1995

Modeling of the head-related transfer functions for reduced computation and storage

Charles Duncan Lueck
Iowa State University

Modeling of the head-related transfer functions for reduced computation and storage

by

Charles Duncan Lueck

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Electrical and Computer Engineering
Major: Electrical Engineering (Communications and Signal Processing)

Approved:

Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1995

TABLE OF CONTENTS

1. INTRODUCTION

Recent advances in audio and video processing have accelerated the emergence and development of the three-dimensional virtual display. The main objective of the virtual display is to present aural and visual information in such a manner that it is realistically perceived by the user. Hence, the term "virtual reality" has become a popular buzzword used to describe the technology. Still in its infancy, the virtual display shows signs of great potential, opening the door to a host of innovative applications in areas ranging from control systems to education to entertainment. However, the technology required to fully achieve this potential has not yet been realized, and will not be realized without significant additional research.

The research presented in this dissertation will be directed toward the implementation of the *virtual acoustic display*, a device for the synthesis of three-dimensional sound via headphones. The virtual acoustic display typically involves the implementation of digital filters which emulate the head-related transfer functions (HRTFs) [1-3]. These functions represent the transformation of sound pressure from a sound source in free space to the eardrums of the listener. Current implementations generally rely on FIR filtering techniques which result in high computation and storage demands, particularly for real-time systems [2]. Because these demands directly translate into increased cost of implementation, reduction of computation and storage requirements is of great concern.

1.1 Applications

Three-dimensional audio technology has generally not received the attention of its three-dimensional visual counterpart. However, this is also an area of great importance and intensive research. Scientists and engineers have been working to understand the complex physical and psychological principles involved in human sound localization and to apply this understanding toward the development of a virtual acoustic display for conveying three-dimensional sound through headphones. Such an acoustic display would prove beneficial to a wide variety of applications. Among these might be:

- **Entertainment/Recording** [4,5]. Undoubtedly, virtual acoustics will find its way into entertainment applications, both at the consumer level and commercially. Potential exists for the enhancement of movies, musical recordings, video games, and other multimedia applications. True three-dimensional acoustic reproduction would create a level of realism unmatched by today's audio reproduction techniques.
- **Tool for Scientific Research** [2,3]. Headphone-delivered stimulus allows researchers precise control over the sound pressure at the eardrums of the listener. The ability to provide simulated signals equivalent to those heard in a realistic environment is unquestionably a valuable scientific tool in the study of the physical and psychological aspects of sound localization.
- **Aircraft control and communications** [1,4,6]. Aviation applications would particularly benefit from the incorporation of three-dimensional acoustic displays. In air traffic control and collision-avoidance systems, where the operator's spatial awareness is critically

important, three-dimensional acoustic cues would likely increase efficiency and quicken response times. In the cockpit, localized acoustic cues would instantly convey to the pilot information regarding the proximity and location of other aircraft, obstacles, and targets.

- **Teleconferencing/Telecommunications** [5]. The effectiveness of remote communication could certainly be enhanced by the application of three-dimensional acoustics. By providing directional cues to speech, individuals would have the perception of being in the same room as other participants, when, in fact, they might actually be separated geographically by large distances.
- **Guidance systems for the blind** [4]. Another application which has been proposed is the incorporation of a virtual acoustic display as a guidance system for the blind. In this system, the global positioning system (GPS) would be used to provide location information to the system. The guidance system would then provide audible cues indicating one's position in relation to landmarks, to the location of buildings, or to the proximity of a busy street or intersection.

In general, three-dimensional acoustic displays would be valuable in any situation in which one's spatial awareness is important, or in any situation in which the added dimension of localized audio enhances the efficiency, performance, impact, or enjoyment of the intended environment.

1.2 Objectives

The currently employed FIR filtering techniques require the use of costly, high-speed signal processors for real-time implementation [1]. The primary purpose of the research presented herein was to develop digital filters having reduced computation and storage requirements which would be capable of emulating the head-related transfer functions without significantly sacrificing perceptual performance.

With this ultimate goal in mind, the following objectives need to be achieved:

- **Computation reduction.** The amount of computation required to implement the head-related transfer functions must be reduced. One approach to achieving this goal would be to attempt to reduce the order of the filters being used. A second approach might be to improve the computational efficiency of the filtering process itself.
- **Data reduction.** Storing the coefficients of high-order filters for a large number of source positions would require a large amount of memory. A simple technique for reducing storage requirements would again be to reduce the order of the filters being used. Another useful technique would be to interpolate coefficients over a sparse set of filters.
- **Demonstration of perceptual validity.** The reduction of computation and storage requirements must not be made at the risk of perceptual performance. Perceptual listening tests must be performed using the developed filtering techniques to demonstrate the preservation of localization performance.

- **Demonstration of feasibility.** The demonstration of feasibility in the form of a real-time filter implementation must be performed to show that the developed filtering methods can be successfully applied to real-world applications.

An outline of the approach used in this dissertation to achieve these objectives is contained in the summary below.

1.3 Summary

Chapter 2 presents a brief background on sound localization theory. The purpose of this presentation is to familiarize the reader with past research involving the physical and psychological aspects of human sound localization. A number of auditory cues associated with sound localization and the extent to which each is a factor are discussed. Two commonly known cues are the interaural time differences (ITDs) and the interaural intensity differences (IIDs). The ITDs represent the direction-dependent difference in time of arrival of a sound pressure wave between the right and left ears. Similarly, IIDs represent the overall intensity difference between the ears. Pinna cues, resulting from the interaction of an incoming sound wave with the complex folds of the outer ear, are another important factor. The head-related transfer functions (HRTFs), which describe the position-dependent transformation of sound pressure between a sound source and the eardrums, are introduced and the role in which the HRTFs play in the synthesis of three-dimensional sound is described.

In Chapter 3, a brief introduction to system modeling is presented. The purpose of this introduction is to demonstrate the application of conventional modeling techniques to the modeling of an arbitrary system. The system modeling problem essentially involves two major issues: estimation of model parameters and determination of the system model order. For parameter estimation, the least-squares method, the extended Prony's method, and the iterative prefiltering method are discussed. In addition, system modeling based on balanced state-space realization and model reduction techniques is described. The various strengths and weaknesses of each technique are presented. For model order estimation, a relatively new method referred to as the minimum eigenvalue model order estimation technique is presented. In this technique, the model order is selected based on eigenvalues of a covariance matrix formed from the observed excitation and response data of the system. Details are provided in Section 3.3.

Chapter 4 presents techniques for modeling the head-related transfer functions. Section 4.1 discusses the approximation of the head-related transfer functions as minimum-phase systems. The transfer function of a minimum-phase system will have all of its poles and zeros located within the unit circle of the complex plane. Many systems can be realized which have the same magnitude response, but only one such system will be minimum phase. The minimum-phase approximation represents a processing convenience because the phase spectrum can essentially be ignored. It is shown that neglecting the phase spectrum can reduce the order required to model a system.

Section 4.2 examines model order estimation of the HRTFs. Application of the minimum eigenvalue model order estimation technique indicates that the HRTFs are essentially autoregressive systems. A factorization of the HRTFs is presented which factors the HRTFs into directional components, referred to as the directional transfer functions (DTFs), and a position-independent component, referred to as the diffuse-field. Application of the minimum eigenvalue model order estimation technique to the directional transfer functions indicates that the DTFs are primarily autoregressive moving-average systems with equal order numerator and denominator polynomials. Since the diffuse-field is independent of source position, the system modeling effort focuses on the modeling of directional transfer functions.

Section 4.3 proposes a number of techniques for parameter estimation of DTF system models. Conventional parameter estimation techniques which minimize an objective error criterion do not take into account the perceptual error associated with the human hearing process. Standard modeling techniques which minimize an unweighted error energy criterion tend to provide a uniform spectral fit across a linear frequency scale, but will provide a poor spectral fit across an auditory frequency scale. As a result, conventional modeling techniques typically result in models which perform poorly on an auditory basis. The auditory scale is a nonlinear function of frequency, i.e., the frequency resolution of the human ear decreases with increasing frequency. For example, the perceptual difference between a 500 Hz tone and a 1000 Hz tone is much more significant than the perceptual difference between a 15,000 Hz tone and a 15,500 Hz tone, even though the tones differ by 500 Hz in both cases. In Section

4.3.2, a critical band frequency scale is described which more closely resembles the auditory scale.

The process which leads to critical band effects can be viewed as a convolution, in the frequency domain, of the magnitude spectrum of the signal incident on the eardrum with a spectral smoothing window that increases in width as frequency is increased. In Section 4.3.3, a technique referred to herein as critical band smoothing is introduced to simulate the spectral spreading process associated with hearing. Details are discussed in Section 4.3.3. The effect that critical band smoothing has on the system modeling process will be studied in detail. Application of several of the modeling techniques presented in Chapter 3, the least-squares method, the iterative prefiltering method, and the balanced model reduction method, to a sample directional transfer function which has been smoothed using critical band smoothing is illustrated. In addition, the application of two other techniques, the least-squares weighted error method and the weighted iterative prefiltering method, is introduced. Each of these techniques estimates system model parameters based on a weighted error criterion in which the weighting function is designed to provide a better model fit on an auditory frequency scale.

In Section 4.3.4, a critical band distance measurement is introduced as a objective error measurement for evaluating the performance of various modeling techniques. The critical band distance measurement measures the Euclidean distance between two spectra expressed as critical band levels in dB. Details of this procedure are described in Section

4.3.4. The goal of this critical band distance is to provide an objective measurement which directly reflects the perceptual error between two spectra.

Results obtained using the modeling techniques described in Chapter 4 are presented in Chapter 5. These results include objective measurements, in the form of critical band distances, and subjective measurements, in the form of localization tests. In Section 5.1, critical band distance measurements will be examined for a wide range of model orders and varying degrees of smoothing so that a comparison of modeling techniques can be made. It is shown that critical band smoothing tends to increase the ability of a given modeling technique to fit a desired spectrum on an auditory frequency scale. However, it is also shown that critical band smoothing introduces a certain amount of perceptual distortion to the original system.

In Section 5.2, the computational requirements, in terms of number of multiply operations, for each technique will be presented for a number of filter implementation structures. The amount of computation required to meet a given critical band distance constraint will be used to rank each technique. Although many of the described system modeling techniques do not result in a significant reduction in computation in comparison to FIR filtering, systems modeled using critical band smoothing followed by balanced model reduction are found to result in a considerable reduction in computational effort.

Results from localization tests demonstrate the degree to which the localization performance is preserved by the system models. The localization tests compare the

performance of high-order FIR filters to the performance of low-order models designed using two of the techniques of Chapter 4. A description of these tests is provided in Section 5.3.

Conclusions drawn from these results are presented in Chapter 6. A discussion of future work is also be included.

2. SOUND LOCALIZATION THEORY

The purpose of this chapter is to provide the user with a brief background in sound localization theory. Within the context in which it will be used here, *sound localization* refers to the process by which the location (distance and/or direction) of an auditory event is related to specific attributes of a sound event [7]. Here, a distinction is made between a sound event and an auditory event. The term *sound event* is associated with that event which causes an audible disturbance. The term *auditory event* refers to the *perception* of that event. That is, the location of an auditory event is the perceived location of a sound event. The position of auditory events and the position of sound events do not always coincide. In fact, a goal of the virtual acoustic display is to place the auditory event at a predefined location, even though the actual sound sources in a pair of headphones are positioned just outside the entrance to the ear canals. The position of the auditory event is directly dependent on both the way in which the sound pressure is transformed by the head, torso, and outer ear and the way in which the signals received at the eardrums are interpreted by the brain.

In a free-field environment with real sources, one might express localization performance as a function of how well the position of an auditory event matches the position of the sound event. In a virtual environment, localization performance might similarly be expressed as a function of how well the position of the auditory event matches the intended source position. The success of the auditory display depends on the extent to which the

localization performance of the virtual environment matches that of the real-world environment.

For many years researchers have been studying the physical and psychological aspects of sound localization. Important to the understanding of sound localization is the determination of the specific attributes, or auditory cues, which directly affect the location of the auditory event. In this chapter, a brief review of localization literature will be presented. In so doing, a number of auditory cues important for sound localization will be revealed. Also, in this chapter, the definition of the head-related transfer functions will be introduced and the theory behind the simulation of free-field listening will be presented.

