Performance Analysis of 3-D Direction Estimation Based on Head-Related Transfer Function

Satyabrata Sen, *Student Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—We analyze the performance of a sensor system based on the head-related transfer function (HRTF) modeling of human hearing, in finding the direction of a source in 3-D space. Thus, this system has only two sensors with a few nearby reflectors. First, we present a parametric measurement model that incorporates the diffractions associated with the head shape and the multipath reflections related to the external ear (pinna). Then, we compute the asymptotic Cramér–Rao bound on the 3-D direction estimate for a stochastic source signal of unknown parametrized spectrum. We provide a few numerical examples to illustrate our analytical results.**

***Index Terms*—3-D direction estimation, head-related transfer function (HRTF), mean-square angular error (MSAE) bounds, multipath reflections.**

I. Introduction

SOURCE direction estimation is a central problem in radar, sonar, navigation, and robotics. Over the past few decades, a wide variety of techniques have been proposed to solve these problems using sensor arrays. Since the accuracy of the estimation improves with increasing number of sensors, many systems often use a large number of sensors. However, biological systems, particularly the human hearing system, perform remarkably well in finding the direction of the sound sources in 3-D with just two ears. Therefore, the development of a sensor system inspired by human hearing could be useful in many applications.

Research on human hearing capabilities is over a century old. Lord Rayleigh was one of the pioneers in this field [1]. He attempted to explain the direction finding capability in terms of two interaural cues: interaural time difference (ITD) and interaural level difference (ILD). ITD arises because of the difference in time of a signal arriving at the two ears, whereas ILD occurs because of the loss of high frequencies caused by head shadowing [2]. However, both ITD and ILD provide information about the lateral angle only, *viz.*, the angle between a ray to the sound source and the vertical median plane. For simplicity, in this paper, we use the terms azimuth and elevation angles for the lateral and polar angles, respectively, of the interaural polar coordinate system.

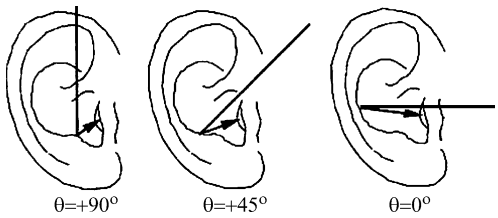

Fig. 1. Pinna reflections of sound for different elevations [21].

About 40 years ago, Batteau emphasized that the external ears (pinnae) play an important role in determining the elevation angle [4]. When the wavelength of the incident sound is on the same order as or shorter than the dimension of the ear, the pinnae start to interact significantly with the incoming wave, as shown in Fig. 1. Over the years, researchers have tried to find experimental evidence that pinna reflections are the major cues for finding elevation angle. Roffler and Butler [5] experimented with a Plexiglas headband to flatten the pinnae against the head. Their results show that listeners wearing the headband could no longer locate broadband noise stimuli accurately. Oldfield and Parker [6] conducted an experiment with wax molds to alter the pinna cues and they noticed that the insertion of the wax molds substantially increases the elevation error. In recent years, Hofman *et al.* found similar results indicating the effect of the wax molds on the pinna acoustic responses in a direction dependent way; see [7, Fig. 1].

Today, it is well established that, in addition to the pinnae, the human torso, shoulders, and head diffract the incoming sound waves. Collectively, these propagation effects are termed the head-related transfer function (HRTF). Note that HRTF can be either measured experimentally or modeled analytically. Brown and Duda have made an extensive experimental study and provide some empirical formulae for the multipath delays produced by the pinna [8]. Although more sophisticated HRTF models have been proposed [9], the Brown–Duda model has the advantage that it provides an analytical relationship between the multipath delays and the azimuth and elevation angles.

In this paper, we analyze a 3-D direction finding system, with only two sensors, inspired by the human auditory system. We consider the Brown–Duda HRTF model as an example to model the frequency dependent head-shadow effects and the multipath delays close to the sensors. However, for mathematical tractability, we do not consider the reflections from the shoulder and torso. We also do not consider any white noise gain error or spatial aliasing error in our model. To analyze the performance of this system for zero-mean wide-sense stationary Gaussian

Manuscript received July 10, 2008; revised September 04, 2008. Current version published March 18, 2009. This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Grant HR0011-07-1-0036. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Malcolm Slaney.

The authors are with the Department of Electrical and Systems Engineering Washington University, St. Louis, MO 63130 USA (e-mail: ssen3@ese.wustl.edu; nehorai@ese.wustl.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2008.2008235

source signals, we compute the asymptotic frequency domain Cramér–Rao bound (CRB) on the error of the 3-D direction estimate. We model the spectrum of the source signal in terms of few unknown parameters. Our numerical examples illustrate the ability to estimate the azimuth and elevation angles.

II. Measurement Model

In this section, we derive our measurement model, integrating the diffraction effects related to the head shape and the multipath propagation effects associated with the pinnae [10].

A. Head-Shadow Filter and ITD

The incident sound waves get diffracted due to the shape of the head before reaching the ears. Brown and Duda [10] provided an approximate but simple formulation of the head-shadow filter in cascade with a delay element corresponding to the ITD. In the frequency domain, this filter has the following form:

$$H_{\mathrm{HS}}(\omega,\phi)=\frac{1+j\frac{\alpha(\phi-\phi_{\mathrm{ear}})\omega}{\frac{2c}{a}}}{1+j\frac{\omega}{\frac{2c}{a}}}\,e^{-j\omega\tau_0(\phi-\phi_{\mathrm{ear}})} \tag{1}$$

where a is the radius of the head, c is the speed of sound, ϕ denotes the azimuth angle, $\alpha(\phi)=1.05+0.95\cos(1.2\phi)$, and ϕ_{ear} specifies the direction of the entrance to the ear canal, which is $+100^\circ$ for the right ear and -100° for the left ear. In (1), $\tau_0(\phi)$ corresponds to Woodworth and Schlosberg's frequency-independent ITD formula [11], which can be written as

$$\tau_0(\phi)=\frac{a}{c}-\frac{a}{c}\cos\phi,\quad \text{for}\quad -90^\circ\le\phi\le 90^\circ. \tag{2}$$

B. Pinna Multipath Filter

The multipath effects due to the pinna can be summarized as follows [10].

