INFORMATION-THEORETIC PERFORMANCE ANALYSIS OF AZIMUTHAL LOCALIZATION FOR SPATIAL-AUDITORY DISPLAY OF BEAMFORMED SONAR DATA

Jason E. Summers

Applied Research in Acoustics (ARiA) LLC
Washington, D.C.
jason.e.summers@ariacoustics.com

ABSTRACT

An information-theoretic model of azimuthal localization is presented. The number of distinct source locations that can be encoded by a set of head-related impulse response functions (HRIR) is predicted in terms of information transfer as a function of the properties of the source signal and a quantization interval that is related to the level of internal perceptual noise. The model also predicts how source locations should be distributed in azimuth in order to maximize the information transferred through the set of HRIR for a given set of input conditions. The predictions are related to design considerations for a spatial-auditory display of beamformed sonar data in which time series associated with fixed beams of a one-dimensional array are mapped to virtual sources located at fixed radius from the listener in the horizontal plane.

1. INTRODUCTION

A recent review by Arrabito et al. cites, "the ability to present sonar beams in a three-dimensional auditory display where the spatial position of each sonar beam corresponds to the actual position of the source in the ocean," as a key research area for enhancing the role of the auditory modality in processing of sonar data [1]. While such an auditory equivalent of a low-level geographic situation (GEOSIT) display is ultimately limited by the beamforming algorithm and the physical receiver array, the number and spacing of the virtual-source locations to which beams are mapped should be governed by the spatial resolution of human hearing. Here, a model is developed that predicts and bounds human azimuthal localization performance based on the amount of spatial information encoded in a set of head-related impulse responses (HRIR). A well-defined signal-processing algorithm is described which determines the information content of the HRIR as a function of several variables, including internal perceptual noise and external source-signal spectrum.

Though direct methods for binaural presentation of sonar signals have been used since the earliest days of sonar [2, 3], spatial auditory display of sonar beam data in which beams are mapped to virtual sources was first mentioned in the open literature much more recently by McFadden and Taylor [4] and has been a topic of ongoing interest [5, 6, 7, 1] .

A simple and obvious approach for presentation of sonar beams from a one-dimensional array via a spatial-auditory display is mapping of the time-series associated with each beam to a virtual source located at a particular azimuth on a circle in the horizontal plane. Such a scheme raises two primary technical concerns. First, it must be known how many independent (virtual) source locations can be identified, which determines the number of beams that can be mapped. Second, it must be known how the virtual sources associated with the beams should be distributed in azimuth in order to realize the desired level of performance. Two theoretical questions underlie these technical concerns: How much azimuthal information about source location can a set of HRIR encode and how much of this information can a listener extract? Further, how is the information density of a set of HRIR allocated in azimuth?

To address these questions, spatial hearing in the horizontal plane is here recast as a communication problem in which scattering from the head and torso, described by the set of HRIR, encodes source location. This representation of the problem provides an information-theoretic framework for the analysis of localization performance. By postulating coding and decoding in terms of the coefficients of a particular orthogonal decomposition of the set of HRIR, the amount of information transferred and, consequently, localization performance is mathematically determined as a function of the particular set of HRIR, the source-signal, and, through a resulting quantization interval, the perceptual signal-to-noise ratio (SNR). For each set of input parameters the model yields a map of information density as a function of azimuth, which indicates the locations of virtual sources required to realize the maximum information transfer possible for a given set of conditions.

2. INFORMATION-THEORETIC MODEL

The task of identifying the direction of a sound source from the received binaural signal is similar to the problem of localizing a radiator in a sound channel from measurements of the sound field in the channel (i.e., matched-field processing, see, e.g., [8]). In both cases spatially dependent variations in the impulse response (of the channel or the scattering from the head and torso) encode information about source position and one estimates the location of the source from measurements of the sound field at the receiver. By recasting this problem as a gridded search in which the task is identifying which cell of the grid contains the source, Buck et al. [9, 10, 11] formulated source localization as an unconventional communication problem and developed an information-theoretic framework for characterizing localization performance. Though Buck et al. investigated the standard matched-field problem of a single-frequency continuous source with the measurement of the sound field being the vector of complex pressures received on a vertical array of hydrophones, Gaumond later used a con-

This work was supported in part by the Office of Naval Research and was performed in part while the author held a National Research Council Research Associateship Award at The U.S. Naval Research Laboratory.

structive approach within this information-theoretic framework to characterize localization performance for a band-limited impulsive source with the measurement of the sound field being the pressure time series received on a single hydrophone [12]. It is this latter approach that is followed here.