In this document, a consistent spherical coordinate system has been adopted to indicate source positions as illustrated in Figure 2-1. The position directly in front of the subject corresponds to an azimuth of 0° and an elevation of 0°. Source positions on the subject's right side correspond to positive azimuth angles with 90° representing a source to the subject's immediate right and 180° representing a source behind the subject. Similarly, positions on the subject's left side correspond to negative azimuth angles with -90° indicating a source to the subject's immediate left. For elevation, positive angles correspond to sources above the horizontal plane with 90° corresponding to a source directly above the subject, and negative angles correspond to sources below the horizontal plane. The *horizontal plane* will refer to the horizontal plane which contains the axis joining the entrances to the ear canals. Similarly, the *median plane* will refer to the plane of symmetry of the human head. The plane

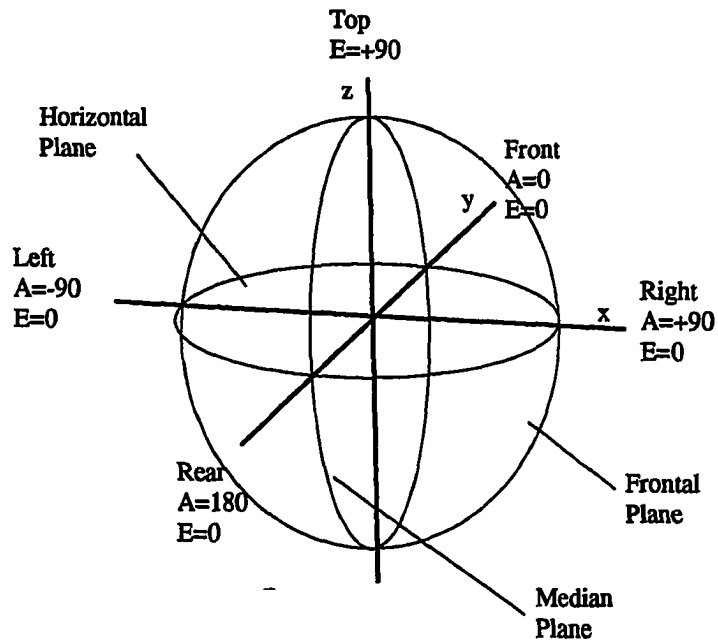


Figure 2-1 Adopted coordinate system.

which is orthogonal to both the horizontal plane and the median plane and which contains the axis joining the entrances to the ear canals will define the *frontal plane.*

2.1 Background

It has long been known that interaural time differences (ITDs) and interaural intensity differences (IIDs) play important roles in the ability to localize sounds [7]. However, these factors are not the only indicators of source location. Interaural time and intensity cues alone cannot account for one's ability to localize sounds at positions on the median plane, where the time and intensity differences between the right and left ears are minimal. Headphone-delivered stimuli strictly utilizing these two cues generally result in the perception of sound

sources inside the listener's head. For this reason, a distinction is often made between lateralization (inside-the-head localization) and localization (outside-the-head localization) [3].

It has been suggested that the folds of the pinna play an important role in sound localization [7-10]. The interaction of a sound wave with the complex folds of the pinnae results in a position-dependent filtering. The presence of pinna cues is also thought to increase localization accuracy (in particular, determination of elevation angle and discrimination between front and back positions) and, in addition, contribute to the externalization of sound.

The importance of pinna cues was demonstrated by Batteau [10] as described in [6]. In Batteau's experiment, a wideband stimulus was recorded dichotically using microphones which were inserted into a pair of artificial pinnae. The pinnae were separated by a distance approximately equal to the width of a human head. Listeners in another room presented with the recorded stimulus were then asked to report apparent azimuth and elevation for a variety of positions. The experiment was also performed using microphones which were not equipped with pinnae. The results indicated that the use of the artificial pinnae resulted in a significantly higher localization accuracy. In addition, when using the artificial pinnae, listeners reportedly perceived the source as being located outside of the head as opposed to inside of the head.

Other researchers have experimented with binaural recordings made using "dummy" heads [6,8]. In these experiments, stimuli were recorded using microphones placed in the ear canals of dummy heads. Listeners were then presented the recorded stimulus using

headphones. These studies also generally reported realistic perception of sound. In one such study by Plenge [8], it was concluded that pinna cues played a primary role in the externalization of sound.

Other factors, such as reflections and head movements, are also thought to contribute to the ability to localize sounds [11-13,37]. In [11], a study of the effect of head movements on sound localization accuracy was presented. The results presented indicated an increased accuracy in the estimation of source elevation and a significant reduction in the number front-back reversals. Additional work has been done to study localization of sound in reverberant fields. In [7], it was reported that binaural recordings in reverberant fields produced a "nearly perfect reproduction" of auditory events. In [37], however, it was concluded that reverberation effects did not significantly improve localization accuracy.

A number of researchers have attempted to measure the transformation in sound pressure which occurs between a source in anechoic space and the ear canals or eardrums of the listener [2,3,10,14]. This transformation is commonly expressed in the form of position-dependent transfer functions, generally referred to as the *head-related transfer functions* (HRTFs). A pair of HRTFs (one for each ear) relates the sound pressure of a source at a particular position in space to the sound pressure at the eardrums of a listener. Measurement of head-related transfer functions typically involves the use of probe tube microphones inserted into the ear canals of a subject seated in an anechoic chamber. In some instances the transfer function measurements are carried out using a dummy head. In most situations,

however, live subjects are used because the measured HRTFs more closely resemble the subjects' actual HRTFs [2,10,14].

2.2 Measurement of the Head-Related Transfer Functions

The specific technique used for measuring the HRTFs presented in this document is that of Wightman and Kistler [2,3]. The steps in this technique will be briefly outlined below; details of this procedure can be found in [2], although some minor modifications have been made since that time. Only the basic steps of the procedure will be summarized here.

To make HRTF measurements, a subject is first fitted with a pair of miniature electret microphones to which have been attached tiny silicone rubber probe tubes. The ends of the probe tubes are carefully inserted into the subject's ear canals. The purpose of this probe microphone system is to measure the sound pressure level near each eardrum of the subject.

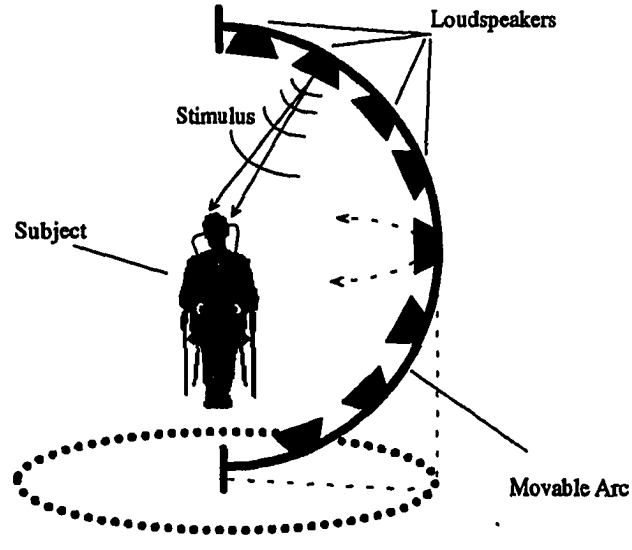


Figure 2-2 HRTF measurement setup.

Once fitted with microphones, the subject is seated in an anechoic chamber at the center of a large semi-circular arc (2.76 m in diameter) upon which a series of loudspeakers have been mounted, as illustrated below in Figure 2-2. The arc can be rotated 360° to produce a stimulus at any azimuth angle. The loudspeakers on the arc have been mounted at fixed intervals, so that stimulus can be provided for a number of discrete elevation angles.

For each source position measured, the arc is positioned at the desired azimuth angle. A wideband stimulus is then produced through the loudspeaker corresponding to the desired elevation angle. The stimulus is digitally constructed to have a flat magnitude spectrum from 200 Hz to 4000 Hz. Above 4000 Hz, the magnitude spectrum increases abruptly by 20 dB and is then flat up to 15,000 Hz. No energy is contained below 200 Hz or above 15000 Hz. The phase spectrum is computed so that the peak factor of the signal is minimized, as described in [15]. A 16-bit, 100 kHz D/A converter is used to produce the stimulus, while a pair of 16-bit, 100 kHz A/D converters are used to simultaneously sample the responses at both eardrums as measured by the probe microphones. For improved signal-to-noise ratio, the ensemble average of the responses is computed over a number of repetitions of the stimulus at each source location.

Once the response data has been collected for all desired source positions, the transfer function characteristics can be computed. Each response is deconvolved with the stimulus (via division of FFTs) to produce the corresponding impulse response. A correction is made for the frequency response properties of the microphones. Sample right and left ear HRTF impulse responses and magnitude response characteristics for one subject are shown below in Figure 2-3 for an azimuth of 90° and an elevation of 0°. For the impulse responses shown,

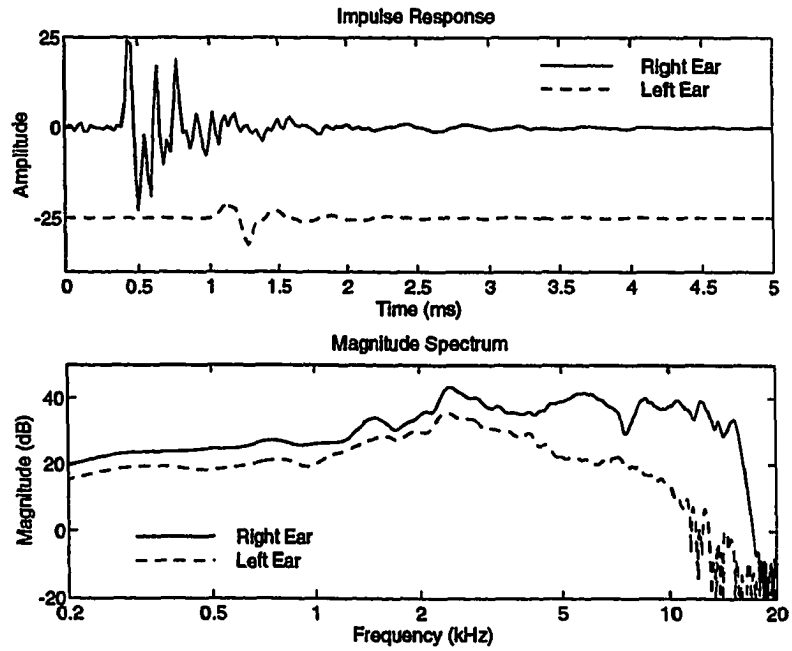


Figure 2-3 Sample right and left ear HRTFs for azimuth = 90° and elevation = 0°.

note the difference in time and amplitude between the left and right ears. Since the sound source in this case was on the subject's right side, the time delay is less for the right ear than it is for the left, as would be expected. The HRTF spectra generally contain a number of uneven peaks and valleys, indicating a rather complex frequency response.

2.3 Synthesis of Localized Sound

To perform the simulation of free-field listening conditions using headphones, a source is digitally filtered using the pair of HRTFs corresponding to the desired source position. This is most simply accomplished using the values of the impulse response as the coefficients of an

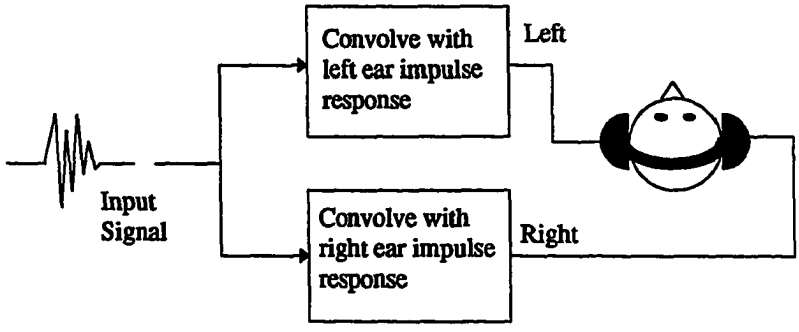


Figure 2-4 Synthesis implementation.

FIR (finite-impulse response) filter. The resulting two-channel signal is then converted via 16-bit, 100 kHz D/A converters. The subject, listening through headphones, will ideally perceive the source as originating from the corresponding position in space. This situation is depicted in Figure 2-4.

The use of this technique to simulate free-field listening conditions with headphone-delivered stimulus has been shown to be largely successful. Such results have been presented in [3]. Using a group of inexperienced listeners, the spatial judgments of a wideband stimulus presented in free-field were compared to the spatial judgments of the same stimulus presented through headphones. The location of headphone-delivered stimuli were judged to be in nearly the same position as the free-field stimuli, indicating the simulation technique to be perceptually sufficient. There was, however, an increased number of front-back reversals (to

be discussed below) than in the free-field case. In addition, the performance in estimation of elevation was generally poorer than in free-field.

2.4 Performance and Implementation Concerns

There are a number of "challenges" remaining before three-dimensional reproduction of sound is successfully developed for widespread use [16]. Among these are:

- **Compensation for inter-subject variability.** Comparison of HRTFs indicates that a fairly large amount of variability exists among subjects. In general, performance is best when a subject uses HRTF filters which have been constructed using his/her own measurements. This observation seems to indicate that individual differences may make it difficult to produce a single set of filters which will work well for all listeners.
- **Reduction of front-back reversals.** A headphone-delivered stimulus intended for a 20° azimuth angle (front, and slightly to the right) may be perceived by the listener as being at 160° (back, and slightly to the right). For a pair of source positions in which interaural time and intensity differences are essentially equal, there is an increased chance that the listener will perceive the unintended source position. Such a reversal is given the term front-back reversal. In a free-field situation, a certain percentage of reversals is expected. In a virtual situation, it would be undesirable to have this percentage increase significantly.
- **Reduction of computational load.** Computational requirements are obviously important in a real-time implementation. At present, FIR filtering using the HRTF impulse responses

requires a large computational effort. More efficient filtering techniques could greatly reduce the computational requirements of such an implementation.

- **Reduction of data requirements.** Storing the filter coefficients for all desired source positions becomes rather impractical. For example, an implementation which stores the HRTF data for 36 azimuth angles (10° intervals) and 12 elevation angles (15° intervals) would require storage for over 400 pairs of filters. Although implementations with a fully stored filter coefficient bank can be constructed, interpolation techniques could be used to reduce the memory requirements of such a system and thereby reduce the cost of implementation.

The research presented herein focuses primarily on reducing the computation and data requirements necessary for implementation. Solutions for dealing with inter-subject variability and front-back reversals, while important, are beyond the scope of the proposed work and, consequently, will not be directly dealt with further.

