On the Role of Head-Related Transfer Function Spectral Notches in the Judgement of Sound Source Elevation

Ewan A. Macpherson
Waisman Center
University of Wisconsin-Madison
1500 Highland Avenue
Madison, WI 53705-2280
Email: macpherson@waisman.wisc.edu

Abstract

Using a simple model of sound source elevation judgment, an attempt was made to predict two aspects of listeners' localization behavior from measurements of the positions of the primary high-frequency notch in their head-related transfer functions. These characteristics were: (1) the scatter in elevation judgments and (2) possible biases in perceived elevation introduced by front-back and back-front reversals. Although significant differences were found among the notch-frequency patterns for individual subjects, the model was not capable of predicting differences in judgment behavior. This suggests that a simple model of elevation perception based on a single spectral notch frequency is inadequate.

1 Introduction

The role of spectral cues in auditory localization is known to be significant but is as yet poorly understood. While it has been established that low-frequency interaural time difference is the primary determinant of the left-right component of perceived source position [1], no simple and reliable cue for the elevation or front-back components has been found.[1] There does exist a regular dependence of spectral notch frequency on position for the head-related transfer functions of the cat [2, 3] and somewhat similar feature motion for humans [4], and some researchers have proposed that this may be the principal elevation cue [5]. Although notch frequency clearly depends on position, it may be the case that the pattern is not as regular for humans as it is for the cat, and no causal relationship between this aspect of the physical acoustics and listener behavior has been confirmed.

The aims of the present study were to examine the differences among the notch frequency patterns of a number of individuals and to attempt to predict patterns in their elevation judgements on the basis of these differences. Predictions were made using the following simple model of elevation perception, which will be referred to as the single-notch model:

> Given that a source is localized to a particular cone of confusion (determined by interaural time difference) and to either the front or rear hemisphere (determined by

[1]The position of a source in space can be defined in a three-pole coordinate system with dimensions of left-right, back-front and elevation (up-down). The left-right dimension corresponds to the angle between the source and the vertical median plane. Sources with equal left-rightness lie on a "cone of confusion," so-called because the interaural time-difference cue is approximately constant and hence ambiguous.

some unknown spectral cue), then perceived elevation is determined by the frequency of the primary high-frequency notch in the head-related transfer function of the ear nearest the source.

Whatever plausibility this model possesses rests on the observation that contours of equal notch frequency tend to intersect each cone of confusion only twice—once in the frontal hemisphere and once in the rear. This is generally true for moderate positive and negative elevations. Musicant and Butler [6] established that spectral features due to the filtering by the near ear are the dominant cues for resolving source position on the cone of confusion. Observations made in our laboratory and by Morimoto and Aokata [7] confirm that listeners are accurate in determining on which cone of confusion a source lies and that errors are primarily made in resolving position on the cone.

Using this model and measured notch patterns, two predictions pertaining to listeners' localization judgments were made. The first concerned the variance of the elevation responses and the second, response bias under conditions of front-to-back or back-to-front confusion. The predictions were evaluated using free-field localization response data.

2 Methods

2.1 Subjects

Data were collected from six members of the Hearing Development Research Laboratory subject pool. There were three female and three male subjects ranging in age from 20 to 24. All reported normal hearing. For each subject head-related transfer functions were measured and free-field localization judgment data were collected as described below.

2.2 HRTF Notch Measurements

The procedure used to measure head-related transfer functions is described in detail by Wightman and Kistler [8]. Using probe-tube microphones positioned as close to the eardrum as possible, source-to-eardrum impulse responses were measured for positions spaced by 10° in both azimuth and elevation.

The location of the primary high-frequency spectral notch in each transfer function was located "by eye" on a computer screen plot of the spectrum and was marked using a mouse input device. Some judgment was required to select the desired notch; care was taken to follow particular features to higher elevations where they tended to dissapear. The primary notch is visible in Figure 1, which shows directional transfer functions (HRTFs normalized by the diffuse-field response) as a function of elevation at 0° azimuth for subject SNF. Note the motion of the high-frequency notches as elevation increases. Since the extraction was a time-consuming task, the analysis was limited to positions spaced by 30° in azimuth and to elevations lying between −50° and +50°. This was done for both left and right ears and resulted in 264 data points for each subject.

2.3 Free-Field Judgments

Free-field localization judgments were collected with participants seated blindfolded in an anechoic chamber. Broadband (200–14000 Hz) noise bursts of 250 ms duration were played from loudspeakers mounted on a moveable arc. Subjects responded verbally with the azimuth and elevation of the perceived source location.