In the information-theoretic representation, the identity of the cell containing the source is the message that is encoded into the sound field and one estimates the identity of the cell containing the source based on noisy measurements of the sound field at the receiver. Note that this formulation, though isomorphic to more conventional communication problems, is fundamentally unlike them in interpretation. Whereas the signal transmitted by the source contains the encoded message in conventional communication problems, in this problem the source location itself is the message.

To characterize azimuthal localization performance, assume that there is a source located on a circle in the horizontal plane with unknown azimuth Θ described by the probability density function $p_\Theta(\theta)$. Following Buck, the continuous set of input conditions is discretized according to $\beta(\theta)$ described by the probability mass function $p_\beta(m)$ where $m = 1, 2, \ldots, M-1, M$.

The time series of acoustic pressure at each ear resulting from a source located at $\Theta = \theta$ are given by

$$x_{left}(\theta; t) = s(t) * h_{left}(\theta; t), \tag{1a}$$

$$x_{right}(\theta; t) = s(t) * h_{right}(\theta; t), \tag{1b}$$

where $s(t)$ is the source waveform and $h_{left}(\theta; t)$ and $h_{right}(\theta; t)$ are the HRIR associated with azimuth θ and the left and right ears, respectively. These can be represented as a single time series by concatenating the responses from the left and right ears in a single time series

$$x(\theta; t) = [\ x_{left}(\theta, t)\ \ x_{right}(\theta, t)\ ]. \tag{2}$$

This received signal is corrupted by internal processes that produce perceptual and criterial noise [13, p. 458], $n(t)$, which can be modeled as additive white Gaussian noise (AWGN). The resulting time series is given by

$$y(\theta; t) = x(\theta; t) + n(t). \tag{3}$$

While the prior equations are expressed in terms of the continuous azimuthal variable Θ, they are equivalent for the discretized case when expressed in terms of the discrete azimuthal index β.

In the discrete case, the complete set of HRIR can be represented as an M-by-N matrix $\mathbf{X}$, where each row of $\mathbf{X}$ is the discrete-time signal of length N given by (2). As in [12] the matrix is constructed so as to exclude initial time-delay because it does not encode any information about source location. However, it is constructed to preserve all aspects of $x(\beta; t)$ that do encode information about source location—interaural time difference (ITD), interaural level difference (ILD), spectral variation and other cues such as those described in [14]. This is accomplished by sequentially time shifting each $x(m; t)$ in order to maximize the cross correlation between the high-energy peaks at the ipsilateral ear with those of the adjacent $x(m-1; t)$. Because each $x(\beta; t)$ is shifted as a whole, ITD is preserved.

In this signal-processing approach, the goal is estimation of the source position as a function of the corrupted signal given by (3)

$$\hat{\beta} = g(y(\beta, t)). \tag{4}$$

In contrast to localization-performance metrics that measure the mean-square error between the estimated and true source position, such as the Cramer-Rao lower bound (CRLB) [8, 15], the information-theoretic framework characterizes performance by the probability of error P_e in assigning the source to a discrete cell

$$P_e = Pr\{\hat{\beta} \neq \beta\}. \tag{5}$$

To characterize the information content of the set of HRIR in terms of a set of discrete states, the time series associated with each discrete source location given by β is expanded in a set of empirical orthogonal functions (EOF) $\nu_n(t)$

$$x(\beta = m; t) = x_m(t) = \sum_{n=1}^{N} \alpha_{mn} \nu_n(t), \tag{6}$$

such that the information content of the set of HRIR is described by the set of coefficients $\{\alpha_{mn}\}$. In this representation each $x_m(t)$ describes the mapping from one of the M discrete positions in azimuth described by β to a position in an N-dimensional space given by the vector of coefficients

$$\mathbf{a}_m = [\alpha_{m1} \cdots \alpha_{mN}]. \tag{7}$$

This expansion is realized by singular value decomposition of the matrix $\mathbf{X}$

$$\mathbf{X} = \mathbf{USV}^T, \tag{8}$$

where the coefficients α_{mn} are given by the product of the matrix of singular vectors and the diagonal matrix of singular values

$$\mathbf{US} = \begin{bmatrix} \alpha_{11} & & \\ & \ddots & \\ & & \alpha_{M,N} \end{bmatrix}, \tag{9}$$

and the EOF $\nu_n(t)$ are given by the matrix of singular vectors

$$\mathbf{V} = \begin{bmatrix} \nu_1(t) \\ \vdots \\ \nu_N(t) \end{bmatrix}. \tag{10}$$

As Gaumond observed [12] for underwater sound channels, coefficient vectors $\mathbf{a}_m$ are not necessarily unique; some may be degenerate. The same holds true for the coefficient vectors of the HRIR. For example, a spherical model of the head yields a set of HRIR that do not resolve source positions lying on cones of confusion. However, HRIR from a realistic head-and-torso model are able to encode substantially more information about source position (see, e.g., [16] and [17, p. 274]) such that the primary source of degeneracy is limited resolution of the encoding due to quantization. For human localization this quantization of coefficients serves as model for the effects of internal noise processes. Expanding $n(t)$ in the same set of EOF

$$n(t) = \sum_{n=1}^{N} \eta_n \nu_n(t), \tag{11}$$

yields a set of independent identically distributed (iid) coefficients η_n that are zero mean with variance σ^2. This corresponds to the uncertainty being uniform in each dimension of the N-dimensional space to which the HRIR maps source position β.