3. INTRODUCTION TO SYSTEM MODELING

3.1 System Modeling Problem

A frequently encountered problem in many science and engineering disciplines is one of determining an appropriate model for an unknown system given observed excitation and response data. In this chapter a brief introduction to system modeling will be presented. For this work, one will be primarily concerned with the modeling of systems based on observed impulse responses. A number of techniques for estimating model parameters will be discussed, one of which will be applicable only to observed impulse response data and several of which will be applicable to arbitrary excitation and response data. Given an observed system impulse response, any of the techniques discussed can be applied by assuming the input to be a discrete-time impulse and the output to be the observed impulse response.

Consider an arbitrary unknown system which produces an output $y(n)$ when an arbitrary input $x(n)$ is applied. This situation is described in Figure 3-1. In system modeling, one wishes to find a parametric model relating the observed excitation to the observed response. A common approach to this problem is to assume that the transfer function of the unknown system can be represented by a rational system and to solve for the system parameters. A rational system can be defined by the following linear, constant-coefficient difference equation of order (p,q):

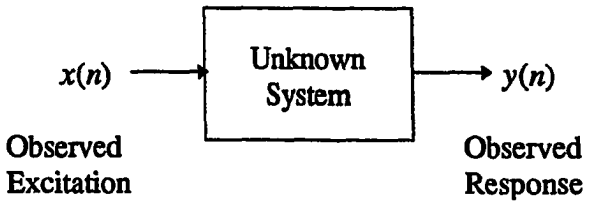


Figure 3-1 Excitation and response of an arbitrary system.

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

where $x(n)$ is the system excitation and $y(n)$ is the system response [17]. The a_k's are commonly termed the denominator or autoregressive (AR) coefficients and the b_k's are termed the numerator or moving average (MA) coefficients. The leading denominator coefficient a_0 is commonly taken to be equal to 1 without any loss of generality. The system described by (1) has a transfer function of

$$H(z)=\frac{B(z)}{A(z)}=\frac{b_0+b_1z^{-1}+b_2z^{-2}+\cdots b_qz^{-q}}{1+a_1z^{-1}+a_2z^{-2}+\cdots a_pz^{-p}}. \tag{2}$$

Such a system can be used to describe linear, time-invariant systems. Factoring the numerator and denominator polynomials yields a pole-zero transfer function of the form

$$H(z)=\frac{b_0(1-d_1z^{-1})(1-d_2z^{-1})\cdots(1-d_qz^{-1})}{(1-c_1z^{-1})(1-c_2z^{-1})\cdots(1-c_pz^{-1})} \tag{3}$$

where the c_i's are the termed the system *poles* and the d_i's are termed the system *zeros*.

Two important system properties can be determined directly from the pole-zero form given by (3). The system will be *stable* if the poles of the system have magnitudes less than one, i.e., all poles lie within the unit circle [18]. If, in addition, all zeros have magnitudes which are less than one, the system is said to be *minimum phase* [18]. Since the transfer function is a rational function, the term *rational system modeling* is often used to describe the modeling process. Another common name is *pole-zero* modeling.

If the order of the denominator polynomial is zero, the system is commonly referred to as an all-zero or moving-average system. Similarly, if the order of the numerator polynomial is zero, the system is referred to as an all-pole or autoregressive system.

Identification of the unknown system essentially involves two major issues. The first is the estimation of the system parameters. The second is the selection of an appropriate model order. In practical situations, the observed excitation and response data will generally not exhibit a perfect pole-zero relationship as indicated by (1). In such cases, the parameters of the model are chosen to minimize a specified error criterion. As will be the shown in the next section, the selection of the error norm has a significant impact on the complexity of the solution and on the characteristics of the resulting model.

3.2 Parameter Estimation

3.2.1 Direct modeling

Perhaps the most intuitive approach to system modeling would be to develop a model which produces an output which approximately matches the observed output of the unknown system for a given excitation. In the case in which the discrete-time impulse is used as the excitation, this approach is equivalent to approximately matching the impulse response of the model to the impulse response of the unknown system.

Define the *modeling error* $e_m(n)$ to be the difference between the output of the unknown system $y(n)$ and the output of the model $\hat{y}(n)$ [17]. The modeling error is thus given by

$$e_m(n)=y(n)-\hat{y}(n) \tag{4}$$

where

$$\hat{y}(n)=-\sum_{i=1}^{p}a_i\hat{y}(n-i)+\sum_{j=0}^{q}b_jx(n-j) \tag{5}$$

is the response of the model due to excitation $x(n)$. This error is depicted in Figure 3-2. In z transform notation the error can be expressed as

$$\begin{aligned} E_m(z) &= Y(z) - \hat{Y}(z) \\ &= Y(z) - \frac{B(z)}{A(z)}X(z) \end{aligned} \tag{6}$$

In the direct modeling problem the criterion for estimation of the model parameters is minimization of the modeling error energy. Unfortunately, however, the modeling error energy in the direct modeling problem is a nonlinear function of the model parameters. This makes the solution of the parameters by direct means difficult, and a direct solution is not feasible. While iterative techniques can be applied to the solution of such a problem, the computational complexity required and the difficulty in obtaining a globally optimum solution makes the direct modeling approach unattractive for many applications.

3.2.2 Indirect modeling

To get around these difficulties, the modeling problem can be reformulated by modification of the error norm into one which has a solution obtainable from a set of linear equations [17]. Define the *equation error* $e_q(n)$ as

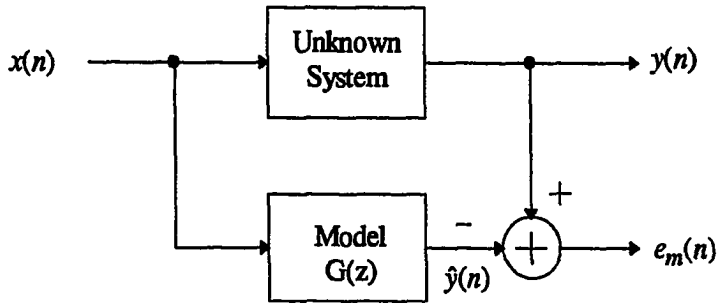


Figure 3-2 Direct modeling problem.

$$e_q(n)=\sum_{i=0}^{p} a_i y(n-i)-\sum_{j=0}^{q} b_j x(n-j). \tag{7}$$

The reader will recognize this error to be the difference between the right and left sides of (1). It is for this reason that it is referred to as the equation error. In the indirect modeling problem, model parameters are selected to minimize the equation error energy. This is often referred to as the *least-squares* error criterion. This situation is depicted in Figure 3-3.

Using z transform notation the equation error can be expressed as

$$E_q(z)=A(z)Y(z)-B(z)X(z) \tag{8}$$

To study the relationship between equation error and modeling error, note that (6) can be rewritten as

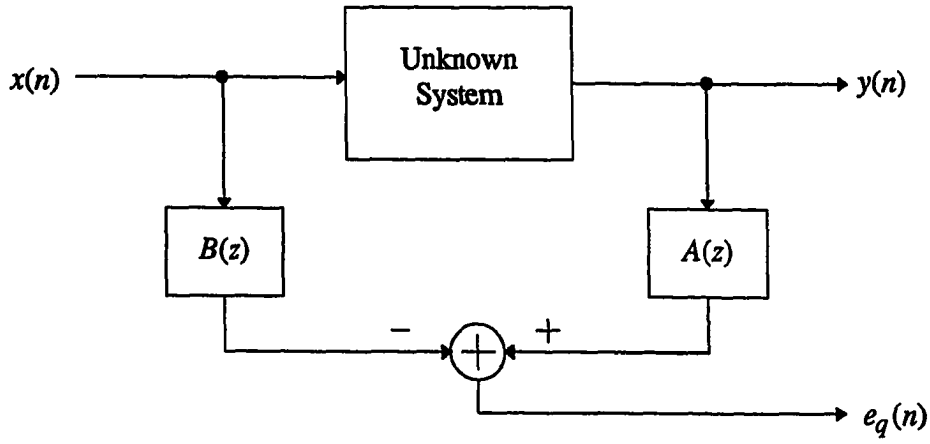


Figure 3-3 Indirect modeling problem.

$$A(z)E_m(z)=A(z)Y(z)-B(z)X(z) \tag{9}$$

A simple comparison between (8) and (9) reveals the relationship between the equation error and the modeling error:

$$E_q(z)=A(z)E_m(z) \tag{10}$$

Thus, minimization of the equation error energy (indirect method) is equivalent to minimization of the energy of a weighted modeling error (direct method). It is important to note the fact that none of the indirect modeling approaches presented will solve the direct modeling problem. The indirect techniques are commonly used because of the simplicity and computational efficiency of the solution process. As will be shown in the next several subsections, a number of related techniques can be used to solve the indirect modeling problem. In addition, a technique referred to as iterative prefiltering will be presented which attempts to estimate the direct modeling solution through repetitive application of indirect techniques.

3.2.2.1 Least squares (LS) method

The method of least squares can be used to determine the set of model parameters which results in minimum equation error energy [17,19-21]. Given a finite-length set of excitation and response data, (7) can be expressed in matrix form as

$$\begin{bmatrix} y(0) & y(-1) & \cdots & y(-p) & x(0) & x(-1) & \cdots & x(-q) \\ y(1) & y(0) & \cdots & y(1-p) & x(1) & x(0) & \cdots & x(1-q) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-p-1) & x(N-1) & x(N-2) & \cdots & x(N-q-1) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \\ -b_0 \\ -b_1 \\ \vdots \\ -b_q \end{bmatrix} = \begin{bmatrix} e_q(0) \\ e_q(1) \\ \vdots \\ e_q(N-1) \end{bmatrix} \quad (11)$$

The solution of the matrix equation requires $N+p$ output samples and $N+q$ input samples. Since $a_0=1$, this can be rearranged as

$$\begin{bmatrix} y(-1) & y(-2) & \cdots & y(-p) & x(0) & x(-1) & \cdots & x(-q) \\ y(0) & y(-1) & \cdots & y(1-p) & x(1) & x(0) & \cdots & x(1-q) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-2) & y(N-3) & \cdots & y(N-p-1) & x(N-1) & x(N-2) & \cdots & x(N-q-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \\ -b_0 \\ -b_1 \\ \vdots \\ -b_q \end{bmatrix} \quad (12)$$

$$= \begin{bmatrix} -y(0) \\ -y(1) \\ \vdots \\ -y(N-1) \end{bmatrix} + \begin{bmatrix} e_q(0) \\ e_q(1) \\ \vdots \\ e_q(N-1) \end{bmatrix}$$

or simply as

$$\boldsymbol{D}\,\phi = -\boldsymbol{y} + \boldsymbol{e} \quad (13)$$

where $\boldsymbol{D}$ is a $N\times(p+q+1)$ data matrix, ϕ is a $(p+q+1)\times1$ parameter vector, $\boldsymbol{e}$ is a $N\times1$ error vector, and $\boldsymbol{y}$ is a $N\times1$ output vector. For $\boldsymbol{e}=\boldsymbol{0}$, (13) will generally represent an overdetermined system of equations which will not have an exact solution. In the least-

squares method, one wishes to find the solution for the parameter vector which minimizes the error energy $e^{T}e$.

This solution can be determined by taking partial derivatives of $e^{T}e$ with respect to each parameter and setting the result equal to zero. A simpler and more common approach, however, would be to apply the *orthogonality principle*, which states that the parameter vector which minimizes the least-squares error will be the one which produces an error e that is orthogonal to the subspace spanned by the columns of D., i.e., $D^{T}e=0$ [17,20,21].

To apply the orthogonality principle, each side of (13) is first multiplied from the left by D^{T}, resulting in

$$D^{T}D\phi=-D^{T}y+D^{T}e. \tag{14}$$

The parameter vector which minimizes the least-squares error will also produce $D^{T}e=0$ as dictated by the orthogonality principle. Thus, from (14), minimization of $e^{T}e$ with respect to the elements of ϕ leads to the following set of equation commonly referred to as the *normal equations* [17]:

$$D^{T}D\phi=-D^{T}y. \tag{15}$$

Solution of the normal equations for ϕ results in the least-squares error solution. The matrix $D^{T}D$ will always be invertible if the columns of D are linearly independent. In this case, the resulting solution is

$$\phi = -\boldsymbol{R}^{-1}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{y} \tag{16}$$

where $\boldsymbol{R}=\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}$ and the minimum error energy is given by [17]

$$\min \boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\phi \tag{17}$$

A number of techniques exist for the computationally efficient solution of (15). However, the focus of this work will not be on the computational aspects of least-squares solutions and, hence, these techniques will not be presented here. The interested reader is encouraged to look at [17,20] for more information.

3.2.2.2 Extended Prony's method

If one is only concerned with matching the impulse response of the model to the impulse response of the unknown system then a technique referred to as the *extended Prony's method* can be used [17,22].

Consider the impulse response $\hat{h}(n)$ of a rational system of order (p,q) given by

$$\hat{h}(n) = -\sum_{i=1}^{p} a_i \hat{h}(n-i) + \sum_{j=0}^{q} b_j \delta(n-j) \tag{18}$$

where $\delta(n)$ is the discrete-time impulse. One should note that only the first $(q+1)$ samples, from $n=0$ to $n=q$, of the impulse response are directly affected by the numerator coefficients. The remaining samples, beginning with $n=q$ and extending to infinity, are a consequence of the denominator coefficients alone.