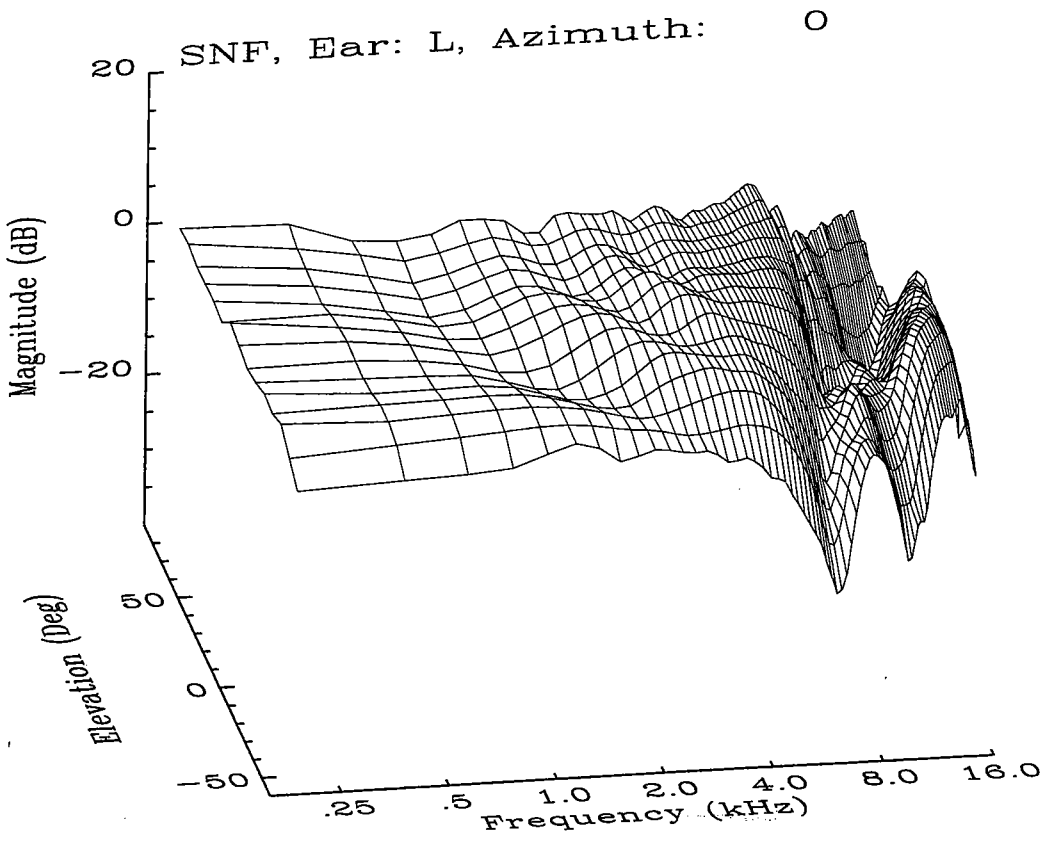

Figure 1: Directional transfer functions as a function of elevation at 0° azimuth for subject SNF.

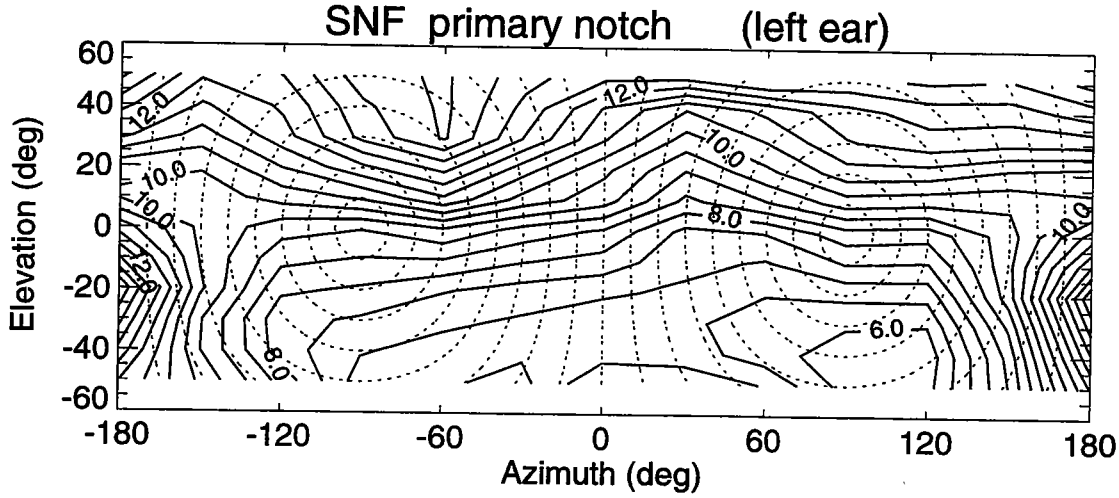

Figure 2: Contours of equal notch frequency (in KHz) for subject SNF. Dotted curves indicate cones of confusion.