Following Gaumond [12], it is assumed that σ^2 defines the quantization interval ρ_0 through a multiplicative constant $\rho_0 = a\sigma^2$. The vector of quantized coefficients is thus given by

$$\mathbf{c}_m = \left[\frac{\alpha_{m1}}{\rho_0} \cdots \frac{\alpha_{mN}}{\rho_0}\right]^T. \quad (12)$$

As in the case of $\mathbf{a}_m$, not all $\mathbf{c}_m$ are unique.

The information input into the channel is given by the entropy of the probability mass function of the random variable describing source position

$$H_{in} = H(\beta) = -\sum_{m=1}^{M} p_\beta(m) log_2 p_\beta(m). \quad (13)$$

Assuming that all source positions are equally probable (maximum entropy) yields

$$p_\beta(m) = 1/M, \quad (14)$$

so that

$$H(\beta) = log_2 M. \quad (15)$$

This corresponds to $2^{H(\beta)}$ discernible source positions. The joint source-channel coding theorem requires that the mutual information $I(\beta;\hat{\beta})$ satisfy

$$H(\beta) \leq I(\beta;\hat{\beta}), \quad (16)$$

in order to estimate β with arbitrary small probability of error ($P_e \to 0$). Mutual information can be expressed as

$$I(\beta;\hat{\beta}) = H(\hat{\beta}) - H(\hat{\beta}|\beta), \quad (17)$$

where $H(\hat{\beta})$ is a measure of prior uncertainty about which azimuthal cell contains the source and the conditional probability $H(\hat{\beta}|\beta)$ represents that uncertainty remaining about $\hat{\beta}$ once the source transmits from cell $\beta = m$. When all $\mathbf{c}_m$ are unique, $H(\hat{\beta}|\beta) = 0$ and $H(\mathbf{c}) = H(\beta)$. Thus, it is required that $H(\hat{\beta}) = H(\beta)$ for ($P_e \to 0$). As in [12] the constructive approach postulates that the decoding from HRIR to $\hat{\beta}$ is done in terms of the coefficients of a particular orthogonal decomposition. This is only one possible decoding scheme, motivated more by mathematical convenience than an underlying verisimilitude to the actual human processes. Other decoding schemes such as those described in [17] that are more closely related to psychophysical localization cues are also possible. However, the decoding scheme presented here has the significant analytical advantage of producing a set of orthogonal coefficients. Therefore it is postulated that $\hat{\beta} = g(\mathbf{c})$. The data-processing theorem [18, Thm. 2.8.1, pp. 34–35], then requires that

$$I(\beta;\hat{\beta}) \leq I(\beta;\mathbf{c}), \quad (18)$$

such that for $H(\mathbf{c}) = H(\beta)$ must hold true for ($P_e \to 0$). Because $\hat{\beta} = g(\mathbf{c})$, the output information is given by the entropy of those L vectors of discrete coefficients that are unique

$$H_{out} = H(\ell) = -\sum_{\ell=1}^{L} p(\ell) log_2 p(\ell) \quad \in [0, H(\beta)], \quad (19)$$

where $p(\ell)$, $\ell = 1, 2, \ldots, L-1, L$, is the sum of the probabilities over all input states $\beta = m$ that result in the ℓth coefficient vector

$$p(\ell) = -\sum_{m|c_\ell = c_m} p_\beta(m). \quad (20)$$

If only noise is present

$$p(\mathbf{c}_m) = \begin{cases} 1 & m = 1 \\ 0 & m > 1 \end{cases}, \quad (21)$$

such that $H(\mathbf{c}) = 0$, indicating that no information is transferred. In contrast, if all $\mathbf{c}$ are unique, $p(\mathbf{c}) = p_\beta(m)$ such that $H(\mathbf{c}) = H(\beta)$, which is $log_2 M$ for equiprobable source positions. However, if all $\mathbf{c}$ are not unique $H(\mathbf{c}) \neq log_2 L$ because equipartition of probability over source positions does not correspond to equipartition of probability over unique coefficient vectors when some coefficient vectors are degenerate. This is made clear by the following reformulation of the information theoretic error metric given in (5)

$$P_e = \sum_{(m,\hat{m})} p_{\hat{\beta}}(\hat{m}|m) p_\beta(m) d(m,\hat{m}), \quad (22)$$

where $p_{\hat{\beta}}(\hat{m}|m)$ is the conditional probability of the source-location estimate given the distribution of source positions and $d(m,\hat{m})$ is the Hamming distortion measure

$$d(m,\hat{m}) = \begin{cases} 0 & m = \hat{m} \\ 1 & m \neq \hat{m} \end{cases}. \quad (23)$$