In the extended Prony's method, the denominator coefficients are determined by minimizing the least-squares equation error over the available impulse response data beyond the first $(q+1)$ samples. The numerator coefficients can then be selected such that the equation error is zero over the first $(q+1)$ samples as shown below.

The extended Prony's method proceeds in much the same manner as the least squares error method with the following exception. Instead of assuming an arbitrary input $x(n)$, a unit sample input $\delta(n)$ is assumed. The corresponding output is taken to be the impulse response of the unknown system $h(n)$. With this in mind (11) can be expressed as

$$\begin{bmatrix} h(0) & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ h(1) & h(0) & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & \cdots & h(N-p-1) & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \\ -b_0 \\ -b_1 \\ \vdots \\ -b_q \end{bmatrix} = \begin{bmatrix} e_q(0) \\ e_q(1) \\ \vdots \\ e_q(N-1) \end{bmatrix} \tag{19}$$

which can then be rewritten as

$$\begin{bmatrix} h(0) & 0 & \cdots & 0 \\ h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & \cdots & h(N-p-1) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} e_q(0) \\ e_q(1) \\ \vdots \\ e_q(q) \\ e_q(q+1) \\ e_q(q+2) \\ \vdots \\ e_q(N-1) \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ 0 \\ b_q \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{20}.$$

This set of equations can be partitioned into two sets, one containing both numerator and denominator coefficients and the other containing only denominator coefficients. Taking the upper $(q+1)$ rows results in

$$\begin{bmatrix} h(0) & 0 & \cdots & 0 \\ h(1) & h(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(q) & h(q-1) & \cdots & h(q-p) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_q \end{bmatrix} + \begin{bmatrix} e_q(0) \\ e_q(1) \\ \vdots \\ e_q(q) \end{bmatrix} \tag{21}$$

and taking the remaining lower rows results in

$$\begin{bmatrix} h(q+1) & h(q) & \cdots & h(1+q-p) \\ h(q+2) & h(q+1) & \cdots & h(2+q-p) \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & \cdots & h(N-p-1) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} e_q(q+1) \\ e_q(q+2) \\ \vdots \\ e_q(N-1) \end{bmatrix} \tag{22}$$

Rewriting (21) in simpler form leads to

$$\boldsymbol{H}_U \begin{bmatrix} 1 \\ \boldsymbol{a} \end{bmatrix} = \boldsymbol{b} + \boldsymbol{e} \tag{23}$$

One will note that if the denominator parameter vector $\boldsymbol{a}$ is known, then the numerator parameter vector $\boldsymbol{b}$ can be solved for exactly ($\boldsymbol{e}=\boldsymbol{0}$).

The reader will also note that the form of (22) is identical to that of (11), except for the fact that denominator coefficients are the only elements contained in the parameter vector. Thus, the least-squares solution for the denominator coefficients can be found in much the same manner of the least-squares method. One can rewrite (22) as

$$\begin{bmatrix} h(q) & h(q-1) & \cdots & h(1+q-p) \\ h(q+1) & h(q) & \cdots & h(2+q-p) \\ \vdots & \vdots & \ddots & \vdots \\ h(N-2) & h(N-3) & \cdots & h(N-p-1) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = -\begin{bmatrix} h(q+1) \\ h(q+2) \\ \vdots \\ h(N-1) \end{bmatrix} + \begin{bmatrix} e_q(q+1) \\ e_q(q+2) \\ \vdots \\ e_q(N-1) \end{bmatrix} \tag{24}$$

or simply

$$\boldsymbol{H_L a = -h + e_L} \tag{25}$$

The least-squares solution is then given as the solution to the following set of normal equations

$$(\boldsymbol{H_L}^{\mathrm{T}}\boldsymbol{H_L})\boldsymbol{a} = -\boldsymbol{H_L^{\mathrm{T}} h} \tag{26}$$

Once $\boldsymbol{a}$ has been determined by solving (26), the numerator parameter vector $\boldsymbol{b}$ is found from (23) using $\boldsymbol{e=0}$.

3.2.2.3 Iterative prefiltering method

Neither of the two previously mentioned methods attempt to solve the direct modeling problem. For this reason, the iterative prefiltering method will be presented [17,23,24]. The iterative prefiltering method is a technique which attempts to approximate the solution to the direct modeling problem through iterative application of indirect methods.

To understand the nature of the iterative prefiltering method, it is important to understand the relationship between the equation error of (8) and the modeling error of (6). In (10), it was shown that the modeling error is equivalent to a weighted equation error. Thus, from (10), the modeling error can be expressed as

$$E_m(z)=\frac{E_q(z)}{A(z)} \tag{27}$$

To apply the iterative prefiltering method, an initial estimate of the model parameters must be obtained. This can be achieved using either the least squares method or the extended Prony's method as previously discussed. Let the numerator and denominator polynomials associated with the model parameter estimates be denoted as $\hat{A}_1(z)$ and $\hat{B}_1(z)$.

Note that if the expression for equation error in (8) is filtered by the filter $1/\hat{A}_1(z)$ the following relationship results:

$$\frac{E_q(z)}{\hat{A}_1(z)} = \frac{A(z)Y(z)}{\hat{A}_1(z)} - \frac{B(z)X(z)}{\hat{A}_1(z)} \tag{28}$$

With the assumption that

$$\frac{E_q(z)}{\hat{A}_1(z)} \cong \frac{E_q(z)}{A(z)} \tag{29}$$

combining (27) and (28) leads to

$$E_m(z) \cong \frac{A(z)Y(z)}{\hat{A}_1(z)} - \frac{B(z)X(z)}{\hat{A}_1(z)} \tag{30}$$

This can be rewritten as

$$E_m(z) \cong A(z)\hat{Y}(z) - B(z)\hat{X}(z) \tag{31}$$

where

$$\hat{Y}(z)=\frac{Y(z)}{\hat{A}_1(z)} \tag{32}$$

is the original output prefiltered by $1/\hat{A}_1(z)$ and

$$\hat{X}(z)=\frac{X(z)}{\hat{A}_1(z)} \tag{33}$$

is the original input prefiltered by $1/\hat{A}_1(z)$.

Note that (31) now has the form of (8), i.e., (31) is an approximate expression for modeling error written in the form of an equation error. The application of the least squares method to the prefiltered input and output data results in a parameter estimate which minimizes an approximate modeling error. This situation is depicted in Figure 3-4. Let the resulting least squares parameter estimates be denoted $\hat{A}_2(z)$ and $\hat{B}_2(z)$.

As before, the new parameter estimate $\hat{A}_2(z)$ can be used to prefilter the original input and output, and the least squares method can be used to obtain a new parameter estimate. This process can be repeated to obtain progressively better parameter estimates in terms of satisfying the minimum modeling error energy criterion. Thus, iterative application of prefiltering and least squares parameter estimation can be used to obtain parameter estimates which minimize an approximate modeling error.

The iterative prefiltering technique can be summarized as follows:

1. Obtain an initial parameter estimate using the least squares method or Prony's method.
2. Prefilter the original input and output sequences using the denominator coefficients estimate using (32) and (33).
3. Compute a new parameter estimate using least-squares method on the prefiltered input and output sequences.
4. Repeat steps 2 through 4 as needed.

In comparison to other indirect methods, the iterative prefiltering method has two major drawbacks. The required computational effort is higher because of multiple application of indirect techniques. The second disadvantage is that the iterative process is not guaranteed to

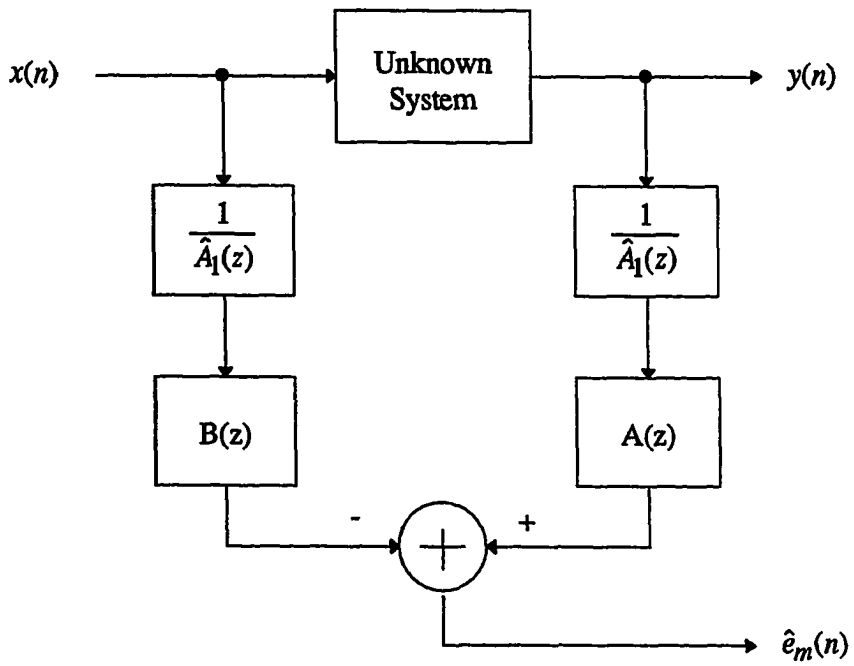


Figure 3-4 First iteration of the iterative prefiltering method.

converge. In practice, however, satisfactory results can be obtained by terminating the iterative process just before the solution begins to diverge, as is sometimes the case. The obvious advantage of the iterative prefiltering method is its ability to estimate the solution to the direct modeling problem. Since the iterative prefiltering method uses the least-squares method as the initial estimate, the performance of the iterative prefiltering method, in terms of minimum modeling error energy, will always be equal to or greater than the performance obtained using the least-squares method.

3.2.3 Balanced Model Reduction

In this subsection, a somewhat different approach to system modeling will be presented based on balanced model reduction [25-31]. In this technique, a state-space realization of the system is formed from the observed system impulse response. The state-space representation is then transformed into what is referred to as a balanced realization via an appropriate state transformation. A system reduction can then be performed by simple truncation of states. A brief overview of balance model reduction will be presented here. For more details on balanced model reduction, including the definition of a balanced state-space and the computation of an appropriate state transformation matrix, the reader is referred to Appendix B.

The state-space model commonly used in control systems theory is the basis for many model reduction techniques [27,32]. The following state-space model can be used to model a linear, finite-dimensional system:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k)+\boldsymbol{B}u(k) \\ y(k) &= \boldsymbol{C}\boldsymbol{x}(k)+Du(k) \end{aligned} \tag{34}$$

where $\boldsymbol{x}(k)$ is the r-dimensional state vector, $\boldsymbol{A}$ is an $r\times r$ matrix, $\boldsymbol{B}$ and $\boldsymbol{C}$ are $r\times 1$ and $1\times r$ vectors respectively, and D is a scalar. These equations are commonly termed the *state equations* of the system, where r is the *system order*. The state equations dictate a mapping from the system input $u(k)$ and the present state $\boldsymbol{x}(k)$ to the output $y(k)$ and the next state $\boldsymbol{x}(k+1)$. The corresponding transfer function is given by [27,32]

$$H(z)=\frac{Y(z)}{U(z)}=\boldsymbol{C}(z\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B}+D \tag{35}$$

and the corresponding impulse response is

$$\begin{aligned} \boldsymbol{h} &= [h_0 \; h_1 \; h_2 \; h_3 \cdots]^{\mathrm{T}} \\ &= [D \; \boldsymbol{CB} \; \boldsymbol{CAB} \; \boldsymbol{CA}^2\boldsymbol{B} \cdots] \end{aligned} \tag{36}$$

For a given transfer function, an infinite number of state-space representations can be realized. Given a high-order, finite-length impulse response, obtaining a state-space realization is a simple procedure. The realization described by

$$\boldsymbol{A}=\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{37}$$

$$B=\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} \tag{38}$$

$$C=[h_1\ h_2\ h_3\ \cdots\ h_n] \tag{39}$$

$$D=h_0 \tag{40}$$

is one such realization [27], which can be easily verified using (36).

Other state-spaces having the same transfer function can be obtained via a state transformation. Given an arbitrary, nonsingular matrix T, the state-space (A, B, C ,D) specified above can be transformed into a new state-space (A', B' ,C' ,D) having the same transfer function and given by [32]

$$A'=T^{-1}AT \tag{41}$$

$$B'=T^{-1}B \tag{42}$$

$$C'=CT \tag{43}$$

In model reduction, one wishes to reduce the order of the system without significantly altering the input-output relationship of the system. For balanced model reduction, one is primarily interested in obtaining systems which are balanced. A balanced system is useful because it allows one to easily reduce the system order by simple truncation of states.

As described in Appendix B, a given state-space realization can be transformed into a balanced realization using a transformation matrix T obtainable through singular value decomposition of the Hankel matrix. The Hankel matrix $\Phi(H)$ is defined as [32]

$$\Phi(H)=\begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_2 & h_3 & h_4 & \cdots & 0 \\ h_3 & h_4 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ h_n & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{44}$$

Performing a singular value decomposition of the Hankel matrix results in

$$\begin{aligned} \Phi(H) &= U\Sigma V^{\mathrm{T}} \\ &= (U\Sigma^{1/2})(\Sigma^{1/2}V^{\mathrm{T}}) \end{aligned} \tag{45}$$

where U and V are unitary matrices and Σ is a diagonal matrix containing the so-called *Hankel singular values* σ_i of $\Phi(H)$. The singular values are nonnegative numbers and typically ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} \geq \cdots \geq \sigma_n$. The desired transformation matrix T which transforms a arbitrary state-space into a balanced state-space is given by [27]

$$T = S_o^{-1}\Sigma^{1/2}U^{\mathrm{T}} \tag{46}$$

where S_o is the observability matrix

$$S_o = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{r-1} \end{bmatrix}. \tag{47}$$

The Hankel matrix and, hence, the Hankel singular values are invariant under state-space transformation. Each singular value represents a ranking of the contribution of a dynamical element of the balanced state-space. The summation of the squared singular values will be equal to the total energy (sum of the squared elements) of the Hankel matrix.