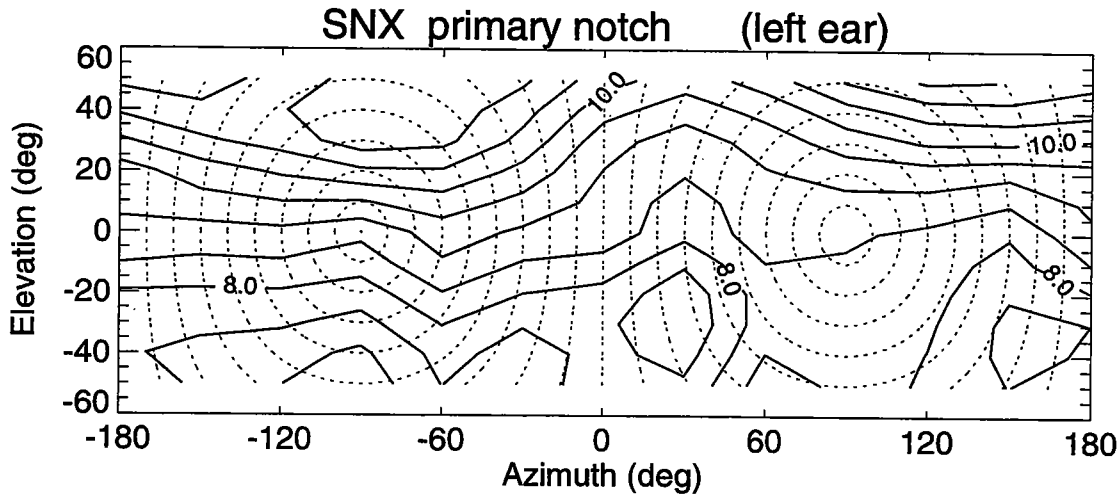

Figure 3: Contours of equal notch frequency for subject SNX.

3 Individual Notch Frequency Patterns

Contour plots of left-ear primary notch frequency as a function of direction are plotted in Figures 2-5 for four representative subjects. The dotted curves in these plots show the cones of confusion. Subjects SNF and SNX show approximately horizontal orientation of the notch contours on the ipsilateral side (negative azimuths). The contours for SNF are generally more closely spaced than those for SNX, revealing that notch frequency varies more slowly with position for the latter. Subjects SNT and SNY show upwards tilting of the contours towards the front. This is extreme in the case of SNY.

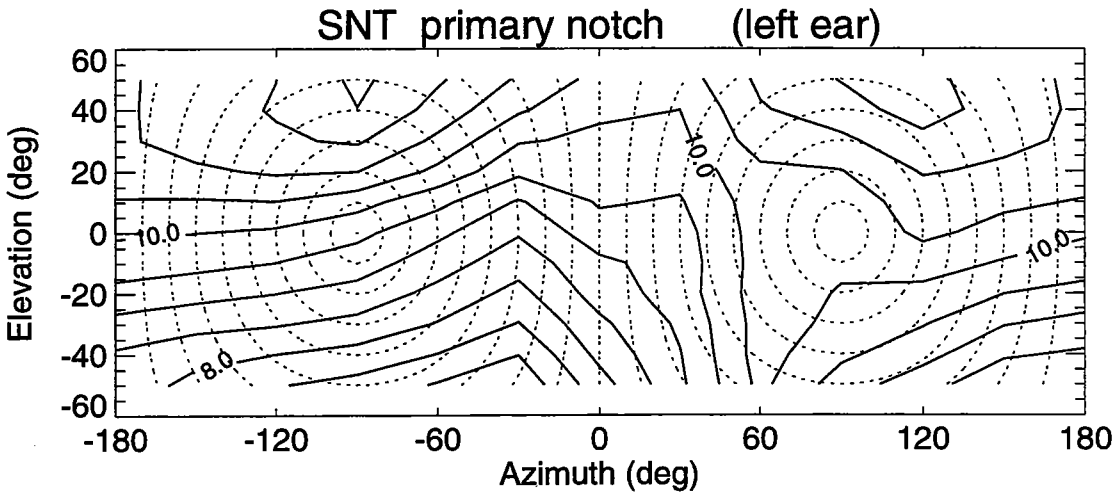

Figure 4: Contours of equal notch frequency for subject SNT.

