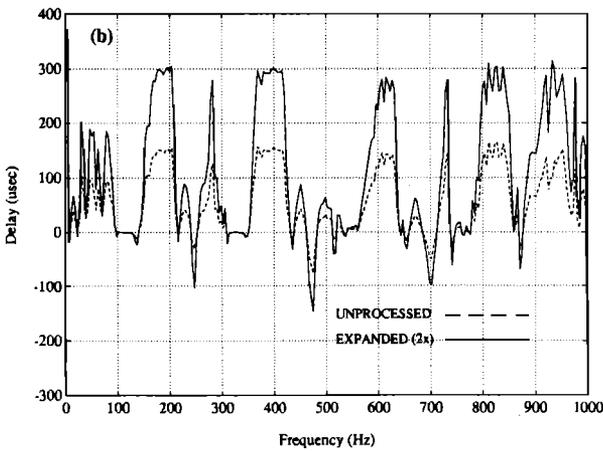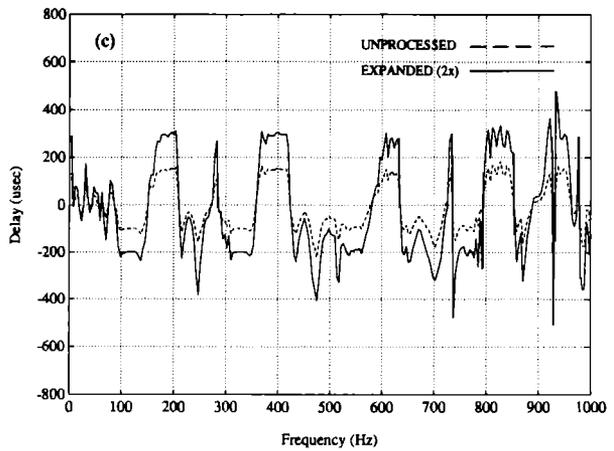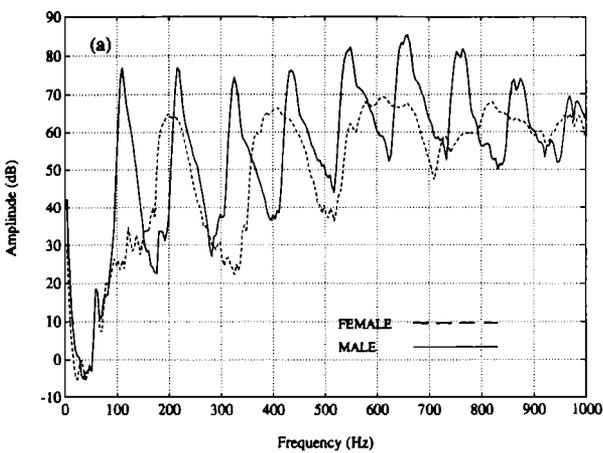Figure 10(a) shows the power spectra of the male and

FIG. 10. (a) Power spectra of /pa/ from the male and female speakers. (b) Transfer function delay (computed from the phase response) between sensors for unprocessed and expanded phase using the FIR filters with response depicted in Fig. 6 to process the speech signals. The binaural delay imposed on the stimuli correspond to 0 μs (center) for the male speaker and 150 μs (right of center) for the female speaker. (c) Transfer function delay with imposed delays of $-$ 100 μs (left of center) and 150 μs (right of center) for the male and female speakers. (d) Transfer function delay with imposed delays of 100 μs (right of center) and 250 μs (right of center) for the male and female speakers.

female utterance of /pa/. The fundamental and its harmonics are clearly defined for both speakers, yet spectrally overlapped to some degree. Figure 10(b)–(d) shows the between channel delay [computed by converting the phase of $H_x(\omega)$ and $H_y(\omega)$ to delay] as a function of frequency for unprocessed and expanded conditions with both speakers as an input. In Fig. 10(b) interaural delays of 0 and 150 μs are imposed on the male and female speech, respectively, corresponding to the mean minimal separation indicated in the experimental results. The regions of nonoverlapping harmonics demonstrate the effectiveness of the expansion algorithm. The interaural time delay is 0 μs at frequencies where the male harmonics are dominant, while the interaural delay expands by a factor of 2, from 150 to 300 μs, at frequencies where the female harmonics are dominant. Figure 10(c) and (d) demonstrates the effectiveness of the expansion algorithm with two off-center sources. Recall that Fig. 6(b) indicates the actual phase response deviates from the ideal at high frequencies for interaural time delays greater than 200 μs ($\tau > 100\,\mu$s). This effect is evident in Fig. 10(d); the expanded delay for the female speaker (interaural delay 250 μs) gradually decreases from 500 μs at low frequencies to 400 μs at high frequencies.

The acoustic analyses in this section indicate that, regardless of the locations of multiple sound sources relative to center, interaural delay expands in a manner consistent with the actual response of the expansion algorithm. They also suggest that an observer could separate the harmonic structure of each speaker with greater probability as a result of the expanded phase or delay.

V. CONCLUSIONS

This research has focused specifically on manipulation of interaural time cues for localization. To this end, we have disregarded interaural magnitude difference cues by manipulating only sound frequencies below 1 kHz. A practical realization of the algorithm obviously requires the integration of a broader range of frequencies. The effects of including interaural magnitude differences and using a larger number of microphones remain as topics for future investigation. Indeed, additional microphones may be required to obtain the additional degrees of freedom necessary for incorporating interaural magnitude variation.

We have presented in this paper the theory and methodology for manipulating auditory space in the horizontal plane using linear filters. Least-squares solutions for filters that are unconstrained and constrained to be FIR are derived and their properties are examined. We have also demonstrated that the algorithm can be instantiated in existing, off the shelf hardware and is capable of running in real-time.

Experimental results indicate that the algorithm effects the desired expansion of auditory space by reducing the minimal separation needed to discriminate between two speakers. Analysis of the speaker data indicates that the algorithm is reliably expanding the interaural delays associated with off-center speakers.

Bendat, J. S., and Piersol, A. G. (1971). *Random Data: Analysis and Measurement Procedures* (Wiley-Interscience, New York).

Bregman, A. S. (1990). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Bregman, A. S., and Campbell, J. (1971). "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Exp. Psych. **89**, 244–249.

Durlach, N. I., and Pang, X. D. (1986). "Interaural magnification," J. Acoust. Soc. Am. **80**, 1849–1850.

Jeffress, L. A. (1948). "A place theory of sound localization," J. Comp. Physiol. Psych. **41**, 25–39.

Jeffress, L. A. (1972). "Binaural signal detection: Vector theory," in *Foundations of Modern Auditory Theory, Vol. II*, edited by J. V. Tobias (Academic, New York).

Kuhn, G. F. (1987). "Physical acoustics and measurement pertaining to directional hearing," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer-Verlag, New York).

Lawson, C. L., and Hanson, R. J. (1974). *Solving Least Squares Problems* (Prentice Hall, Englewood Cliffs, NJ).

Mitchel, O. M., Ross, C. A., and Yates, G. H. (1971). "Signal processing for a cocktail party effect," J. Acoust. Soc. Am. **50**, 656–660.

Stern, R. M., Zeiberg, A. S., and Trahiotis, C. (1988). "Lateralization of complex binaural stimuli: A weighted-image model," J. Acoust. Soc. Am. **84**, 156–165.