Model reduction can be achieved by truncating the states associated with the smallest singular values. Thus, given a balanced state-space (A', B', C', D), a reduced-order state-space (A'_1, B'_1, C'_1, D) can be obtained by truncation of states such that

$$A' = \begin{bmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{bmatrix} \tag{48}$$

$$B' = \begin{bmatrix} B'_1 \\ B'_2 \end{bmatrix} \tag{49}$$

$$C' = [C'_1 \; C'_2] \tag{50}$$

If the system (A', B', C', D) is asymptotically stable and balanced, the truncated system (A'_1, B'_1, C'_1, D) will be asymptotically stable [27,29].

The balanced model reduction method can now be summarized as follows:

1. Obtain a high-order impulse response from the unknown system.
2. Form a state-space realization (A, B, C, D) from the impulse response of the unknown system using (37) through (40).

3. Transform the state-space realization (A, B, C ,D) into a balanced realization (A' ,B' ,C' ,D) according to (41) through (43) using the transformation matrix T given by (46).
4. Form a reduced-order model (A'_{11}, B'_1, C'_1, D) using (48) through (50) based on Hankel singular values.

The computational requirements for the balanced model reduction technique are intensive. In addition, the calculations are often ill-conditioned. For this reason, a number of methods have been proposed for simplifying the design procedure. The reader is encouraged to look at [28-30] for additional information. More will be said about balanced model reduction techniques in the next chapter. The focus of the system modeling discussion will now shift to model order estimation.

3.3 Order Estimation

Determination of the system order has been the focus of considerable research. A number of model order estimation techniques exist in the literature including the Akaike information criteria (AIC) [33], the final prediction error (FPE) [34], and the minimum description length (MDL) [33]. Since each of these is based on the estimated error variance, the model parameters must be estimated for each model order evaluated, and the corresponding equation error must be computed. These techniques require considerable computational effort, and the resulting model order estimate is dependent on the parameter estimation technique that is used.

In this work, a relatively new order estimation approach will be examined which is based on the eigenvalue decomposition of a covariance matrix formed from the observed response data [35,36]. Details of this technique are briefly presented below.

For the noiseless case, assume that the observed excitation and response data satisfies (1) exactly for some undetermined orders p and q. Note that for finite-length data records, (1) can be rewritten in matrix form as

$$\begin{bmatrix} y(0) & y(-1) & \cdots & y(-p) \\ y(1) & y(0) & \cdots & y(1-p) \\ \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-p-1) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} x(0) & x(-1) & \cdots & x(-q) \\ x(1) & x(0) & \cdots & x(1-q) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-1) & x(N-2) & \cdots & x(N-q-1) \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_q \end{bmatrix} \tag{51}$$

or simply as

$$\boldsymbol{Y}_p \boldsymbol{a}_p = \boldsymbol{X}_q \boldsymbol{b}_q \tag{52}$$

in which $\boldsymbol{Y}_p$ is a $N\times(p+1)$ response matrix, $\boldsymbol{X}_q$ is a $N\times(q+1)$ excitation matrix, $\boldsymbol{a}_p$ and $\boldsymbol{b}_q$ are respectively $(p+1)\times 1$ and $(q+1)\times 1$ column parameter vectors. Rearranging (52) yields

$$\left[\boldsymbol{Y}_p \vdots -\boldsymbol{X}_q\right] \begin{bmatrix} \boldsymbol{a}_p \\ \cdots \\ \boldsymbol{b}_q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{53}$$

Let the data matrix $\boldsymbol{D}_{pq}$ be defined as

$$\boldsymbol{D}_{pq}=\left[\boldsymbol{Y}_p \vdots -\boldsymbol{X}_q\right] \tag{54}$$

and the corresponding covariance matrix $\boldsymbol{R}_{pq}$ to be

$$\boldsymbol{R}_{pq} = \boldsymbol{D}_{pq}^T \boldsymbol{D}_{pq} \tag{55}$$

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

$$\boldsymbol{R}_{pq} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \tag{56}$$

in which the $(p+q+2)\times(p+q+2)$ eigenvector matrix $\boldsymbol{Q}$ has as its columns the orthonormal set of eigenvectors of $\boldsymbol{R}_{pq}$ and the eigenvalue matrix

$$\Lambda = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{p+q+1} \end{bmatrix} \tag{57}$$

has as its elements the corresponding $(p+q+2)$ eigenvalues. Typically, the eigenvalues are ordered such that $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{p+q+1}$. Because $\boldsymbol{R}_{pq}$ is positive semidefinite, all of its eigenvalues will be nonnegative.

It can be shown [35,36] that if the "true" order of the unknown system is (n_p,n_q) and if the system order (p,q) in (52) is selected such that $p \geq n_p$ and $q \geq n_q$, then the covariance matrix $\boldsymbol{R}_{pq}$ will have at least one zero eigenvalue, since there will exist at least one exact solution for $\boldsymbol{a}_p$ and $\boldsymbol{b}_q$ in (53).

Based on this result, the model order estimation procedure can be formulated.

1. For all model orders (p,q) of interest, construct the data matrix and compute the corresponding covariance matrix defined in (54) and (55) using available excitation-response data.
2. Compute the minimum eigenvalue of each covariance matrix.
3. Take the model order estimate to be the lowest order (p,q) at which the minimum eigenvalue is zero.

Thus, in the noiseless case, the true order of the system can be selected as the lowest order (p,q) at which the minimum eigenvalue falls to zero. In the noise-contaminated case, however, the estimation procedure becomes more difficult because the eigenvalues of the covariance matrix will be incremented and will no longer equal zero when the true model order has been reached or exceeded.

One can consider the case in which the excitation $x(n)$ and response $y(n)$ are both contaminated with a zero mean, white noise with variance σ^2. It is shown in Appendix A that for large N the corresponding eigenvalues will be incremented by an amount equal to $\sigma^2 N$. Thus, the new eigenvalue matrix $\hat{\Lambda}$ corresponding to a noise-contaminated covariance matrix will be given by

$$\hat{\Lambda}=\Lambda+\sigma^2 NI \tag{58}$$

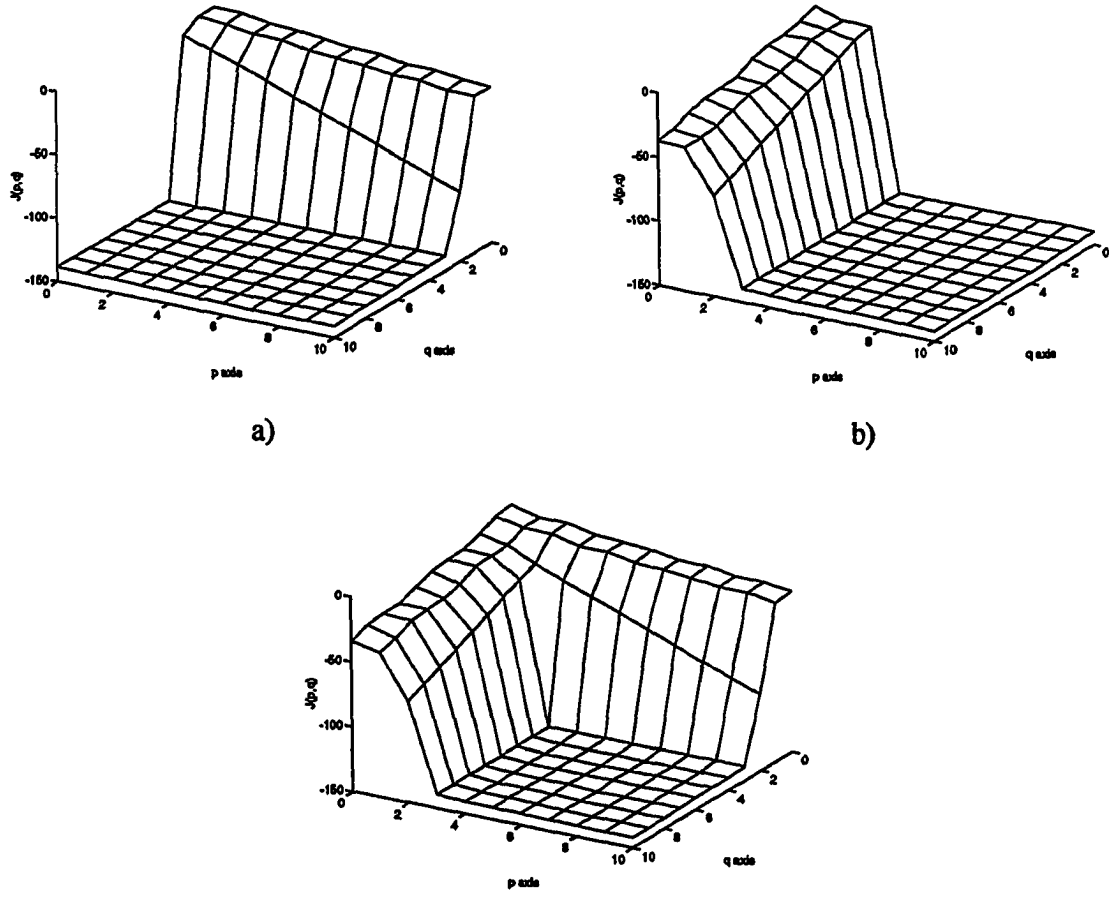
where Λ is the eigenvalue matrix associated with a noise-free covariance matrix as described in Appendix A, and I is the identity matrix. The minimum eigenvalue will no longer be zero but will equal $\sigma^2 N$.

Another useful case is that in which the excitation is assumed to be noise-free and the response is assumed to be contaminated with a zero-mean white noise with variance σ^2. Such a situation would be applicable when $x(n)$ is a discrete-time impulse and $y(n)$ is the observed impulse response data. It is shown in Appendix A that for this case the increment in eigenvalues is bounded by $\sigma^2 N$. The new eigenvalue matrix is upper-bounded such that

$$\hat{\Lambda} \leq \Lambda + \sigma^2 N I \tag{59}$$

This result is useful because it relates the modeling error variance to the value of the minimum eigenvalue.

To demonstrate the minimum eigenvalue estimation technique and the effect of additive output noise, consider the following three model examples. Figure 3-5 shows minimum eigenvalue plots for MA(3), AR(3), and ARMA(3,3) systems. For each system, a 200 coefficient impulse response was obtained to which a very low variance (10^{-16}) white noise was added. The covariance matrix was then formed for all model orders (p,q) such that $0 \leq p \leq 10$ and $0 \leq q \leq 10$. At each order the minimum eigenvalue of the covariance matrix was plotted.



a) b)

c)

Figure 3-5 Log-scale minimum eigenvalue plots for a) an MA(3) system b) an AR(3) system and c) an ARMA(3,3) system.

To accommodate the dynamic range of the data, a logarithmic scale was adopted. Minimum eigenvalues were plotted in dB relative to unity, i.e., $J(p,q)=10\cdot\log_{10}(\lambda_{p+q+1})$ where $J(p,q)$ is the minimum eigenvalue in dB for the order (p,q) model. Since noise was added to the observed system response of all three systems the minimum eigenvalues did not

drop to zero once the true model order had reached, but rather to some small, nonzero "noise floor". The value of this noise floor predicted by (59) given a 200 coefficient impulse response and a noise variance of 10^{-16} is -137 dB, approximately the value indicated in the plots. A noise threshold set to a value just above this noise floor could therefore be used to select the appropriate model order. That is, the lowest order (p,q) which results in a minimum eigenvalue below this threshold could be selected as the model order estimate.

In the next chapter, this technique will be used to estimate appropriate model structures for HRTFs. Given an arbitrary HRTF, comparison of the corresponding minimum eigenvalue plot to the plots shown in Figure 3-5 will reveal the order and structure of an appropriate model.

4. MODELING OF THE HEAD-RELATED TRANSFER FUNCTIONS

The modeling techniques presented in the previous chapter can be applied to model a variety of linear, time-invariant systems. In this work, one will be primarily concerned with the application of these techniques to the modeling of head-related transfer functions. Because of the perceptual importance of the HRTFs, modifications to standard modeling techniques will be proposed for increasing the auditory performance of the resulting model.

A set of abbreviations has been established for making reference to the various modeling techniques presented in this chapter. These abbreviations are shown in Table 4-1. Application of each of these to HRTFs will be discussed in Section 4.3. In this document, the variable p has been and will be consistently used to represent the order of the denominator polynomial of a model transfer function. Likewise, the variable q will be used to represent the numerator polynomial order. Thus, the pair (p,q) will denote a pth order denominator and a qth order numerator. An order pair (p,q) will commonly be used in conjunction with an abbreviation to represent the model type and order. For example, a BMR(12,12) model denotes a model designed using balanced model reduction with a 12th order denominator and a 12th order numerator.