1) Apart from the direct path, there are essentially five additional multipaths of the pinna response.
2) The attenuation factors $\{\rho_p\}$ are independent of the azimuth and elevation angles.
3) The time delays $\{\tau_p\}$ are functions of azimuth angle, elevation angle, and the listener.

Considering the received signal to be the coherent sum of the direct signal and multiple attenuated and delayed versions of the direct signal, the multipath filter response can be expressed as

$$H_{\mathrm{MP}}(\omega,\phi,\theta)=1+\sum_{p=1}^{P}\rho_p e^{-j\omega\tau_p(\phi,\theta)} \tag{3}$$

where P is the number of paths from the source to the receiver, ρ_p and τ_p are the attenuation coefficient and time delay through the pth path, respectively. For a source at azimuth angle ϕ and elevation angle θ, the empirical formula for τ_p is

$$\tau_p(\phi,\theta)=A_p\cos\left(\frac{\phi}{2}\right)\sin\left[D_p\left(90^\circ-\theta\right)\right]+B_p \quad \text{for}\ -90^\circ\le\phi\le 90^\circ,-90^\circ\le\theta\le 90^\circ \tag{4}$$

where A_p is the amplitude, B_p is the offset, and D_p is a scaling factor dependent on the individual listener. Their values are given in Table I for the five significant multipaths. For the numerical analysis, in Section IV, we will use $P=5$ according to [10].

TABLE I
Coefficient Values of the Pinna Model [10]

No. of multipath (p)	ρ_p	A_p (in μs)	B_p (in μs)	D_p
1	0.50	22.67	45.35	1.0
2	-1.00	113.38	90.70	0.5
3	0.50	113.38	158.73	0.5
4	-0.25	113.38	249.43	0.5
5	0.25	113.38	294.78	0.5

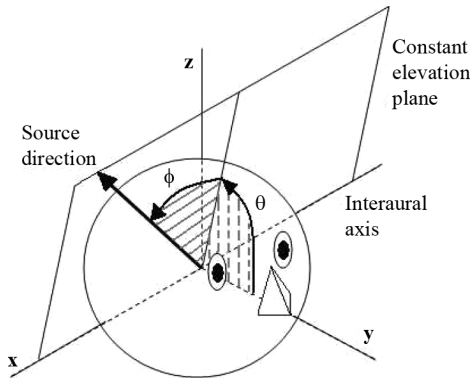

Fig. 2. Interaural polar coordinate system with azimuth angle ϕ and elevation angle θ. The front of the head is in the y direction.

C. Measurement Model

Based on the discussions of the previous subsections, we now describe our measurement model. We use the interaural-polar coordinate system, as shown in Fig. 2, to specify the azimuth and elevation angles. A point directly in front of the head is denoted as being 0° azimuth and 0° elevation. The azimuth coordinate increases in a clockwise direction while the elevation increases upwards.

Consider a far-field source $s(t)$ located at azimuth angle ϕ and elevation angle θ. The received signals at the right and left ears are

$$\begin{aligned} y_{\mathrm{r}}(t)&=h_{\mathrm{r}}(t)\star s(t)+e_{\mathrm{r}}(t)\\ y_{\mathrm{l}}(t)&=h_{\mathrm{l}}(t)\star s(t)+e_{\mathrm{l}}(t)\end{aligned} \tag{5}$$

where $h_{\mathrm{r}}(t)$ and $h_{\mathrm{l}}(t)$ are the impulse responses of the right- and left-ear filters, respectively, and $e_{\mathrm{r}}(t)$ and $e_{\mathrm{l}}(t)$ are the corresponding measurement noises. In the frequency domain, the ear filter responses are denoted as $H_{\mathrm{r}}(\omega,\phi,\theta)$ and $H_{\mathrm{l}}(\omega,\phi,\theta)$, respectively. These can be easily obtained by multiplying the head-shadow filter response $H_{\mathrm{HS}}(\omega,\phi)$ and the multipath filter response $H_{\mathrm{MP}}(\omega,\phi,\theta)$ with appropriate parameters for the right and left ears.

III. Performance Measure

We compute the Cramér–Rao bound (CRB) on the error of the direction estimates (ϕ,θ) for the model presented in Section II. We begin by stating a few statistical assumptions of our model.

A. Statistical Model

We assume that both the signal and noise are zero-mean wide-sense stationary processes, uncorrelated with each other. Therefore, the power spectral density (PSD) matrix $S_{\mathrm{y}}(\omega)$ of the received signal vector $\boldsymbol{y}(t)=[y_{\mathrm{r}}(t),y_{\mathrm{l}}(t)]^T$ can be expressed as

$$S_{\mathrm{y}}(\omega)=\boldsymbol{H}(\omega,\phi,\theta)S_{\mathrm{s}}(\omega,\boldsymbol{a})\boldsymbol{H}^H(\omega,\phi,\theta)+S_{\mathrm{e}}(\omega) \quad (6)$$

where $\boldsymbol{H}(\omega,\phi,\theta)=[H_{\mathrm{r}}(\omega,\phi,\theta),H_{\mathrm{l}}(\omega,\phi,\theta)]^T$, and $S_{\mathrm{s}}(\omega,\boldsymbol{a})$ and $S_{\mathrm{e}}(\omega)$ are the PSDs of the source signal $s(t)$ and the measurement noise process $\boldsymbol{e}(t)=[e_{\mathrm{r}}(t),e_{\mathrm{l}}(t)]^T$, respectively. We parameterize the source spectrum in terms of n unknowns $\boldsymbol{a}=[a_1,a_2,\ldots,a_n]^T$. For simplicity, we further assume that the noise processes $e_{\mathrm{r}}(t)$ and $e_{\mathrm{l}}(t)$ are white and uncorrelated with each other, i.e., $S_{\mathrm{e}}(\omega)=S_{\mathrm{e}}I_2$.