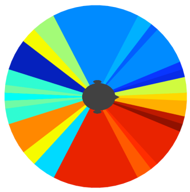
Because $p_{\hat{\beta}}(\hat{m}|m) \neq 0$ for $\hat{m} \neq m$ if all $\mathbf{c}$ are not unique, (22) indicates that $P_e > 0$ if $p(m) = 1/M$.

3. NUMERICAL RESULTS

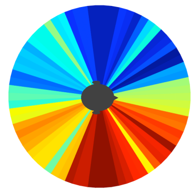
To illustrate the theory developed in Sec. 2, consider the set of HRIR for the KEMAR dummy-head microphone [19]. The set of HRIR were measured in the horizontal plane at a fixed radius from the center of the head in 5 deg. increments of azimuth ($M = 72$). A small loudspeaker served as the source and transmitted maximum-length pseudorandom binary sequences of length 16383, which were recorded at a sampling rate of 44.1 kHz, resulting in a nominal SNR of 65 dB. The impulse response of the measurement loudspeaker was removed from the HRIR measurement using a inverse filter calculated from its measured impulse response using a Mourjopoulos least-squares technique [20], yielding a response that is approximately flat over the bandwidth of the loudspeaker.

As in [12] the source waveform is a band-limited impulse. Four different source spectra are considered, unfiltered, having the full bandwidth of the measurement system, and three octave-band-filtered impulses with center frequencies of 250Hz, 1kHz, and 4kHz. Signal-to-noise ratio is specified indirectly in terms of the maximum number of quantization levels q for any coefficient in $\mathbf{c}$. While this avoids specifying the explicit relation between SNR and quantization interval ρ_0, it means that results for different source waveforms are not directly comparable.

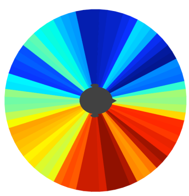
The analysis for each of the source waveforms is the same: $\mathbf{X}$ is expanded in a set of EOF in order to define α_{mn}. The vectors of discrete coefficients are then calculated according to (12) for each of the 72 source positions for $q = 1, 2, 3, 4$. In Tables 1– 4 the number of unique coefficients, the output information given by the entropy of the coefficients $H(\mathbf{c})$, and the corresponding number of distinct source positions that can be identified with arbitrary small probability of error are specified for each value of q, under the assumption of equiprobable source positions. Figures 1 – 4 display the azimuthal distribution of unique coefficients as a color



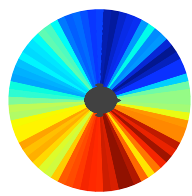
(a) q=1



(b) q=2

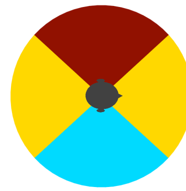


(c) q=3

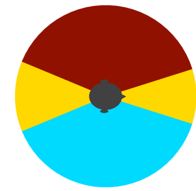


(d) q=4

Figure 1: Azimuthal distribution of unique coefficients for the full-bandwidth impulse.

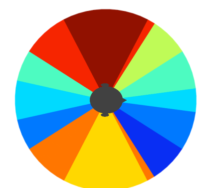


(a) q= 1



(b) q= 2

(c) q=3



(d) q=4

Figure 2: Azimuthal distribution of unique coefficients for the 250 Hz octave-band-filtered impulse.

plot for each of the four q values. In these figures each distinct color corresponds to a unique coefficient vector. Each value of q corresponds to a set of discrete coefficients $\mathbf{c}_{mn}^{(q)}$ and a set of HRIR given by

$$\hat{x}_m^{(q)}(t) = \sum_{n=1}^{N} \mathbf{c}_{mn}^{(q)} \nu_n(t). \quad (24)$$

These data were sonified by sequentially convolving the set of HRIR with a band-limited noise signal in order to generate a series of auditory displays equivalent to Figs. 1 – 4. The stimulus signal was a band-limited, time-windowed noise pulse of 150 ms total duration with 10 ms raised-cosine onset and offset transitions. The noise was band-pass filtered to have the same four source spectra as the band-limited impulses. For each of the four source spectra and each of the four values of q, the stimulus signal was convolved in sequence with the HRIR associated with each of the 72 source positions, beginning with 0 deg. source position and ending at 355 deg. source position. A 150 ms silence was inserted between the signals corresponding to each source position. The resulting binaural sounds are provided for the full-bandwidth case with $q = 1$, 2, 3, and 4, the 250 Hz octave-band-filtered case with $q = 1$, 2, 3, and 4, the 1 kHz octave-band-filtered case with $q = 1$, 2, 3, and 4, and the 4 kHz octave-band-filtered case with $q = 1$, 2, 3, and 4.