4.1 Minimum-Phase Approximation

A considerable amount of research has focused on the approximation of HRTFs as minimum phase systems [38]. In section 3.1, it was stated that a rational system was said to be minimum-phase if all system poles and zeros are inside the unit circle of the complex plane.

Table 4-1 Abbreviations for modeling techniques.

Abbreviation	Description
FIR	FIR filter designed by truncation of impulse response.
LS	Least-squares method.
LSWE	Least-squares weighted error method.
IP	Iterative prefiltering method.
WIP	Weighted iterative prefiltering method.
PZC	Pole-zero cancellation method.
BMR	Balanced model reduction method.

Given an arbitrary, stable, rational system $H(z)$ with corresponding impulse response $h(n)$ and frequency response $H(e^{j\omega})$, there will exist only one system which is minimum-phase and has a magnitude response $|H(e^{j\omega})|$, assuming that the system has no poles and zeros in common. Let this system be denoted $H_{\min}(z)$ and its corresponding impulse response be denoted $h_{\min}(n)$. Thus, specifying a desired magnitude response uniquely specifies a corresponding minimum-phase system.

Research indicates that the approximation of the head-related transfer functions as minimum-phase systems preserves localization performance [38]. This assumption represents a significant processing convenience because the phase spectrum can essentially be ignored.

An initial step in the modeling process, therefore, is to find a minimum phase system given an impulse response or, equivalently, given a desired magnitude response.

A minimum-phase system is said to have a *minimum energy delay* property, which states that the energy of the first n samples of the impulse response of a minimum phase system is greater than the energy of the first n samples of the impulse response of any other system with the same magnitude response. Mathematically, the minimum energy delay property states that [18]

$$\sum_{m=0}^{n}|h(m)|^2 \le \sum_{m=0}^{n}|h_{\min}(m)|^2 \tag{60}$$

where

$$\left|H(e^{j\omega})\right| = \left|H_{\min}(e^{j\omega})\right|. \tag{61}$$

To convert the impulse response $h(n)$ of an arbitrary system into the impulse response $h_{\min}(n)$ of a minimum phase system with the same magnitude response, a technique presented in [18] can be used. It can be shown [18] that the complex cepstrum $\hat{c}_h(n)$ of the impulse response of a minimum-phase system is causal, i.e., $\hat{c}_h(n)=0$ for $n<0$. For real sequences the cepstrum $c_h(n)$ can be defined as the even part of the complex cepstrum $\hat{c}_h(n)$

$$c_h(n)=\frac{\hat{c}_h(n)+\hat{c}_h(-n)}{2}. \tag{62}$$

The cepstrum $c_h(n)$ is actually the inverse Fourier transform of the log-magnitude spectrum of $h(n)$. If one computes the cepstrum of $h(n)$, (62) can be used to compute the complex cepstrum of $h_{\min}(n)$ assuming the complex cepstrum of $h_{\min}(n)$ to be causal. From (62), the causal, complex cepstrum of the minimum-phase system can be expressed as

$$\hat{c}_h(n)=\begin{cases} 0, & n<0 \\ c_h(n), & n=0 \\ 2\cdot c_h(n), & n>0 \end{cases} \tag{63}$$

or equivalently as

$$\hat{c}_h(n)=c_h(n)\cdot g(n) \tag{64}$$

where

$$g(n)=2u(n)-\delta(n)\,. \tag{65}$$

where $u(n)$ is the unit step response and $\delta(n)$ is the discrete-time impulse.

The complex cepstrum $\hat{c}_h(n)$ of $h_{\min}(n)$ in terms of $h_{\min}(n)$ is given by.

$$\hat{c}_h(n)=\Im^{-1}\{\log[\Im(h_{\min}(n))]\} \tag{66}$$

where $\Im$ represents the discrete Fourier transform and $\Im^{-1}$ represent the inverse discrete Fourier transform. Thus, $h_{\min}(n)$ can be obtained from $\hat{c}_h(n)$ as

$$h_{\min}(n)=\Im^{-1}\{\exp[\Im(\hat{c}_h(n))]\} \tag{67}$$

A block diagram of the minimum-phase conversion procedure is shown in Figure 4-1. Given an impulse response, the real cepstrum is computed. The complex cepstrum of $h_{\min}(n)$ is then computed using (64). Finally, the sequence $h_{\min}(n)$ is found from $\hat{c}_h(n)$ using (67).

A system $H(z)$ can be factored into a minimum-phase system $H_{\min}(z)$ with magnitude response $|H(e^{j\omega})|$ and an all-pass system $H_{ap}(z)$ with unity magnitude response such that

$$H(e^{j\omega}) = H_{\min}(e^{j\omega}) H_{ap}(e^{j\omega}) \tag{68}$$

where $H_{\min}(e^{j\omega})$ is the frequency response of the minimum-phase system and $H_{ap}(e^{j\omega})$ is the frequency response of the all-pass system.

A minimum-phase system has a lower order than a mixed-phase system with the same

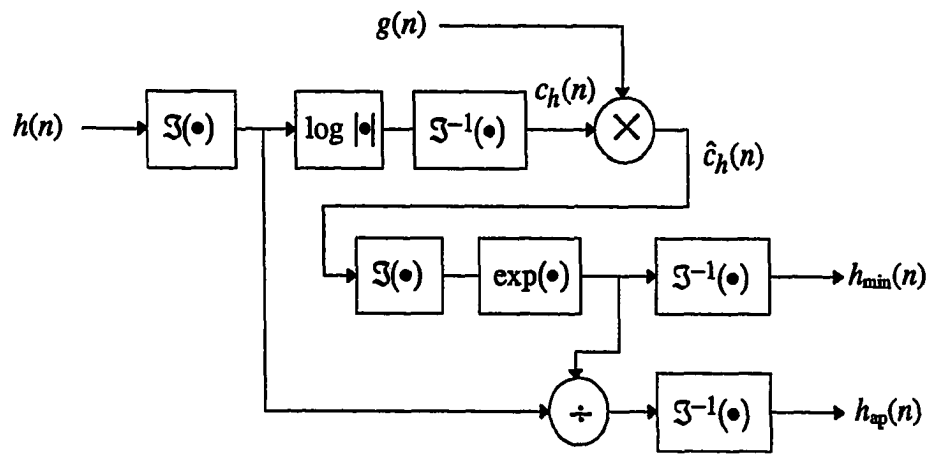


Figure 4-1 Minimum-phase computation block.

magnitude response since the all-pass system is not required. Therefore, a minimum-phase system can generally be modeled more efficiently than a corresponding mixed-phase system. For this reason, and for the previously mentioned processing convenience, HRTFs will be modeled exclusively as minimum phase functions for the remainder of this work. The fundamental assumption, of course, is that the minimum-phase approximation will not affect the localization performance of the filters, as previously stated.

4.2 Model Order Estimation

4.2.1 Application of minimum eigenvalue order estimation to HRTFs

The minimum eigenvalue technique discussed in Chapter 3 and in Appendix A can be applied to the head-related transfer functions to identify appropriate model structures. 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 HRTF impulse response. HRTF impulse responses measured using the technique described in [2] are first converted to minimum-phase sequences before applying the order estimation procedure.

Figure 4-2 shows the minimum eigenvalue plots for several left ear HRTFs of one subject for positions on the horizontal plane. By comparing the theoretical plots of Figure 3-5 to the HRTF plots of Figure 4-2, it appears as though the HRTFs are primarily all-pole (autoregressive) systems. Indicative of an all-pole system is the fact that the minimum eigenvalues drop sharply in the p axis direction and slowly in the q axis direction. This

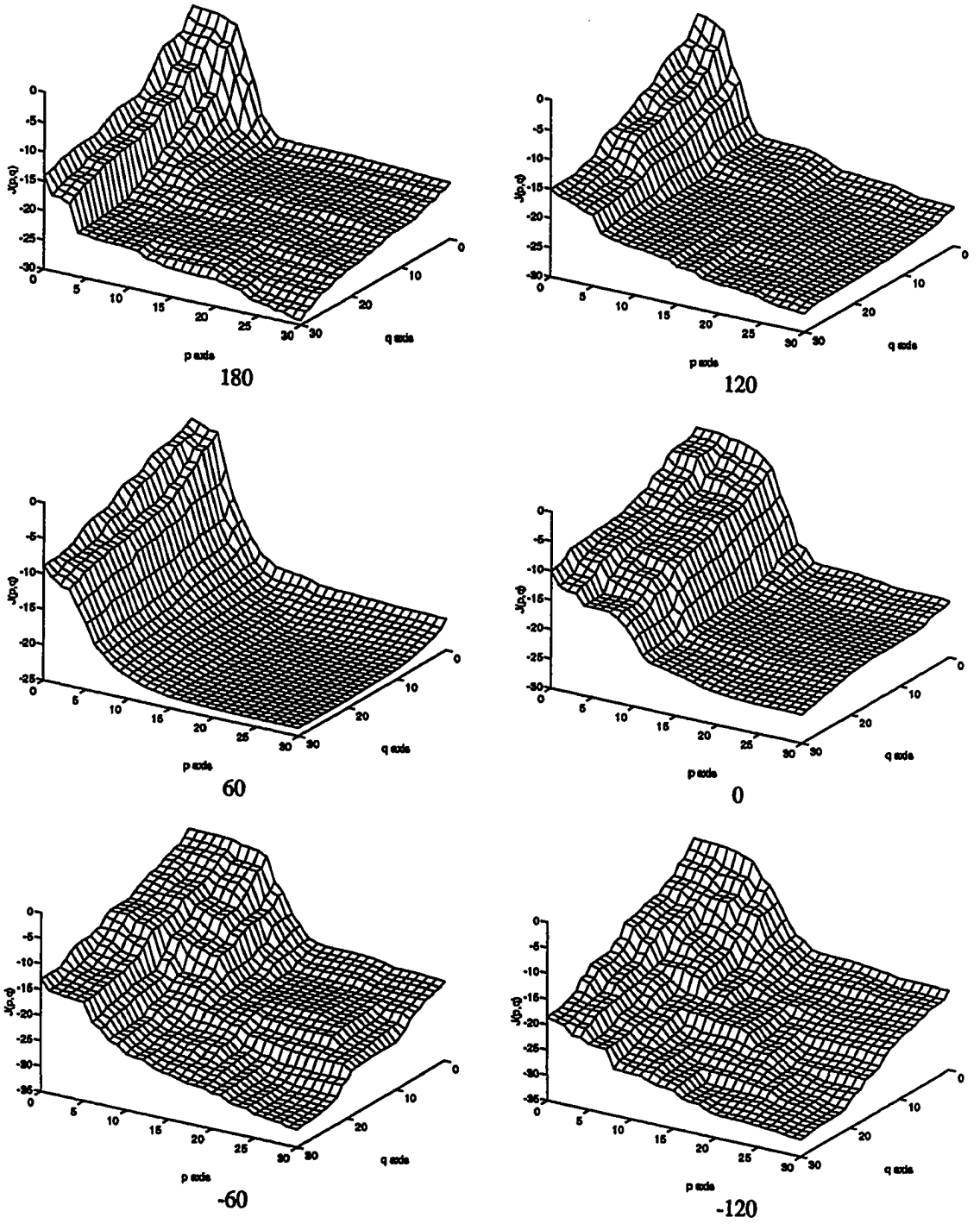


Figure 4-2 Minimum eigenvalue plots for left ear, horizontal plane HRTFs of one subject. Azimuth angle is shown.

effect is observed for the HRTFs shown in Figure 4-2, and for HRTFs in general.

To estimate the order of each transfer function, a noise threshold can be selected. The model order at which the minimum eigenvalue drops below this threshold is selected as the order estimate. Because the observed impulse response is not a perfect impulse response from a rational system, the eigenvalues will not clearly fall to zero as in the noise-free case. Selection of an appropriate threshold for head-related transfer functions, therefore, can only be accurately selected through the use of subjective listening tests. For the time being, a comparison of all-pole and all-zero model orders required to reach a fixed threshold will be made.

Based on a -15 dB threshold, AR model order estimates were computed for 450 minimum-phase, left ear HRTFs of one subject for positions distributed evenly about a spherical shell. Figure 4-3a shows a histogram of the AR model order estimates for all 450 transfer functions. The mean order is 9.8. The highest order estimate is 14.

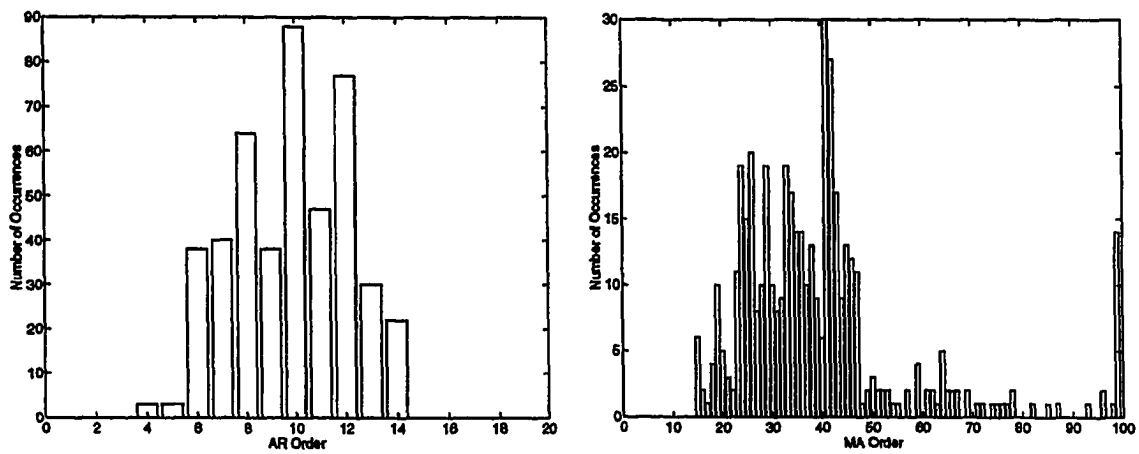
For comparison, MA model order estimates for the -15 dB threshold were also computed for the same 450 HRTFs. The results are illustrated by the histogram of Figure 4-3b. It is clear that, at a -15 dB threshold, the MA model order estimates are significantly higher than the AR model estimates.