B. Cramér–Rao Bound

Let $\{\boldsymbol{y}(t),t=1,2,\ldots N\}$ be a measurement vector that is a function of an unknown M-element parameter vector $\boldsymbol{\psi}$, and let $\hat{\boldsymbol{\psi}}$ be an unbiased estimate of $\boldsymbol{\psi}$. Under mild regularity assumptions, $\mathrm{CRB}(\hat{\boldsymbol{\psi}})\geq J^{-1}(\boldsymbol{\psi})$, where $J(\boldsymbol{\psi})$ denotes the Fisher information matrix (FIM). When $\boldsymbol{y}(t)$ is a zero-mean wide-sense stationary Gaussian process, a convenient formula for computing the asymptotic form (i.e., large N) of the FIM was given by Whittle [12], as shown by equation (7) at the bottom of the page, where $S_{\mathrm{y}}(\omega,\boldsymbol{\psi})$ denotes the power spectral density of $\boldsymbol{y}(t)$; see [14, Ch. 3] for details.

To apply Whittle's formula to our model, we assume that both the signal and noise are zero-mean wide-sense stationary Gaussian processes and $\tilde{S}_j(\omega)=\partial S_{\mathrm{s}}(\omega)/\partial a_j$ exists for $1\leq j\leq n$. Denoting $\boldsymbol{\psi}=[\theta,\phi,\boldsymbol{a}^T]^T$, we compute CRB on $n+2$ unknown parameters as follows:

$$\mathrm{CRB}(\boldsymbol{\psi})=\begin{bmatrix} J_{\theta\theta} & J_{\theta\phi} & J_{\theta\boldsymbol{a}} \\ J^*_{\theta\phi} & J_{\phi\phi} & J_{\phi\boldsymbol{a}} \\ J^*_{\theta\boldsymbol{a}} & J^*_{\phi\boldsymbol{a}} & J_{\boldsymbol{a}\boldsymbol{a}} \end{bmatrix}^{-1} \quad (8)$$

where you find equations (9)–(12), as shown at the bottom of the next page, where $S_{\mathrm{T}}\triangleq S_{\mathrm{e}}+S_{\mathrm{s}}\boldsymbol{H}^H\boldsymbol{H}$. See the Appendix for detailed derivations.

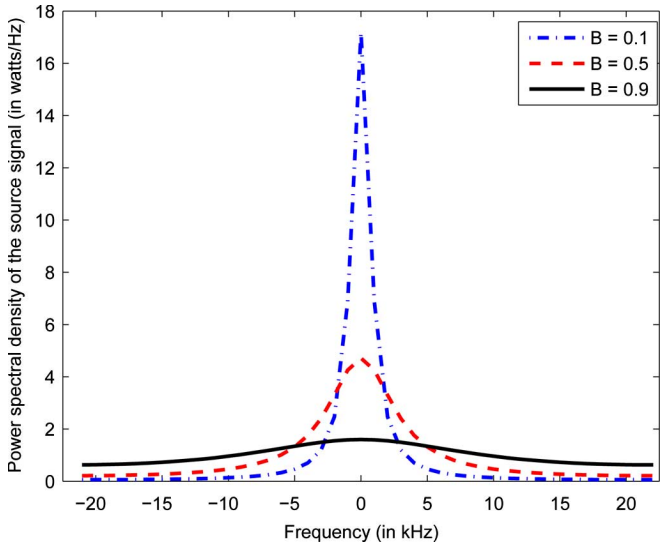

Fig. 3. Power spectral density of the source signal at different normalized bandwidths.

IV. Numerical Results

To illustrate the analytical results of the bounds on the accuracy of azimuth and elevation angle estimations, we evaluated these bounds for a variety of scenarios. For all the results shown in Figs. 4–6, we considered the speed of sound $c=334$ m/s, the distance between two ears $=17$ cm, and sampling frequency $f_{\mathrm{s}}=44.1$ kHz; thus, $N=44\,100$ corresponds to observation length of 1 s. The PSD of the noise process was assumed to be flat over the whole frequency range $[-\pi,\pi]$. We considered the source signal as a first-order autoregressive (AR) process whose PSD is modeled as $S_{\mathrm{s}}(\omega)=\left(1-a_1^2\right)/\left(1-2a_1\cos\omega+a_1^2\right)$. This also ensures that $\int_{-\pi}^{\pi}S_{\mathrm{s}}(\omega)d\omega=1$ and hence $\mathrm{SNR}=\left(\int_{-\pi}^{\pi}S_{\mathrm{s}}(\omega)d\omega\right)/(2\pi S_{\mathrm{e}})=1/(2\pi S_{\mathrm{e}})$. Furthermore, we defined normalized bandwidth $B=\omega_0/\pi$, where ω_0 was computed in such a way that 95% of signal energy lies within $-\omega_0\leq\omega\leq\omega_0$. Fig. 3 depicts the PSD of the

$$J(\boldsymbol{\psi})_{kl}=\frac{N}{4\pi}\int_{-\pi}^{\pi}\mathrm{tr}\left\{\frac{\partial S_{\mathrm{y}}(\omega,\boldsymbol{\psi})}{\partial\psi_k}S_{\mathrm{y}}^{-1}(\omega,\boldsymbol{\psi})\frac{\partial S_{\mathrm{y}}(\omega,\boldsymbol{\psi})}{\partial\psi_l}S_{\mathrm{y}}^{-1}(\omega,\boldsymbol{\psi})\right\}d\omega,\quad 1\leq k,\,l\leq M \quad (7)$$