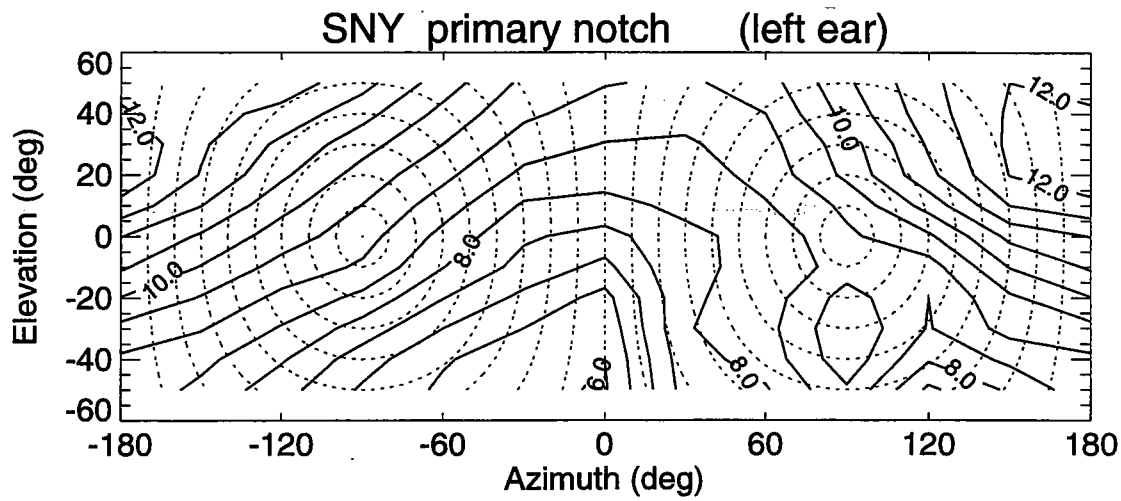

Figure 5: Contours of equal notch frequency for subject SNY.

4 Scatter of Elevation Judgments

4.1 Single-Notch Model Prediction

The first prediction generated using the single-notch model concerns the relationship between the rate at which notches change frequency with position and the variance of the elevation component of subject's responses. If notch frequency determines elevation, then subjects for whom notches move more rapidly should show less spatial scatter in their responses to individual real source locations. Uncertainty about notch frequency should correspond to relatively greater uncertainty about elevation for subjects with "slow" notches.

4.2 Data analysis

To evaluate this prediction, some measure of the spatial dependence of notch frequency for each subject was required. The magnitude of the near-ear notch frequency gradient averaged over the region of the sphere under consideration was chosen as a suitable metric of overall notch"speed." This value (∇) was calculated over all available positions and also for positions within the region of the upper hemisphere lying between $-30°$ and $+30°$ (the high-front case). The latter region was considered to be of particular interest since, unlike in the coronal plane, elevation judgment is likely to be almost entirely spectrally-based near the median plane due to the near-zero values of interaural difference cues.

To characterize the degree of scatter in subjects' judgments, the standard deviation of the elevation responses elicited by each physical source position was calculated and then averaged over the region of interest, yielding the value σ. Only responses classified as unconfused were analyzed; those deemed to be examples of front-back, back-front, or up-down reversal were excluded. The number of responses remaining at each location ranged from 4 to 7. The criteria for these classifications are discussed in Section 5.2.

4.3 Results

The results of these analyses are presented in Table 1. Linear regression revealed the correlations between ∇ and σ to be -0.21 in the overall case and -0.61 in the high-front case.

5 Bias in Front-Back Confused Elevation Judgments

5.1 Single-Notch Model Prediction

The second prediction made by the single-notch model concerns the effects of front-to-back and back-to-front confusions on elevation error. If, as in the cases of subjects SNT and SNY, the contours of constant notch frequency are tilted significantly away from the horizontal, and if the single-notch model is correct, then a front-back or back-front reversal should have a significant and consistent effect on elevation judgment errors. For example, it might be expected that SNY would experience over-elevation in cases of back-to-front reversal since the notch contours sweep upwards towards the front. Similarly, front-to-back reversals should be under-elevated.

5.2 Data analysis

The available responses for each physical source location in the left hemisphere were classified as one of: correct front (F), correct back (B), front-to-back reversed (FB), back-to-front reversed

	All Positions		High-Frost	
Subject	∇ (Hz/deg)	σ (deg)	∇ (Hz/deg)	σ (deg)
SNF	67.1	13.0	91.6	11.3
SNJ	61.9	15.5	59.6	16.0
SNR	43.5	16.5	42.5	21.4
SNT	47.0	16.6	37.2	13.2
SNX	40.1	11.8	36.3	17.4
SNY	39.1	16.3	22.9	17.8
	correlation = -0.21		correlation = -0.11	

Table 1: Averaged standard deviation of elevation judgements.

Response Type	Subjects with Horizontal Contours		Subjects with Tilted Contours	
	SNF	SNX	SNT	SNY
F	12.1	10.0	26.7	0.7
BF	17.1	1.1	↑24.7	↑2.7
B	8.4	6.1	10.3	-4.3
FB	−1.2	5.0	↓11.2	1↓2.1

Table 2: Elevation judgment bias (bias in degrees).