At lower thresholds, however, the difference between AR and MA model order estimates is reduced due to the fact that the MA model has a gradual decrease in minimum eigenvalues (along the q axis) while the AR model shows little increase above roughly 15th order. 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. As shown in Appendix A, lowering the eigenvalue threshold corresponds roughly to lowering the allowable modeling error variance, an objective criterion which does not necessarily reflect the perceptual performance of the model. Consequently, perceptual listening test are needed to fully asses the significance of these results. It does appear, however, that the use of an autoregressive filter would be required to efficiently model the HRTFs.

4.2.2 The diffuse field and directional transfer functions (DTFs)

The HRTFs seem to be highly autoregressive for most, if not all, transfer function



a) b)

Figure 4-3 a) AR model order estimates b) MA model order estimates.

positions. There is research [38] which suggests that the head-related transfer functions can be modeled as a single, position-independent all-pole filter in parallel with a position-dependent all-zero filter with promising results. With this in mind, it would seem advantageous to find a factorization of the HRTFs which would divide each one into a position-dependent function and a position independent function.

One can define the *diffuse field response* to be the RMS magnitude spectrum of the HRTFs of all source positions. Thus, the diffuse field computed over N source positions can be expressed as

$$DF(\omega)=\left(\frac{1}{N}\sum_{i=0}^{N-1}\left|H_i(e^{j\omega})\right|^2\right)^{1/2} \tag{69}$$

where $DF(\omega)$ is the diffuse field response and $H_i(e^{j\omega})$ is the frequency response of the head-related transfer function associated with the ith source position.

The diffuse field response can be viewed as the overall response of the ear to a wideband, diffuse sound field, i.e., wideband sound of equal intensity from all directions. The diffuse field represents a non-directional component of the HRTFs. The diffuse field is typically computed separately for each ear. Dividing the magnitude response of each left ear HRTF by the diffuse field response of the left ear (in the frequency domain) extracts the part of the left ear which is dependent on direction. A similar procedure can be done for the right ear. After extracting the diffuse field from an HRTF, by division of magnitude spectrums, the residual is generally referred to as the *directional transfer function* (DTF) since it represents the direction-dependent component of the HRTF.

When HRTF-filtered stimuli are presented over conventional headphones, the headphone-to-eardrum transfer function contributes an additional frequency response to the received signal. This headphone-to-eardrum transfer function will generally have a similar frequency response characteristic to that of the diffuse-field response. In fact, some headphones claim to be "diffuse field equalized", indicating that the headphone-to-eardrum transfer function matches the user's response to a diffuse sound field. Therefore, it is approximately correct in an actual implementation to use the DTFs for three-dimensional sound synthesis instead of using the HRTFs.

Elimination of the additional computation associated with the diffuse field filtering represents perhaps the greatest benefit of using DTFs for sound synthesis. Although it would be ideal to use the HRTFs corrected by the actual headphone-to-eardrum transfer functions characteristic as measured in a laboratory, it will be assumed that the headphone-corrected head-related transfer functions essentially match the directional transfer functions. Because of this, the focus of the modeling process will now shift to examining the directional transfer functions.

4.2.3 Application of minimum eigenvalue order estimation to DTFs

To examine appropriate model structures for DTFs, the minimum eigenvalue technique was applied. The directional transfer functions were first computed using the procedure outlined above. The DTFs were then bandlimited by specifying a flat magnitude response (0 dB) above 15 kHz, since the original HRTF measurements were only valid up to that frequency. Using the minimum-phase DTFs of one subject for positions on the horizontal

plane, the minimum eigenvalues were computed for a wide range of model orders. The results are shown in Figure 4-4.

By comparing the plots in Figure 4-4 to the theoretical plots of Figure 3-5, it is apparent that the DTFs are not autoregressive systems but rather autoregressive moving-average systems. The minimum eigenvalues decay slowly in the p and q axes directions, but fall sharply along the diagonal. The DTFs appear to be high order systems since the "noise floor" is not reached until a high order. Because the drop in eigenvalues as the model order is increased is so considerable, it is assumed that a low-order model may be acceptable. Even for 10th order models, the drop in minimum eigenvalues exceeds 50 dB. The exact threshold that can be tolerated without sacrificing perceptual performance is difficult to determine without using localization tests.

The minimum eigenvalue plots also seem to indicate a rather consistent model structure for all positions. The symmetry of the plots and the value of the noise floor seem to be essentially the same for all positions shown. Because of the autoregressive moving-average appearance of the minimum eigenvalue plots, the parameter estimation techniques of the next section will focus on estimating the parameters of models with equal order numerator and denominator polynomials.

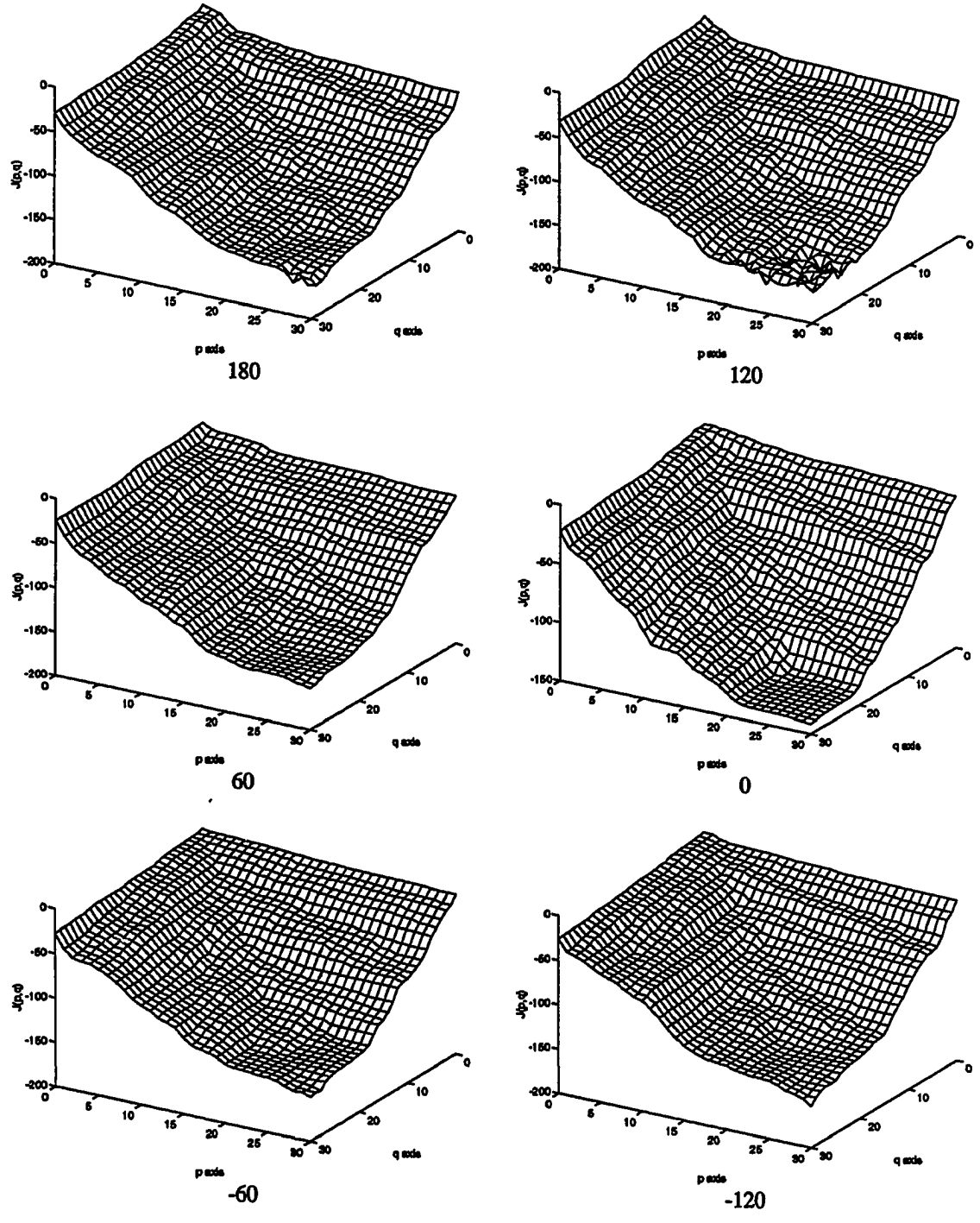


Figure 4-4 Minimum eigenvalue plots for left ear, horizontal plane DTFs of one subject. Azimuth angle is shown.

4.3 Parameter Estimation

4.3.1 Least-squares (LS) method

In Chapter 3, the least-squares method was presented for computing the parameters of a pole-zero model. In this subsection, the least squares method will be applied to the modeling of HRTFs. Figure 4-5 shows the magnitude response of a sample HRTF and the magnitude response of a corresponding model fit using the least-squares method. It is apparent that the model fits the magnitude spectrum well at high frequencies, but poorly at lower frequencies.

The question of concern is whether or not this difference can affect the localization performance of the model. Perhaps a more fundamental question might be whether or not the difference is perceptually transparent. To answer these questions, it becomes necessary to have some understanding of the hearing process, which is the topic of the next subsection.

4.3.2 The hearing process and psychoacoustics

Important to any perceptually valid modeling process is an understanding of the hearing process. The human ear is fundamentally a frequency analysis instrument. A sound incident upon the ear ultimately results in a disturbance along the cochlear partition, or cochlear duct [39]. In humans, the cochlea is a tiny, spiral cavity which resides in the inner ear. An "unrolled" diagram of the cochlea is shown in Figure 4-6. The cochlear partition is an organ which divides cochlea into an upper gallery (scala vestibuli) and a lower gallery (scala tympani). The cochlear partition runs the entire length of the cochlea.

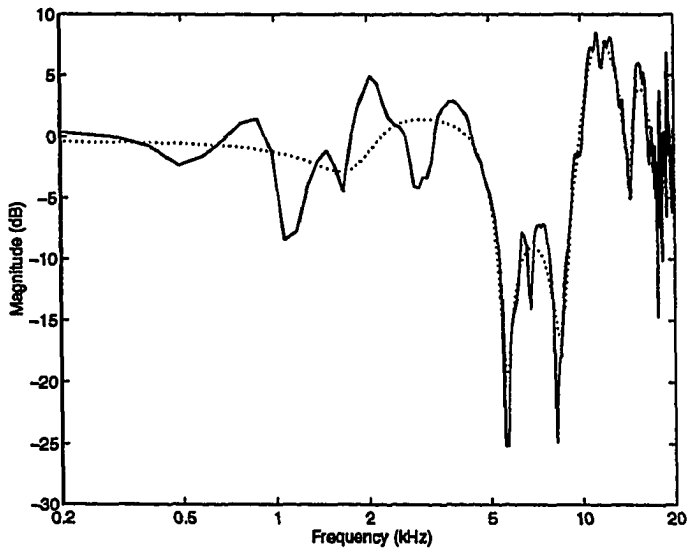


Figure 4-5 Comparison between sample DTF magnitude response (solid line) and magnitude response of a 12th order model fit using least-squares method.

The cochlear partition is bounded along the lower gallery by the basilar membrane. An incident sound produces a disturbance along the cochlear partition. The frequency of an incident sound determines the physical area of disturbance along the basilar membrane. In general, high frequency tones result in a maximum disturbance of the basilar membrane near the basal end, while low frequency tones result in a maximum disturbance of the basilar membrane near the apex. The detection of pitch is determined by one's ability to sense the area of maximum disturbance. Increasing the frequency of the incident sound widens the area of disturbance.

The term *masking* refers to the phenomenon by which one sound interferes with the perception of another sound [39]. Consider the situation in which a single tone is masked by

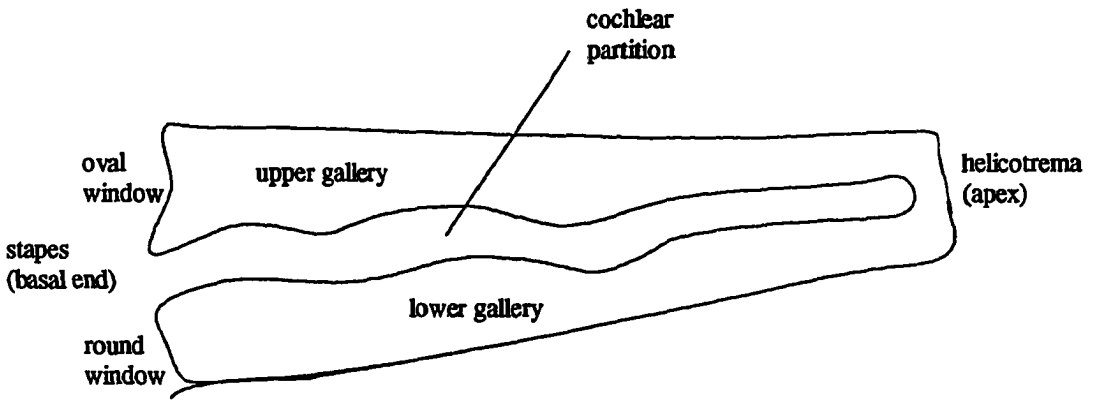


Figure 4-6 Simplified diagram of the unrolled cochlea.

another tone. Each tone corresponds to a physical excitation of a region along the basilar membrane. Figure 4-7 shows a plot of the amplitude of the disturbance along the basilar membrane due to two tones, one at 500 Hz and the other at 1000 Hz. The horizontal axis is labeled in frequency corresponding to distance along the cochlea partition. If the level of the 500 Hz tone is raised such that the response at 1000 Hz due to the 500 Hz tone is greater than the response at 1000 Hz due to the 1000 Hz tone itself, then the 1000 Hz tone will become inaudible. The 500 Hz tone is said to *mask* the 1000 Hz tone.