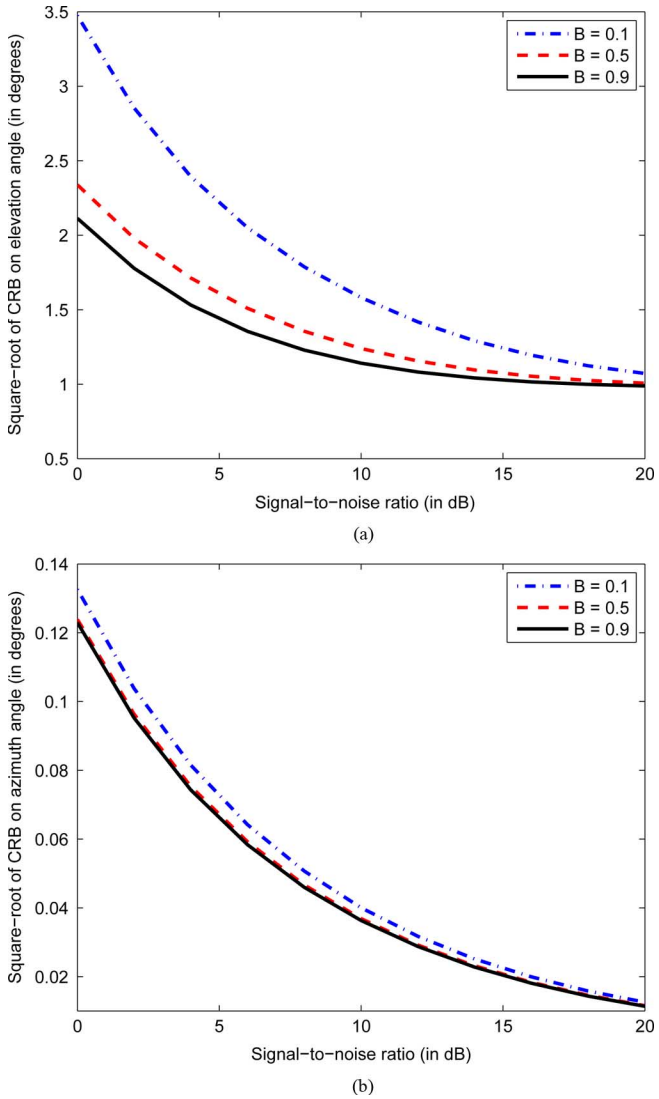

Fig. 4. Square-root of the Cramér–Rao bound on (a) elevation angle and (b) azimuth angle estimation as a function of SNR at three different bandwidths.

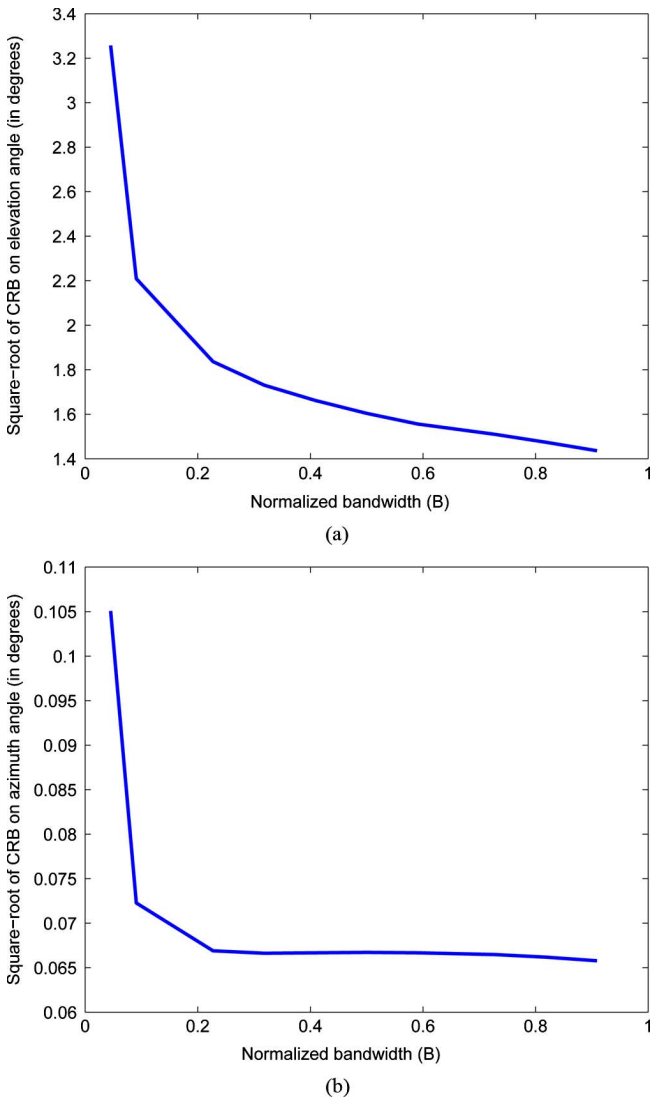

Fig. 5. Square-root of the Cramér–Rao bound on (a) elevation angle and (b) azimuth angle estimation as a function of normalized bandwidth of the source spectrum.

source signal at three different normalized bandwidths. The integrations of (9)–(12) were numerically evaluated using the approximation $(N/2\pi)\int_{-\pi}^{\pi} f(\omega)d\omega \approx \sum_{m=0}^{N-1} f(\omega_m)$, where $\omega_m = 2\pi m/N,\ m = 0,\ 1, \ldots, N-1$, and $f(\omega)$ denotes the corresponding integrand expression.