For the 250 Hz octave-band-filtered impulse, the only coefficient that carries azimuthal information is essentially a measure of interaural time difference (ITD), as shown in Fig. 5a. Finer quantization simply allows for more values of ITD to be encoded. At higher frequencies, there are multiple coefficients that carry azimuthal information, some of which are symmetric about the median plane and others that are antisymmetric, as shown in Fig. 5b. As observed in [16], 2 kHz is roughly the dividing point between the low-frequency region in which temporal cues dominate and the high-frequency region in which spectral cues dominate. However, in general, the coefficients do not correspond directly to psychophysical cues such as those described in [14].

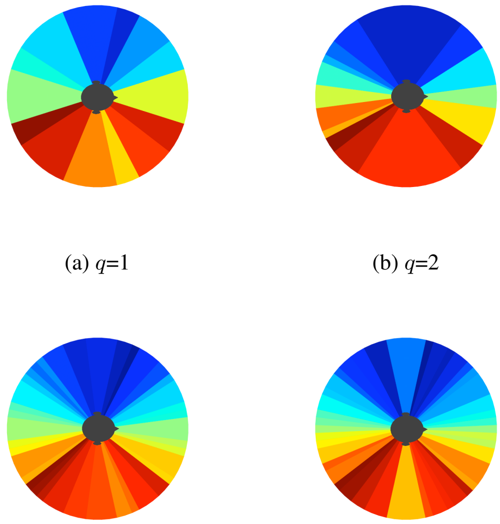
Finally, it is important to note that, because a uniform quantization interval is not optimal for the distribution of coefficient values, the increase in the number of unique coefficients with decreasing quantization interval is not strictly monotone.

4. DISCUSSION

Head-related impulse responses encode a substantial amount of information about azimuthal source location but, in general, the information density is not uniformly distributed in azimuth. Both the amount of information and the azimuthal distribution vary as a function of source signal and quantization interval. At low frequencies HRIR contain primarily ITD cues while, at higher frequencies, HRIR contain additional ILD and spectral cues. However, given a sufficiently small quantization interval, the azimuthal capacity of the HRIR exceeds that of the measured set of discrete HRIR in $\mathbf{X}$, even for the 250Hz octave band considered.

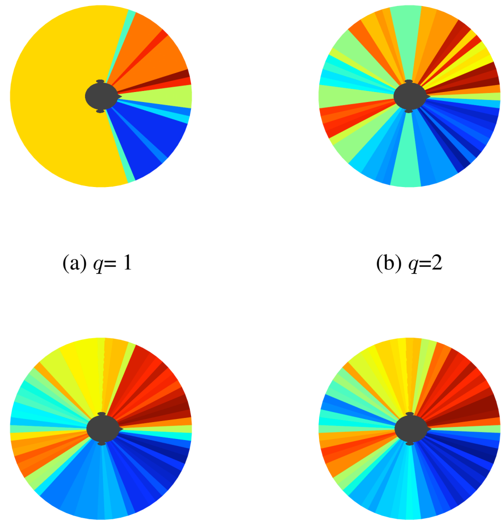
levels	unique coefficients	entropy (bits)	source positions
1	21	3.6767	12.788
2	60	5.7810	54.988
3	66	5.9823	63.219
4	72	6.1699	72

Table 1: Performance metrics for the full-bandwidth impulse



(a) q=1 (b) q=2 (c) q= 3 (d) q=4

Figure 3: Azimuthal distribution of unique coefficients for the 1 kHz octave-band-filtered impulse.



(a) q= 1 (b) q=2 (c) q=3 (d) q=4

Figure 4: Azimuthal distribution of unique coefficients for the 4 kHz octave-band-filtered impulse.

In those cases for which performance is noise (i.e., quantization-interval) limited, the number of unique coefficients is less than the number of source positions. Because the azimuthal distribution of those unique coefficients does not correspond to the distribution of source positions, the spatial information transmitted through the channel is less than its capacity. There are two possible methods for realizing the capacity of the set of HRIR to encode azimuthal information (i.e., maximize $I(\hat{\beta};\beta)$). A new partition of azimuth $\beta'(\theta)$ can be defined such that there are L discrete sources located at θ_ℓ. Alternately, the probability distribution $p_\beta(m)$ can be modified to be nonuniform so that $p(\ell)$ corresponding to the azimuths with unique coefficient vectors $\mathbf{c}_\ell$ are uniform with probability $p(\ell) = 1/L \ \ \forall \ell$. The first of these two options is related to the design of a virtual-source array for spatial-auditory display.