(BF), or up-down reversed (UD). If a judged position lay closer to the real location when reflected in the coronal plane, it was deemed to be a BF or FB confusion. Errors of elevation of greater than 90° were classed as up-down confusions and were excluded from the analysis. The mean difference between the reported and actual elevations was calculated at each position, and then these mean differences were averaged over the region of interest. This procedure was carried out for subjects SNF, SNX, SNT, and SNY, who all had orderly notch patterns and made significant numbers of front-back reversals.

5.3 Results

The results are presented in Table 2, in which arrows in cells indicate the predicted direction of the bias. Both of the subjects with more horizontal contours tend to over-elevate, although their bias patterns differ. Subject SNT does show significant over-elevation of back-to-front reversed judgments, but also over-elevates sources correctly localized in the front. SNY, for whom the notches were even more strongly tilted, shows no significant bias in any condition. The striking result is that for both of these listeners there are no differences between confused and unconfused judgments.

6 Discussion and conclusions

Although significant differences exist among the notch patterns for different subjects, attempting to predict localization behavior on the basis of these differences using the single-notch model cannot be termed a success. There appears to be no strong relationship between the average magnitude of the notch frequency gradient and response scatter either for the all-positions or the high-front case. Although the correlation coefficient of −0.61 is suggestive it is not a convincing

result, and its magnitude is due mainly to one outlying point (subject SNF). There appears to be little evidence of a relationship when the quantities are averaged over all positions. It is not surprising that the observed correlations, while low, were in the appropriate direction since the rate of notch movement with position must be positively correlated with the rate of change of overall spectral shape with position.

In the case of front-back and back-front reversals, the predictions of the simple notch model were not observed. The two subjects (SNT and SNY) with tilted notch contours had very different error bias patterns and, more importantly, showed no effect of front-back reversals on elevation judgments.

The results of these analyses clearly do not support the single-notch model of elevation perception. The observed individual differences in notch-frequency variation do not yield strong predictive power for localization behavior when coupled with this model. Therefore, elevation judgments must depend on additional spectral cues which have yet to be identified and verified.

Acknowledgments

The author is grateful to Dr. Frederic Wightman and Dr. Doris Kistler for their advice and assistance. The tedious extraction of notch-frequency from the HRTF plots was done by Kristin Andersen. This research was supported in part by grants from NIH, NASA, and ONR.

References

[1] Wightman, F. L., and D. J. Kistler. "The Dominant Role of Low-Frequency Interaural Time Differences in Sound Localization." *J. Acousti. Soc. Am.* **91** (1992): 1648–1661.

[2] Rice, J. J., B. J. May, G. A. Spirou, and E. D. Young. "Pinna-Based Spectral Cues for Sound Localization in Cat." *Hearing Res.* **58** (1992): 132–152.

[3] Neti, C., E. D. Young, and M. H. Schneider. "Neural Network Models of Sound Localization Based on Directional Filtering by the Pinna." *J. Acousti. Soc. Am.* **92** (1992): 3140–3156.

[4] Kuhn, G. F. "Physical Acoustics and Measurements Pertaining to Directional Hearing." In *Directional Hearing*, edited by W. A. Yost and G. Gourevitch, 3–25. New York: Springer-Verlag, 1987.

[5] Butler, R. A., and K. Belendiuk. "Spectral Cues Utilized in the Localization of Sound in the Median Sagittal Plane." *J. Acoust. Soc. Am.* **61** (1977): 1264–1269.

[6] Musicant, A. D., and R.A. Butler. "The Influence of Pinnae-Based Spectral Cues on Sound Localization." *J. Acoust. Soc. Am.* **75** (1984): 1195–1200.

[7] Morimoto, M., and H. Aokata. "Localization Cues of Sound Sources in the Upper Hemisphere." *J. Acoust. Soc. Japan* **5** (1984): 165–173.

[8] Wightman, F. L., and D. J. Kistler. "Headphone Simulation of Free-Field Listening I: Stimulus Synthesis." *J. Acoust. Soc. Am.* **85** (1989): 858–867.

Efficient Modeling of the Head-Related Transfer Functions

Charles Lueck
Iowa State University
Ames, IA 50011
Email: cdl@iastate.edu

Abstract

Efficient representations of the head-related transfer functions (HRTFs) based on autoregressive moving average (ARMA) modeling are explored in this chapter. A relatively new technique for ARMA order estimation, based on the minimum eigenvalues of a covariance matrix, is used to estimate efficient model structures. A major advantage of this method is that no prior estimation of the model parameters is required. Examples will be given which indicate that the HRTFs are primarily autoregressive (AR) systems.