$$J_{\theta\phi} = \frac{N}{2\pi}\int_{-\pi}^{\pi}\left[\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)^2 \mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\frac{\partial \boldsymbol{H}^H}{\partial\phi}\boldsymbol{H}\right\} + \left(\frac{S_{\mathrm{s}}^2}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\phi}\right\}\right]d\omega \tag{9}$$

$$J_{\theta a_j} = \frac{N}{2\pi}\int_{-\pi}^{\pi}\tilde{S}_j\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}^2}\right)\boldsymbol{H}^H\boldsymbol{H}\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\right\}d\omega \quad \text{for } 1 \le j \le n \tag{10}$$

$$J_{\phi a_j} = \frac{N}{2\pi}\int_{-\pi}^{\pi}\tilde{S}_j\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}^2}\right)\boldsymbol{H}^H\boldsymbol{H}\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\phi}\boldsymbol{H}\right\}d\omega \quad \text{for } 1 \le j \le n \tag{11}$$

$$J_{a_i a_j} = \frac{N}{4\pi}\int_{-\pi}^{\pi}\tilde{S}_i\tilde{S}_j\left(\frac{\boldsymbol{H}^H\boldsymbol{H}}{S_{\mathrm{T}}}\right)^2 d\omega \quad \text{for } 1 \le i,\ j \le n \tag{12}$$

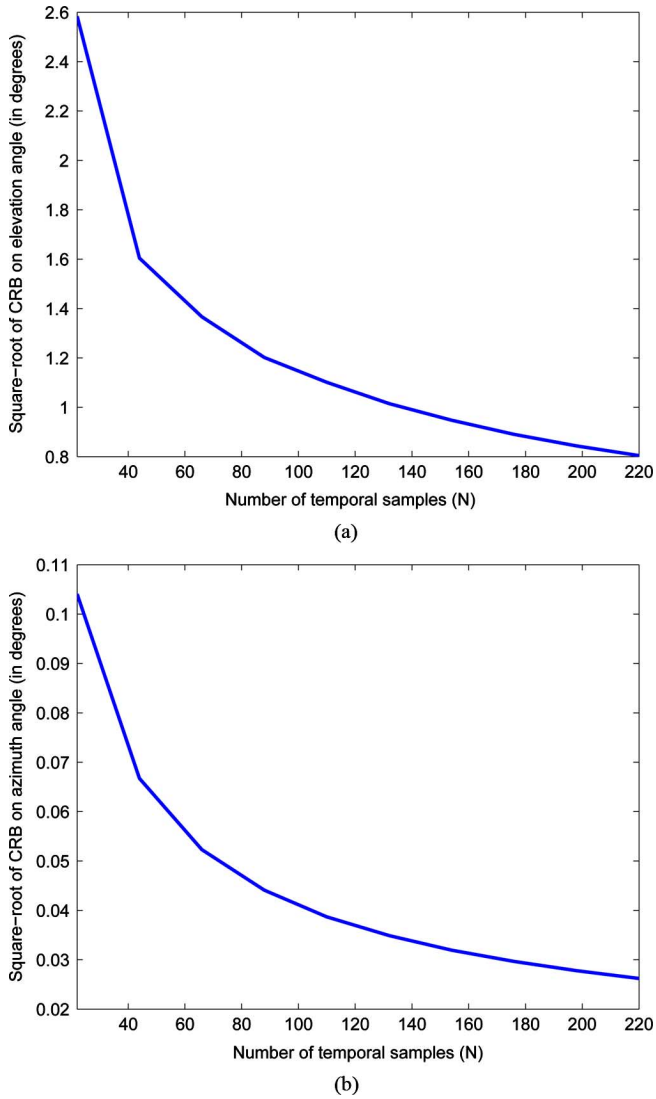

Fig. 6. Square-root of the Cramér–Rao bound on (a) elevation angle and (b) azimuth angle estimation as a function of the number of temporal samples. Here, the X-axis also corresponds to 0.5–5 ms.

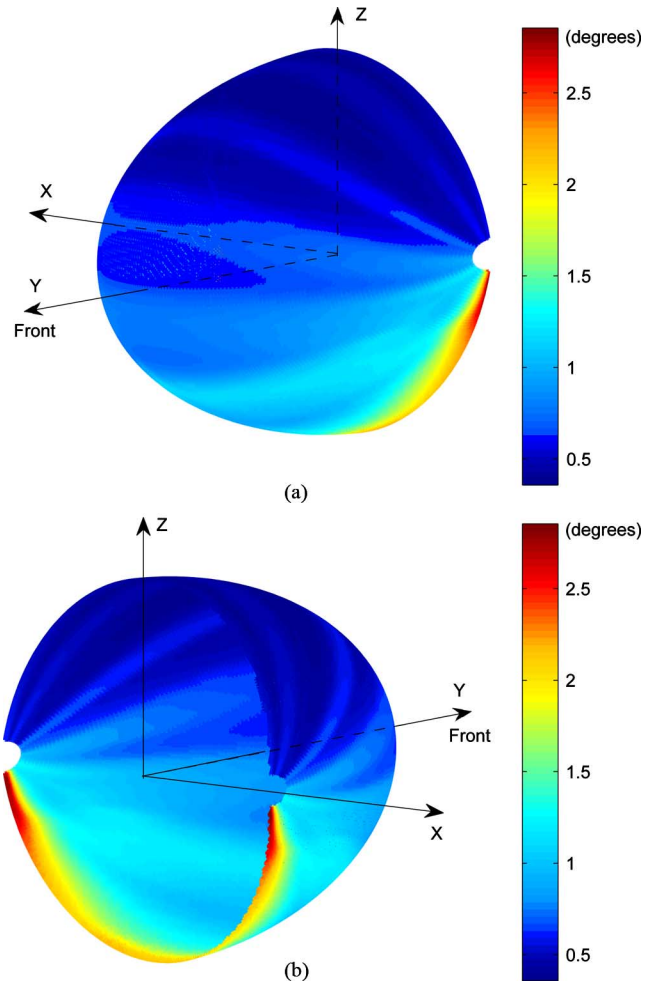

Fig. 7. Square-root of the mean-square angular error lower bound on source direction estimation at different azimuth and elevation angles in the frontal hemisphere of the head. (a) Front-side view. (b) Back-side view. The axes are marked in accordance to the coordinate system defined in Fig. 2.