For example, suppose there are eight equiprobable, uniformly distributed sources but, for a given set of conditions, two source positions have coefficient vectors that are degenerate so that there are only seven unique coefficient vectors. Though the output information $H(\mathbf{c})$ is necessarily less than the input information $H(\beta) = log_2 8 = 3$, given a uniform probability distribution for each source, $H(\mathbf{c}) = 0.25 log_2 0.25 + 6(0.125 log_2 0.125) = 2.75$ is less than the maximum possible information $H(\mathbf{c}) = log_2 7 = 2.8074$ and it would not be possible to identify seven source positions with arbitrary small P_e.

As a practical example, consider the case of the full-bandwidth impulse with $q = 1$. For this level of quantization there are 21 unique coefficient vectors, as indicated in Table 1. The distribution of these coefficient vectors in azimuth, as shown in Fig. 1a, indicates that coefficient degeneracy will lead to a number of ambiguities in localization with multiple noncontiguous source locations corresponding to a single coefficient vector. This is well illustrated by the sonification of these data described previously. If, however, the number of source locations is reduced to 21 so that each of the L source positions corresponds to a single unique coefficient, the ambiguity is removed and all source positions can be resolved, as demonstrated by a sonification of the modified configuration.

Because the distribution of unique coefficients is not generally uniform in azimuth, increasing coefficient degeneracy due to decreasing SNR does not lead to a uniform loss of azimuthal resolution. Rather, information density is generally greater for frontal

levels	unique coefficients	entropy (bits)	source positions
1	3	1.5256	2.8790
2	3	1.5420	2.9119
3	7	2.7983	6.9563
4	9	3.1187	8.6860

Table 2: Performance metrics for the 250 Hz octave-band-filtered impulse

levels	unique coefficients	entropy (bits)	source positions
1	12	3.4345	10.811
2	14	3.3804	10.414
3	38	5.0441	32.994
4	40	5.1042	34.398

Table 3: Performance metrics for the 1 kHz octave-band-filtered impulse

source locations than for lateral source locations, as has been observed in psychoacoustic experiments [21]. This also can cause ambiguity in the encoding of azimuth, which leads to such phenomena as cones of confusion or the front-back confusion shown in Fig. 1a. In such cases maximum information transfer is achieved by placing only one source associated with a degenerate coefficient in one of the contiguous azimuthal sectors associated with that coefficient.

The information-theoretic framework on which the localization theory is predicated requires that all predictions be made relative to quantization interval, which is here assumed to be associated with a particular SNR. Noise in this context represents all aspects of audition that prevent perfect recovery of the spatial information encoded in the HRIR. Consequently, it cannot be measured directly and must be inferred from localization performance. To do so requires establishing a link between a model of human perception and a model of human decision. This is traditionally supplied by signal-detection theory (see, e.g., [22]).

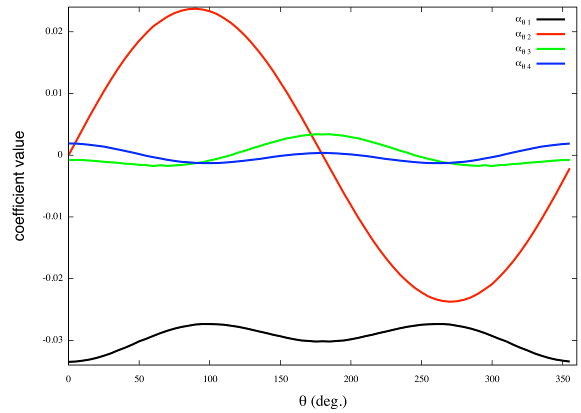
4.1. Relationship to psychoacoustic metrics of localization

While there is some recent work that applies a signal-processing method to compute performance bounds on HRIR-based localization [15], localization performance is most typically characterized through psychoacoustic metrics based on experiments with human subjects. In particular, azimuthal localization performance is characterized by absolute and relative localization thresholds. Absolute-localization experiments measure the accuracy and precision of source-location estimates. In comparison, relative-localization experiments measure acuity: the minimum audible angle of difference that can be perceived between two successive stimuli [22]. These two thresholds are linked in that the width of the distribution of absolute localization judgments (i.e., precision) is related to the relative-localization threshold [23], though in a somewhat more complex manner than suggested by a straightforward application of signal-detection theory [22], as discussed by [24].