1 Introduction

The head-related transfer functions (HRTFs) describe the position-dependent transformation in sound pressure which occurs between a source in anechoic space and the eardrums of a listener. HRTF measurements are commonly obtained as a series of time-domain impulse responses using techniques similar to those found in Wightman and Kistler [1, 2], with a pair of responses being measured for each desired source position. Simulation of virtual acoustic environments typically involves finite impulse response (FIR) filtering based on measured HRTF impulse responses. Such implementations are, however, generally not computationally efficient.

In this chapter, autoregressive moving average (ARMA) systems will be employed to identify more efficient representations of the HRTFs. An ARMA system can be described by its input-output relationship, given by the following linear, constant-coefficient difference equation of order (p, q):

$$\sum_{i=0}^{p} a_i y(n-i) = \sum_{j=0}^{q} b_j x(n-j) \tag{1}$$

where $x(n)$ is the system excitation and $y(n)$ is the system response.

Given excitation-response records for an unknown system, the system identification problem essentially becomes a two-part process. Firstly, one needs to determine the order (p, q) of the system and, secondly, one must estimate the system parameters (a_k's and $b_k's$). Estimating the order (p, q) will be the primary concern of this chapter. A recently developed technique for model order estimation based on minimum eigenvalues of a covariance matrix will be used to estimate appropriate model orders. Results will be presented which indicate that HRTFs are primarily AR systems for most source positions.

2 Minimum Eigenvalue Model Order Estimation

Assume for the moment that the observed excitation-response data satisfies Eq. (1) exactly for some undetermined orders p and q. Note that for finite-length data records, Eq. (1) can be

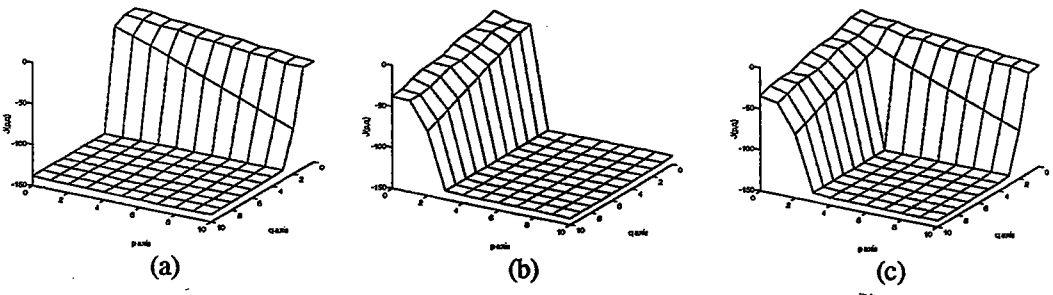

Figure 1: Minimum eigenvalue plots (in dB) for (a) an MA(3) system, (b) an AR(3) system, and (c) an ARMA(3,3) system.

rewritten in matrix form as

$$\begin{bmatrix} y(0) & y(-1) & \dots & y(-p) \\ y(1) & y(0) & \dots & y(1-p) \\ \vdots & \vdots & \vdots & \vdots \\ y(N) & y(N-1) & \dots & y(N-p) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} x(0) & x(-1) & \dots & x(-q) \\ x(1) & x(0) & \dots & x(1-q) \\ \vdots & \vdots & \vdots & \vdots \\ x(N) & x(N-1) & \dots & x(N-q) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_q \end{bmatrix} \quad (2)$$

or simply as

$$\mathbf{Y}_p \mathbf{a}_p = \mathbf{X}_q \mathbf{b}_q \quad (3)$$

in which $\mathbf{Y}_p$ is a $(N+1) \times (p+1)$ response matrix, $\mathbf{X}_q$ is a $(N+1) \times (q+1)$ excitation matrix, and $\mathbf{a}_p$ and $\mathbf{b}_q$ are respectively $(p+1) \times 1$ and $(q+1) \times 1$ column parameter vectors. Rearranging Eq. (3) yields

$$\begin{bmatrix} \mathbf{Y}_p & \vdots & -\mathbf{X}_q \end{bmatrix} \begin{bmatrix} \mathbf{a}_p \\ \cdots \\ \mathbf{b}_q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Let the data matrix $\mathbf{D}_{p,q}$ be defined to be

$$\mathbf{D}_{p,q} = \begin{bmatrix} \mathbf{Y}_p & \vdots & -\mathbf{X}_q \end{bmatrix}$$

and the corresponding covariance matrix $\mathbf{R}_{p,q}$ to be

$$\mathbf{R}_{p,q} = \mathbf{D}_{p,q}^T \mathbf{D}_{p,q}. \quad (4)$$

Using eigenvalue decomposition, the covariance matrix can be decomposed into the form

$$\mathbf{R}_{p,q} = \mathcal{Q} \Lambda \mathcal{Q}^T \quad (5)$$

in which the eigenvector matrix $\mathcal{Q}$ has as its columns the set of orthonormal eigenvectors of $\mathbf{R}_{p,q}$ and the diagonal eigenvalue matrix

$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{\text{min}}) \quad (6)$$

has as its elements the corresponding eigenvalues. Typically, the eigenvalues are ordered such that $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{\text{min}})$.