Fig. 4 depicts the square-root of CRB on the elevation and azimuth angle estimation as a function of SNR, for different values of B. For this example, we considered $\phi_0 = 30°$, $\theta_0 = 30°$, and $N = 44$. Fig. 5 shows the dependence of CRB on elevation and azimuth angle estimation on B. In this case, we fixed $\phi_0 = 30°$, $\theta_0 = 30°$, SNR $= 5$ dB, and $N = 44$. The variations of the bounds with respect to N are presented in Fig. 6 while keeping $\phi_0 = 30°$, $\theta_0 = 30°$, $B = 0.5$, and SNR $= 5$ dB. In Fig. 7, we plot the square-root of the mean-square angular error lower bound $(\mathrm{MSAE}_{\mathrm{CR}})$ at different values of azimuth and elevation angles shown as points on a unit sphere. The color on the spherical surface denotes the values of the bound. The $\mathrm{MSAE}_{\mathrm{CR}}$ is a useful performance measure of 3-D direction estimation [15]. The angular error is defined as the angular difference between the directional vector to the source $\boldsymbol{u} = [\sin\phi, \cos\phi\cos\theta, \cos\phi\sin\theta]^T$ and its estimate. Following [15], for our model we evaluated that $\mathrm{MSAE}_{\mathrm{CR}}(\phi,\theta) = \left[\mathrm{CRB}(\phi) + \cos^2\phi\,\mathrm{CRB}(\theta)\right]$, where $\mathrm{CRB}(\phi)$ and $\mathrm{CRB}(\theta)$ are the (scalar) CRBs on the azimuth and elevation angles, respectively. Note that for consistency with our other figures, we did not premultiply the $\mathrm{MSAE}_{\mathrm{CR}}$ with N as in [15], and therefore, here $\mathrm{MSAE}_{\mathrm{CR}}$ is still proportional to $1/N$. For this example, we fixed SNR $= 5$ dB, $B = 0.9$, and $N = 44$. From these results, we summarize the following.

- From Fig. 4(a) and (b) it is evident that the azimuth estimation is more accurate than the elevation. Note that this is similar to the experimental results on human hearing [16], [17].
- Fig. 5 suggests that the bounds are relatively small for the wideband source spectrum. This means that the frequency diversity improves the location estimation. This is also consistent with the experimental results on human hearing [18], [19].
- Fig. 6 shows the number of temporal samples required to achieve a certain accuracy in elevation and azimuth angle estimations. For example, to attain $\sqrt{\mathrm{CRB}_{\theta\theta}} = 1°$ we need approximately 136 samples, which corresponds to about 3 ms for the chosen f_s.
- From Fig. 7, we have the following observations: 1) the bound is symmetric about the vertical median plane $(\phi = 0°)$ but not about the visuoaural plane $(\theta = 0°)$, 2) the

accuracy of direction estimation is bad towards the side of the head and worst at negative elevations, and 3) the accuracy is slightly reduced at certain mid-elevation angles indicated by the patches of dark and light blue color. However, for the sources that are more than 45° below the horizontal plane the observation (2) may be different from the experimental results as the Brown–Duda model does not include the effect of torso.

V. Conclusion

We have presented a simple model for direction estimation in 3-D space using only two sensors and a small number of nearby reflectors, based on the human auditory system, namely HRTF-based modeling. To analyze the performance, we compute the Cramér–Rao bound on the direction estimate for wideband stochastic Gaussian source signals of unknown parametrized spectrum. We are extending our analysis to propose a 3-D direction-finding system that exploits multipath reflections close to the sensors inspired by human hearing [20]. This fits practical systems where there are close-range reflections from the parts of the platform on which the sensor is mounted. In our future work, we will analyze our model under more general scenarios incorporating nonstationary source signals. We will also analyze the performance for sources in the backside of the head. We will explore the identifiability issue for the unknown source spectrum. We also plan to conduct experiments with real data to verify our analytical results.

Appendix

In this section, we demonstrate the steps for computing the expressions of $J_{\theta\phi}$, $J_{\theta a_j}$, and $J_{a_i a_j}$. The other entries of the