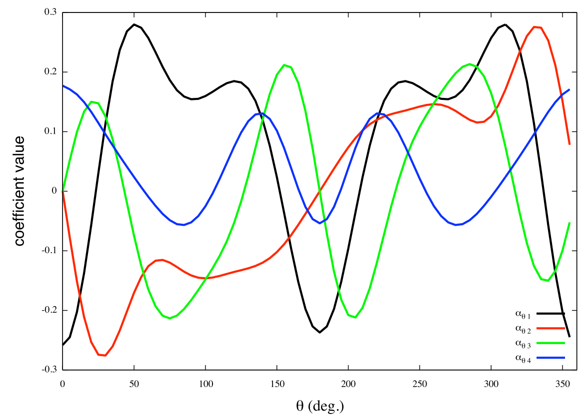
The information-theoretic performance analysis described in this paper characterizes absolute-localization. In particular it bounds the performance of the source-identification method [25], which formulates the absolute-localization task as source identification over a grid of equal-azimuth sectors. In contrast, the Cramer-Rao lower bound [15] characterizes the mean-square error of absolute localization and thus is more closely related to relative localization. One of the primary spatial-hearing tasks that an operator could perform using an azimuthal spatial auditory display is aurally detecting and estimating the bearing of a signal associated with one of many virtual sources. For this task the relevant measure of performance is absolute localization, particularly the source-identification method.

levels	unique coefficients	entropy (bits)	source positions
1	9	1.9874	3.9652
2	42	5.1175	34.7152
3	58	5.7391	53.412
4	68	6.0588	66.663

Table 4: Performance metrics for the 4 kHz octave-band-filtered impulse



(a) 250 Hz octave-band filtered impulse



(b) 2 kHz octave-band filtered impulse

Figure 5: Coefficient α_{mn} plotted as a function of azimuth (m) for $n = 1, 2, 3, 4$.

4.2. Relationship to models of localization

Unlike some other models of localization, the objective of the information-theoretic model developed here is providing insight and upper bounds on performance rather than specific predictions of localization performance. In Colburn and Kulkarni's taxonomy [17, p. 272], the model follows a signal-processing approach to localization, as contrasted with psychophysical and physiological approaches. Like other signal-processing approaches such as [22, 25], the model is not congruent with human processes. Though the coefficients α_{mn} correspond to localization cues such as ITD in some instances, the relation between HRIR and location estimate is not explicitly made in terms of psychoacoustic parameters, as in [14] or neurological correlates of localization [26]. Moreover, the model does not account for "biologic constraints" due to the auditory periphery [16] and neural processing (see, e.g., [26]) that limit the spectral and temporal resolution with which localization information can be extracted from HRIR [17]. While the internal noise level and corresponding quantization interval limit the resolution with which information is extracted from a set of HRIR, homoskedastic noise in a multidimensional space of coefficients may not be a good model for the limitations of human perception. For these reasons the model cannot provide verisimilitude to human localization, particularly its more subtle aspects. For example, it does not address performance variations due to spectral integration that arise when the spectrum of the stimulus exceeds a critical band.

5. CONCLUSIONS

Spatial hearing and localization in the azimuthal plane can be interpreted as a communication problem in which scattering from the head and torso, as described by the HRIR, encodes information about source location. Information theory gives bounds on the performance of this communication channel as a function of source signal and places an upper limit on the number of sources that can be identified for a given set of conditions comprising source signal and internal noise level. Further, it indicates how to maximize performance under those conditions. In particular, an array of virtual sources distributed uniformly in azimuth does not maximize the amount of spatial information that can be encoded from noisy observations of the set of HRIR. Instead, maximum information transfer is realized when virtual source positions correspond to those azimuthal locations which are uniquely encoded by the HRIR given the quantization interval associated with a particular set of conditions.

In future work it would be possible to extend this model to elevation-angle localization in the medial plane or to more general localization of source varying in range and over the full 4π steradian of solid angle. Because the model extracts spectral features without the need for explicitly defining them, it may be of particular use to the study of localization in the medial plane for which there is not consensus regarding the spectral cues used by human listeners [16]. Similarly, the model may offer some insight into localization in the presence of multiple reflections and reverberation, as discussed in [27].

6. ACKNOWLEDGMENT

Portions of this work have been presented at the 151st meeting of the Acoustical Society of America [J. Acoust. Soc. Am. 119, 3396 (A) (2006)]. J.E.S. gratefully acknowledges seminal contributions to this work by and ongoing discussion with C. F. Gaumond.

7. REFERENCES

[1] G. R. Arrabito, B. E. Cooke, and S. M. McFadden, "Recommendations for enhancing the role of the auditory modality for processing sonar data," *Appl. Acoust.*, vol. 66, pp. 986–1005, 2005.

[2] H. C. Hayes, "World War I—submarine detection," *Sound*, vol. 1, no. 5, pp. 47–48, 1962.

[3] R. D. Fay, "Underwater-sound reminiscences: *Mostly* binaural," *Sound*, vol. 2, no. 6, pp. 37–42, 1963.