It can be shown [3] that if the "true" order of the unknown system is(n_p, n_q) and the system order (p, q) in Eq. (2) is selected such that $p \geq n_p$ and $q \geq n_q$, then the corresponding covariance matrix $\mathbf{R}_{p,q}$ will have at least one zero eigenvalue (the minimum eigenvalue) since there will exist at least one exact solution for $\mathbf{a}_p$ and $\mathbf{b}_q$ in Eq. (4). Thus, the true order of the system can be selected as the lowest order (p, q) at which the minimum eigenvalue is zero.

Typically, however, the observed excitation and response data will not exhibit a perfect ARMA relationship as indicated by (1). In such cases, the minimum eigenvalue will not equal zero once the true system order has been reached but will equal some small, nonzero value. For example, Figure 1 shows typical minimum eigenvalue mesh plots for ARMA(3,3), MA(3), and AR(3) systems. For each system, a 200 coefficient impulse response is obtained to which a low-variance, white noise is added. The excitation is assumed to be a noise-free discrete-time impulse. The covariance matrix is then formed for all model orders (p, q) such that $0 \leq p \leq 10$ and $0 \leq q \leq 10$. For each order (p, q) the minimum eigenvalue of the corresponding covariance matrix is computed and plotted in dB relative to unity, i.e., $10 \times \log_{10}(\lambda_{\min})$. Since noise has been added to the observed system response of all three systems the minimum eigenvalues do not drop to zero once the true model order has been reached, but rather to some small, nonzero value. It can be shown [3] that for large N, the increment in minimum eigenvalues will be upper bounded by a value proportional to the noise variance, although an in-depth discussion of this topic is beyond the scope of this chapter.

3 Application to HRTF Data

This technique can be applied to HRTFs to estimate the appropriate ARMA model orders. For HRTF data, the system excitation $x(n)$ is taken to be the discrete-time impulse, and the observed system response $y(n)$ is taken to be the measured HRTF impulse response, converted to a minimum-phase sequence before applying the order estimation procedure.

Figure 2 shows the minimum eigenvalue plots for left ear HRTFs of a single subject for several positions on the horizontal plane. Here, an azimuth angle of 0° indicates a position directly in front of the subject, negative azimuth angles indicate positions to the subject's left, and positive azimuth angles indicate positions to the subject's right.

Comparing the theoretical mesh plots of Figure 1 to the HRTF plots of Figure 2, it appears as though the HRTFs are primarily AR systems. Indicative of an AR system is the fact that the minimum eigenvalues drop sharply in the p axis direction and slowly in the q axis direction. This effect is observed for the HRTFs shown in Figure 3, and for HRTFs in general. To select ARMA model order estimates, an eigenvalue threshold can be used. The order at which the minimum eigenvalue drops below this threshold is selected as the estimate.

To demonstrate the AR nature of the HRTFs, a -15 dB eigenvalue threshold was used. AR model order estimates were computed using this -15 dB threshold for 450 minimum-phase, left ear HRTFs of a single subject for positions distributed evenly about a spherical shell by considering minimum eigenvalues along the p axis. Figure 4a shows a histogram of the AR model order estimates for all 450 transfer functions. The mean order estimate is 9.8 and the highest order estimate is 14. Similarly, MA model order estimates for the -15 db threshold were also computed for the same 450 HRTFs by examining minimum eigenvalues along the q axis. The results are illustrated by the histogram of Figure 4b. It is clear that the MA model order estimates are significantly higher at this threshold than the AR model order estimates.

At lower thresholds, however, the difference between AR and MA model order estimates is reduced due to the fact that the MA model shows a gradual decrease in minimum eigenvalues