$$
\begin{aligned}
\frac{\partial S_{\mathrm{y}}}{\partial\theta}S_{\mathrm{y}}^{-1}\,\frac{\partial S_{\mathrm{y}}}{\partial\phi}S_{\mathrm{y}}^{-1} &= \frac{\partial \boldsymbol{H}}{\partial\theta}\left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}}+\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H\,\frac{\partial \boldsymbol{H}}{\partial\phi}\left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}}+\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H \\
&\quad+\frac{\partial \boldsymbol{H}}{\partial\theta}\left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}}+\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H\boldsymbol{H}S_{\mathrm{s}}\frac{\partial \boldsymbol{H}^H}{\partial\phi}S_{\mathrm{y}}^{-1} \\
&\quad+\boldsymbol{H}S_{\mathrm{s}}\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\phi}\left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}}+\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H \\
&\quad+\boldsymbol{H}S_{\mathrm{s}}\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\boldsymbol{H}S_{\mathrm{s}}\frac{\partial \boldsymbol{H}^H}{\partial\phi}S_{\mathrm{y}}^{-1} \\
\frac{\partial S_{\mathrm{y}}}{\partial\theta}S_{\mathrm{y}}^{-1}\,\frac{\partial S_{\mathrm{y}}}{\partial a_j}S_{\mathrm{y}}^{-1} &= \tilde{S}_j\frac{\partial \boldsymbol{H}}{\partial\theta}\left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}}+\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H\boldsymbol{H}\boldsymbol{H}^HS_{\mathrm{y}}^{-1} \\
&\quad+\tilde{S}_j\boldsymbol{H}S_{\mathrm{s}}\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\boldsymbol{H}\boldsymbol{H}^HS_{\mathrm{y}}^{-1} \\
\frac{\partial S_{\mathrm{y}}}{\partial a_i}S_{\mathrm{y}}^{-1}\,\frac{\partial S_{\mathrm{y}}}{\partial a_j}S_{\mathrm{y}}^{-1} &= \tilde{S}_i\tilde{S}_j\boldsymbol{H}\boldsymbol{H}^HS_{\mathrm{y}}^{-1}\boldsymbol{H}\boldsymbol{H}^HS_{\mathrm{y}}^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{tr}\left\{\frac{\partial S_{\mathrm{y}}}{\partial\theta}\,S_{\mathrm{y}}^{-1}\frac{\partial S_{\mathrm{y}}}{\partial\theta}\,S_{\mathrm{y}}^{-1}\right\} &= \left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)^2\boldsymbol{H}^H\frac{\partial \boldsymbol{H}}{\partial\theta}\boldsymbol{H}^H\frac{\partial \boldsymbol{H}}{\partial\theta}+\left(\frac{S_{\mathrm{s}}^2}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\theta} \\
&\quad+\left(\frac{S_{\mathrm{s}}^2}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\theta}+\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)^2\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H} \\
&= 2\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)^2\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\right\}+2\left(\frac{S_{\mathrm{s}}^2}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\theta}\right\} \\
\mathrm{tr}\left\{\frac{\partial S_{\mathrm{y}}}{\partial\theta}\,S_{\mathrm{y}}^{-1}\frac{\partial S_{\mathrm{y}}}{\partial a_j}\,S_{\mathrm{y}}^{-1}\right\} &= \tilde{S}_j\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\left\{\boldsymbol{H}^HS_{\mathrm{y}}^{-1}\frac{\partial \boldsymbol{H}}{\partial\theta}\right\}+\tilde{S}_j\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}}\right)\boldsymbol{H}^H\boldsymbol{H}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}S_{\mathrm{y}}^{-1}\boldsymbol{H}\right\} \\
&= 2\tilde{S}_j\left(\frac{S_{\mathrm{s}}}{S_{\mathrm{T}}^2}\right)\boldsymbol{H}^H\boldsymbol{H}\,\mathrm{Re}\left\{\frac{\partial \boldsymbol{H}^H}{\partial\theta}\boldsymbol{H}\right\} \\
\mathrm{tr}\left\{\frac{\partial S_{\mathrm{y}}}{\partial a_i}\,S_{\mathrm{y}}^{-1}\frac{\partial S_{\mathrm{y}}}{\partial a_j}\,S_{\mathrm{y}}^{-1}\right\} &= \tilde{S}_i\tilde{S}_j\boldsymbol{H}^HS_{\mathrm{y}}^{-1}\boldsymbol{H}\boldsymbol{H}^HS_{\mathrm{y}}^{-1}\boldsymbol{H}=\tilde{S}_i\tilde{S}_j\left(\frac{\boldsymbol{H}^H\boldsymbol{H}}{S_{\mathrm{T}}}\right)^2
\end{aligned}
$$

FIM can be obtained following a similar procedure. We know that

$$\frac{\partial S_{\mathrm{y}}}{\partial \theta} = \frac{\partial \boldsymbol{H}}{\partial \theta} S_{\mathrm{s}} \boldsymbol{H}^H + \boldsymbol{H} S_{\mathrm{s}} \frac{\partial \boldsymbol{H}^H}{\partial \theta},$$
$$\frac{\partial S_{\mathrm{y}}}{\partial \phi} = \frac{\partial \boldsymbol{H}}{\partial \phi} S_{\mathrm{s}} \boldsymbol{H}^H + \boldsymbol{H} S_{\mathrm{s}} \frac{\partial \boldsymbol{H}^H}{\partial \phi} \tag{A-1}$$
$$\frac{\partial S_{\mathrm{y}}}{\partial a_j} = \frac{\partial S_{\mathrm{s}}}{\partial a_j} \boldsymbol{H}\boldsymbol{H}^H = \tilde{S}_j \boldsymbol{H}\boldsymbol{H}^H \quad \text{for} \quad 1 \leq j \leq n. \tag{A-2}$$

Further, using Woodbury's matrix inversion identity in (6), we have

$$S_{\mathrm{y}}^{-1} = \frac{1}{S_{\mathrm{e}}} \left[I_2 - \boldsymbol{H} \left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}} + \boldsymbol{H}^H \boldsymbol{H} \right)^{-1} \boldsymbol{H}^H \right]. \tag{A-3}$$

Using (A-1)–(A-3), after some algebraic manipulation we get

$$\frac{\partial S_{\mathrm{y}}}{\partial \theta} S_{\mathrm{y}}^{-1} = \frac{\partial \boldsymbol{H}}{\partial \theta} \left(\frac{S_{\mathrm{e}}}{S_{\mathrm{s}}} + \boldsymbol{H}^H \boldsymbol{H} \right)^{-1} \boldsymbol{H}^H + \boldsymbol{H} S_{\mathrm{s}} \frac{\partial \boldsymbol{H}^H}{\partial \theta} S_{\mathrm{y}}^{-1} \tag{A-4}$$
$$\frac{\partial S_{\mathrm{y}}}{\partial a_j} S_{\mathrm{y}}^{-1} = \tilde{S}_j \boldsymbol{H}\boldsymbol{H}^H S_{\mathrm{y}}^{-1}. \tag{A-5}$$

Therefore, we get the equation shown at the bottom of the previous page. Defining $S_{\mathrm{T}} \triangleq S_{\mathrm{e}} + S_{\mathrm{s}} \boldsymbol{H}^H \boldsymbol{H}$, we get $\left(S_{\mathrm{e}}/S_{\mathrm{s}} + \boldsymbol{H}^H \boldsymbol{H} \right)^{-1} = S_{\mathrm{s}}/S_{\mathrm{T}}$ and $S_{\mathrm{y}}^{-1} \boldsymbol{H} = \boldsymbol{H}/S_{\mathrm{T}}$. Then, using the properties of trace operator, we get the equation as shown at the bottom of the previous page. Substituting these into (7), we obtain the expressions given in (9)–(12).