[4] S. M. McFadden and M. M. Taylor, "Human limitations in towed arrray sonar recognition," in *Proc. 24th Defense Research Group Seminar, The Human as a Limiting Element in Military Systems*, ser. DS/A/DR(83) 170, vol. 1, May 1983, pp. 431–451.

[5] S. M. McFadden and R. Arrabito, "Proposals for enhancing the auditory presentation of sonar information," Defense and Civil Institute of Environmental Medicine, Report 94-52, November 1994.

[6] S. Richardson and C. Loeffler, "Three-dimensional auditory display of passive sonar data," in *Collected Papers, 137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustical Association*, Berlin, 14–19 March 1999.

[7] B. E. Cooke, "Uses of spatial audio in sonar," Defense and Civil Institute of Environmental Medicine, Contract Report CR 2002-054, Feb 2002.

[8] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," *J. Acoust. Soc. Am.*, vol. 83, no. 2, pp. 571–587, 1988.

[9] J. R. Buck, "Information theoretic bounds on source localization performance," in *Proc. 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM-2002)*, Washington, D.C., 2002, pp. 184–188.

[10] ——, "Fading channel capacity and passive sonar performance bounds," in *Proc. 4th IEEE Workshop on Sensor Array and Multichannel Processing (SAM-2006)*, Waltham, MA, July 2006, pp. 294–298.

[11] T. Meng and J. R. Buck, "Rate distortion bounds on passive sonar performance," in *Proc. 4th IEEE Workshop on Sensor Array and Multichannel Processing (SAM-2006)*, Waltham, MA, July 2006, pp. 636–640.

[12] C. F. Gaumond, "Broadband information transfer from oceanic sound transmission," *Acoust. Res. Lett. Online (ARLO)*, vol. 5, pp. 44–49, 2004. [Online]. Available: link.aip.org/link/?ARLOFJ/5/44/1

[13] F. G. Ashby, "Multidimensional models of categorization," in *Multidimensional models of perception and cognition*, F. G. Ashby, Ed. Hillsdale, N.J.: Lawrence Erlbaum, 1992, pp. 449–483.

[14] C. L. Searle, L. D. Braida, M. F. Davis, and H. S. Colburn, "Model for auditory localization," *J. Acoust. Soc. Am.*, vol. 60, pp. 1164–1175, 1976.

[15] S. Sen and A. Nehorai, "Perfromance analysis of 3-D direction estimation based on head-related transfer functions," *IEEE Trans. Audio Speech Lang. Process.*, vol. 17, no. 4, pp. 607–613, May 2009.

[16] C. Jin, M. Schenkel, and S. Carlile, "Neural system identification model of human sound localization," *J. Acoust. Soc. Am.*, vol. 108, pp. 1215–1235, 2000.

[17] H. S. Colburn and A. Kulkarni, *Sound Source Localization*, ser. Springer Handbook of Auditory Research. New York: Springer, 2005, vol. 25, ch. Models of sound localization, pp. 272–316.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New york: Wiley, 1991.

[19] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Lab, MIT Media Lab Perceptual Computing Technical Report 280, May 1994. [Online]. Available: http://sound.media.mit.edu/KEMAR.html

[20] A. Farina. Inverse filter of the MIT Medialab (sic.) measurement setup. [Online]. Available: http://pcfarina.eng.unipr.it/Public/Sursound/

[21] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *J. Acoust. Soc. Am.*, vol. 87, pp. 2188–2200, 1990.

[22] W. M. Hartmann and B. Rakerd, "On the minimum audible angle—a decision theory approach," *J. Acoust. Soc. Am.*, vol. 85, pp. 2031–2041, 1989.

[23] G. H. Recanzone, S. D. D. R. Makhamra, and D. C. Guard, "Comparison of relative and absolute sound localization ability in humans," *J. Acoust. Soc. Am.*, vol. 103, pp. 1085–1097, 1998.

[24] J. M. Moore, D. J. Tollin, and T. C. T. Yin, "Can measures of sound localization be related to the precision of absolute localization estimates?" *Hear. Res.*, vol. 238, pp. 94–109, 2008.

[25] W. M. Hartmann, B. Rakerd, and J. B. Gaalaas, "On the source-identification method," *J. Acoust. Soc. Am.*, vol. 104, pp. 3546–3557, 1998.

[26] N. H. Salminen, H. Tiitinen, S. Yrttiaho, and P. J. C. May, "The neural code for interaural time difference in human auditory cortex," *J. Acoust. Soc. Am. Express Lett.*, vol. 127, pp. EL60–EL65, 2010.

[27] J. E. Summers, "Information transfer in auditoria," in *Proc. 19th Intl. Congress Acoust. (ICA)*, Madrid, 2–7 September 2007.