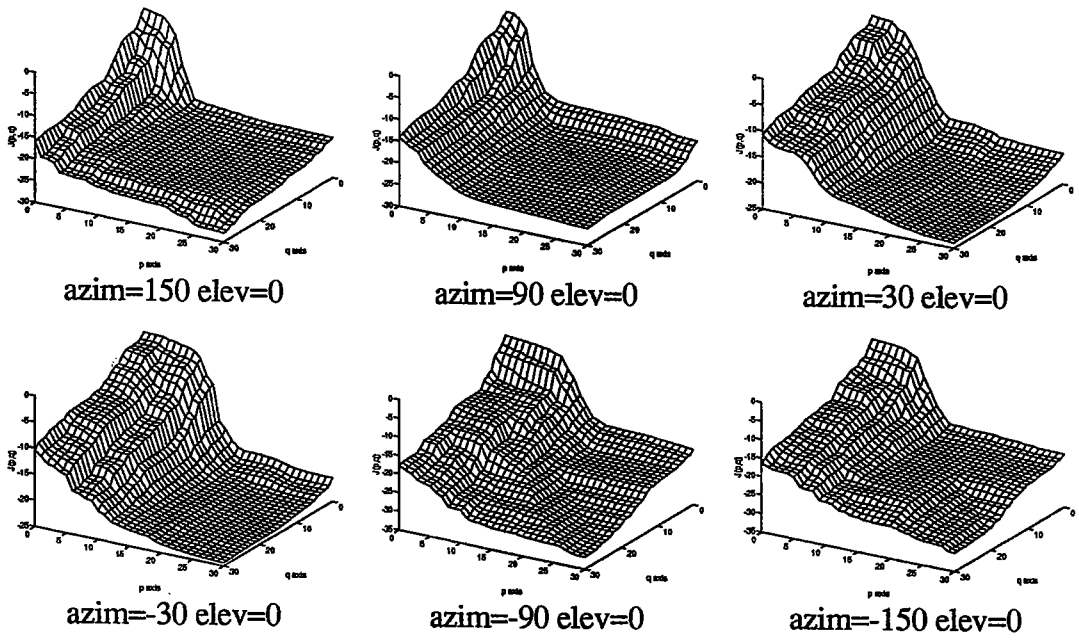

Figure 2: Log-scale minimum eigenvalue ploits for HRTF functions.

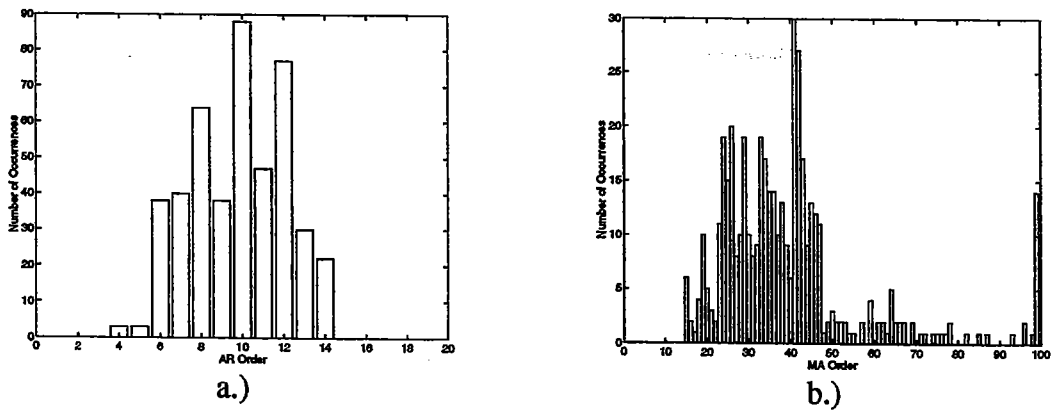

Figure 3: Histograms of -15 dB (a) AR order estimates (b) MA order estimates.

(along the q axis) while the AR model shows little decrease above 15th order (along the p axis). Thus, the AR model appears to have its greatest advantage over the MA model at low model orders. The question arises as to the perceptual contribution of the AR model for sound localization. Lowering the eigenvalue threshold corresponds roughly to lowering the allowable modeling error variance, an objective criteria which may not directly reflect the perceptual performance of the model. Consequently, perceptual listening tests are needed to fully assess the significance of these results.

Conclusions

The use of autoregressive moving average models for efficient representation of the head-related transfer functions has been explored in this chapter. Using a minimum eigenvalue model order estimation technique, the HRTFs have been shown to be primarily AR systems. These results suggest the implementation of HRTFs using autoregressive filters or possibly the use of an autoregressive prefilter for efficient computation, although the validity of such an implementation has yet to be subjectively demonstrated.

Acknowledgments

This work was supported, in part, by a Rockwell International Corporation Fellowship. The author would like to thank Fred Wightman and Doris Kistler at the University of Wisconsin for providing the HRTF data used in this chapter.

References

[1] Wightman, F., and D. Kistler. "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis." *J. Acoust. Soc. Am.* **85** (1989): 858–867.

[2] Wightman, F., and D. Kistler. "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation." *J. Acoust. Soc. Am.* **85** (1989):868-878.

[3] Liang, G., D. Wilkes, and J. Cadzow. "ARMA Model Order Estimation Based on the Eigenvalues of the Covariance Matrix." *IEEE Transactions on Signal Processing* **41** (1993): 3003–3009.