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the insightful comments that improved the presentation.

REFERENCES

[1] L. Rayleigh, "On our perception of sound direction," *Philos. Mag.*, vol. 13, pp. 214–232, 1907.
[2] W. M. Hartmann, "How we localize sound," *Phys. Today*, vol. 52, pp. 24–29, Nov. 1999.
[3] J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," *Annu. Rev. Psychol.*, vol. 42, pp. 135–159, 1991.
[4] D. W. Batteau, "The role of the pinna in human localization," in *Proc. R. Soc. London, Series B, Biol. Sci.*, Aug. 15, 1967, vol. 168, pp. 158–180.
[5] S. K. Roffler and R. A. Butler, "Factors that influence the localization of sound in the vertical plane," *J. Acoust. Soc. Amer.*, vol. 43, no. 6, pp. 1255–1259, Jun. 1968.
[6] S. R. Oldfield and S. P. A. Parker, "Acuity of sound localisation: A topography of auditory space. II. Pinna cues absent," *Perception*, vol. 13, no. 5, pp. 601–617, 1984.
[7] P. M. Hofman, J. G. A. V. Riswick, and A. J. V. Opstal, "Relearing sound localization with new ears," *Nature Neurosci.*, vol. 1, no. 5, pp. 417–421, Sep. 1998.
[8] C. P. Brown, "Modeling the Elevation Characteristics of the Head-Related Impulse Response," M.S. thesis, San Jose State Univ., San Jose, CA, 1996.
[9] P. Satarzadeh, V. R. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Conv. Audio Eng. Soc.*, Vienna, Austria, May 5–8, 2007, paper no. 7098.
[10] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 476–488, Sep. 1998.
[11] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1983.
[12] P. Whittle, "The analysis of multiple stationary time series," *J. R. Statist. Soc. , Ser. B (Methodol.)*, vol. 15, no. 1, pp. 125–139, 1953.
[13] A. Zeira and A. Nehorai, "Frequency domain Cramér-Rao bound for Gaussian processes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 6, pp. 1063–1066, Jun. 1990.
[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
[15] A. Nehorai and E. Paldi, "Vector-sensor array processing for electromagnetic source localization," *IEEE Trans. Signal Process*, vol. 42, no. 2, pp. 376–398, Feb. 1994.
[16] S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," *Hear. Res.*, vol. 114, pp. 179–196, 1997.
[17] S. Carlile, *Virtual Auditory Space: Generation and Applications*. Austin, TX: R. G. Landes, 1996.
[18] S. K. Roffler and R. A. Butler, "Localization of tonal stimuli in the vertical plane," *J. Acoust. Soc. Amer.*, vol. 43, no. 6, pp. 1260–1266, Jun. 1968.
[19] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Amer.*, vol. 53, no. 2, pp. 400–408, Feb. 1973.
[20] S. Sen and A. Nehorai, "Exploiting close-to-the-sensor multipath reflections using a human–hearing-inspired model," *IEEE Trans. Signal Process*, vol. 57, no. 2, pp. 803–808, Feb. 2009.
[21] C. P. Brown, "Modeling the elevation characteristics of the head-related impulse response," M.S. thesis, San Jose State Univ., San Jose, CA, May 1996.

Satyabrata Sen (S'07) received the B. E. degree in electronics and telecommunication engineering from Jadavpur University, India, in 2002, and the M.Tech. degree in electrical engineering (with specialization in communication and signal processing) from the Indian Institute of Technology, Bombay, India, in 2005. He is currently pursuing the Ph.D. degree in the Department of Electrical and Systems Engineering, Washington University, St. Louis, MO.

His research interests are in the area of statistical signal processing, detection and estimation theory, and their applications in radar, communications, and sensor arrays.

Arye Nehorai (S'80-M'83-SM'90-F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion, Haifa, Israel, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

From 1985 to 1995, he was a faculty member with the Department of Electrical Engineering, Yale University, New Haven, CT. In 1995, he joined as a Full Professor the Department of Electrical Engineering and Computer Science, The University of Illinois, Chicago (UIC). From 2000 to 2001, he was Chair of the department's Electrical and Computer Engineering (ECE) Division, which then became a new department. In 2001, he was named University Scholar of the University of Illinois. In 2006, he became Chairman of the Department of Electrical and Systems Engineering, Washington University, St. Louis, MO. He has been the inaugural holder of the Eugene and Martha Lohman Professorship and the Director of the Center for Sensor Signal and Information Processing (CSSIP) at WUSTL since 2006.

Dr. Nehorai was Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING during the years 2000 to 2002. In the years 2003 to 2005, he was Vice President (Publications) of the IEEE Signal Processing Society (SPS), Chair of the Publications Board, member of the Board of Governors, and member of the Executive Committee of this Society. From 2003 to 2006, he was the founding editor of the special columns on Leadership Reflections in the *IEEE Signal Processing Magazine*. He was corecipient of the IEEE SPS 1989 Senior Award for Best Paper with P. Stoica, c-author of the 2003 Young Author Best Paper Award and corecipient of the 2004 Magazine Paper Award with A. Dogandzic. He was elected Distinguished Lecturer of the IEEE SPS for the term 2004 to 2005 and received the 2006 IEEE SPS Technical Achievement Award. He is the Principal Investigator of the new multidisciplinary university research initiative (MURI) project entitled Adaptive Waveform Diversity for Full Spectral Dominance. He has been a Fellow of the Royal Statistical Society since 1996.