If the frequency of the tone at 500 Hz is increased, there exists a point at which a single tone will be heard instead of two distinct ones. The perception of a single tone occurs when the areas of disturbance of the two tones along the cochlea partition overlap. The shaded regions in Figure 4-7 represent *critical bandwidths*, which indicate the width of the region of disturbance. If the critical bandwidths do not overlap, two distinct tones will be

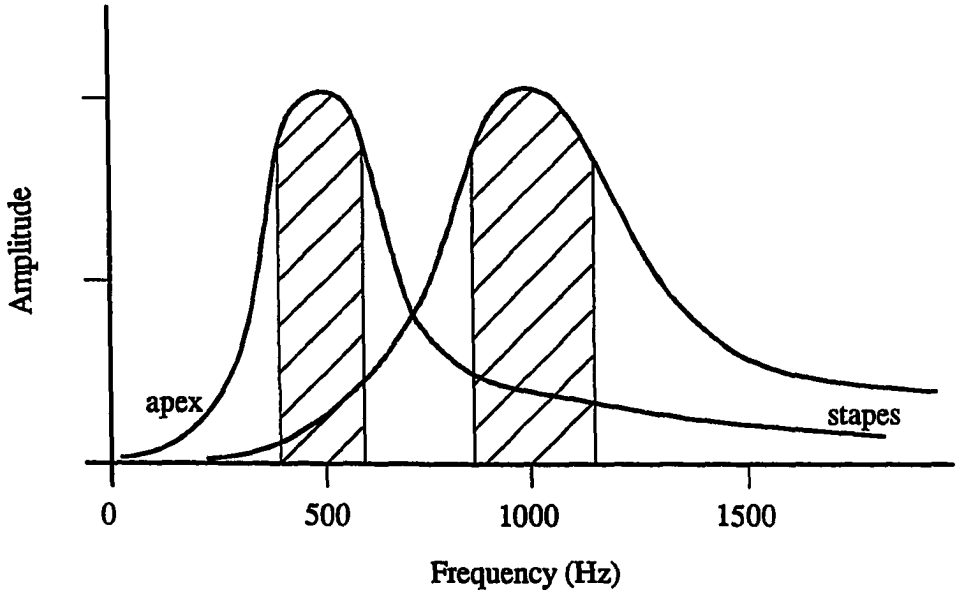


Figure 4-7 Amplitude of disturbances along the cochlea partition.

heard. Otherwise, a single tone will be heard. A table of critical bands and corresponding bandwidths as described by Scharf [10] is shown in Table 4-2.

4.3.3 Critical band smoothing

The process which leads to masking and critical band phenomena can be viewed as a convolution, in the frequency domain, of the power spectrum of the signal incident upon the ear with a spectral spreading function [40,41]. The width of this spreading function varies with frequency according to the width of a critical band. An alternate viewpoint of this process would be a convolution of a spreading function of constant bandwidth with the power spectrum of the incident signal which has been "warped" along its frequency axis according to a critical band scale.

Table 4-2 Critical band data.

Band number	Lower edge (Hz)	Center (Hz)	Upper edge (Hz)
1	0	50	100
2	100	150	200
3	200	250	300
4	300	350	400
5	400	450	510
6	510	570	630
7	630	700	770
8	770	840	920
9	920	1000	1080
10	1080	1170	1270
11	1270	1370	1480
12	1480	1600	1720
13	1720	1850	2000
14	2000	2150	2320
15	2320	2500	2700
16	2700	2900	3150
17	3150	3400	3700
18	3700	4000	4400
19	4400	4800	5300
20	5300	5800	6400
21	6400	7000	7700
22	7700	8500	9500
23	9500	10500	12000
24	12000	13500	15500
25	15500	19500	

To simulate this processing, the technique illustrated in Figure 4-8 was used. The magnitude spectrum of the specified incident signal is computed via a fast Fourier transform. The resulting magnitude spectrum is then smoothed using a variable-width Guassian window. The bandwidth of the smoothing window at a frequency f is selected such that 3 dB attenuation corresponds to a specified fraction of the critical bandwidth of the ear at frequency

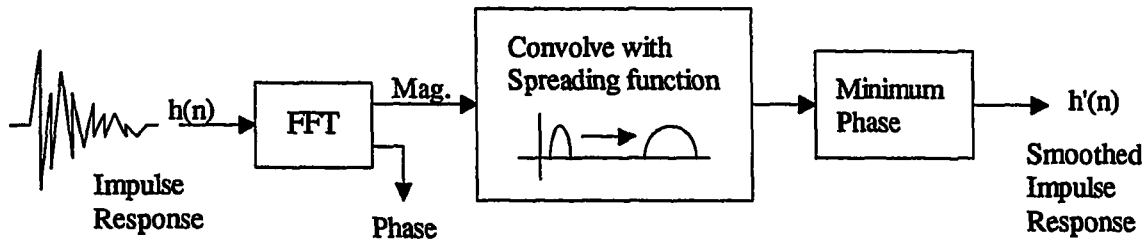


Figure 4-8 Critical band smoothing processing block.

f. The overall width of the window is selected based on a 80 dB bandwidth. For the remainder of this document, this process will be referred to simply as *critical band smoothing*.

To compute the critical bandwidth at an arbitrary frequency *f*, the critical band data listed in Table 4-2 is fit using a second order polynomial. A plot of this relationship, critical bandwidth versus frequency, based on a second-order polynomial fit is shown in Figure 4-9.

Critical band smoothing attempts to simulate the smoothing process associated with hearing. To demonstrate the effect of critical band smoothing on the frequency response of the DTFs, an impulse response of a sample DTF is processed using critical band smoothing.

Figure 4-10 shows a comparison between the original magnitude response (solid line) and the magnitude response after smoothing (dotted line). The amount of smoothing used was 0.5 critical bandwidths, meaning that the width of the smoothing window at a frequency *f* was 0.5 times the critical bandwidth at that frequency. The width of the smoothing window

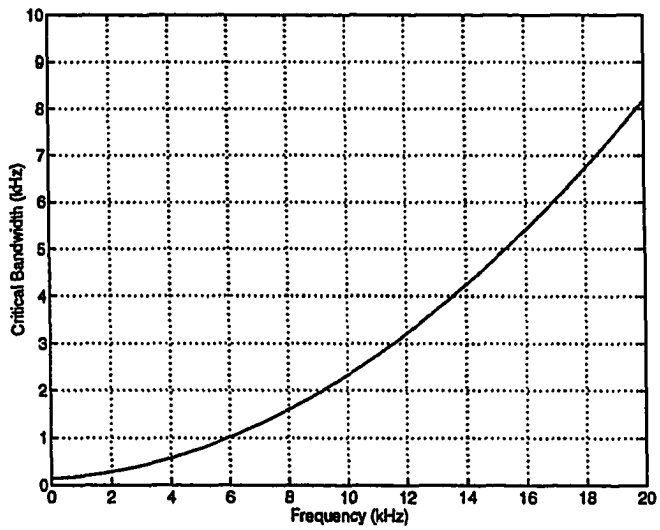


Figure 4-9 Critical bandwidth versus frequency.

expressed as a fraction of the critical bandwidth (in this case 0.5) will be referred to as the *smoothing factor*. It is clear that the critical band smoothing preserves the response at lower frequencies but smoothes the response at higher frequencies. Two important question arise as to the effects of critical band smoothing:

1. What effect does critical band smoothing have on the system modeling process?
2. How much smoothing is allowable before it becomes perceptually significant?

The first question is concerned with the effects of critical band smoothing on the ability of a model to fit a desired magnitude spectrum. It is anticipated that critical band smoothing will provide for a better fit by the model because the spectral resolution at higher frequencies has been reduced.

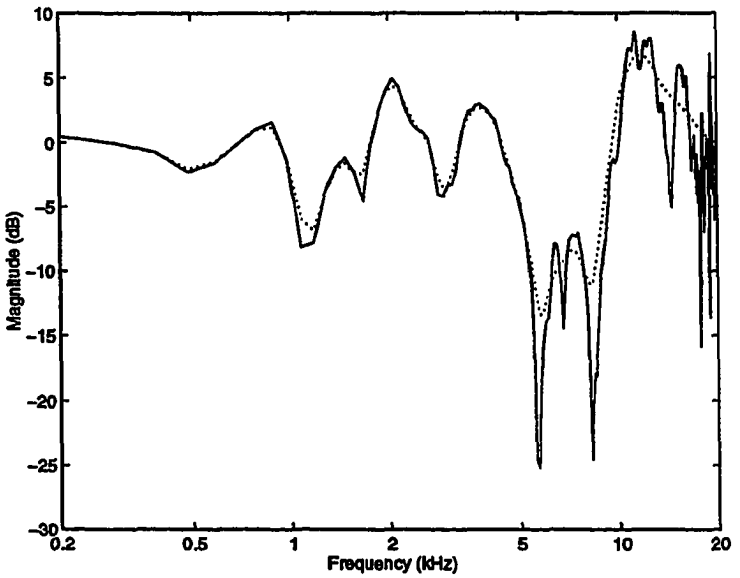


Figure 4-10 Response of a sample DTF compared to the smoothed response obtained using a smoothing factor of 0.5.

The second question deals with the amount of smoothing that can be allowed. If a signal is smoothed before it reaches the ear, then the ear will perform an additional smoothing on the already smoothed signal. Increasing the amount of smoothing may increase the ability of the model to fit the spectrum, but, at the same time, will introduce a potential perceptual error. Critical band smoothing provides for an increasingly better model fit of a gradually degrading impulse response. Obtaining the appropriate smoothing factor which will provide the best model fit without introducing perceptual distortion is certainly a topic of concern. More will be said about this topic in the next chapter.

4.3.4 Critical band distance measurement

An auditory-based criterion for objective evaluation of model performance is presented here. The critical band distance measurement measures the Euclidean distance between two power spectrums expressed in critical band levels. Figure 4-11 shows the processing involved in computing the critical band distance.

First, the power spectrums are computed for each transfer function. The critical band values are then computed by summing the energy in each critical band according to the bands given in Table 4-2. The energy in each critical band is expressed in dB. The Euclidean distance between each critical band spectrum expressed in dB is the resulting critical band distance. For this work, critical bands 3 through 24 will be used for computing critical band distance, giving a cumulative frequency range from 200 Hz to 15kHz. This measurement will be used to preliminarily demonstrate the performance of several techniques presented in the upcoming subsections of this chapter. Results from a more thorough evaluation, in terms of critical band distance, will be discussed in Chapter 5.

4.3.5 Least squares weighted error (LSWE) method

What is the effect of critical band smoothing on the least squares method of parameter estimation? Figure 4-12a shows a spectral fit of a 12th order model designed using the least-squares error method compared to that of a sample DTF smoothed using a 0.5 smoothing factor. It is apparent that, even though the filter specification was smoothed, the lower frequencies are still modeled poorly.

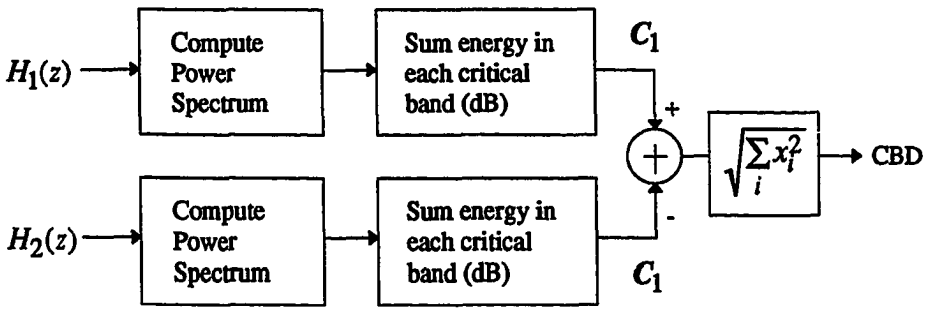


Figure 4-11 Critical band distance processing block.

The least-squares method minimizes an equation error energy. According to Parseval's theorem, the total energy in the time domain can be related to the total energy in the frequency domain. Thus,

$$\sum_{n=-\infty}^{\infty} \left| e_q(n) \right|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| E_q\left(e^{j\omega}\right) \right|^2 d\omega . \tag{70}$$

The total error energy can be computed in the frequency domain by integrating over the power spectrum. The least-squares method has a tendency to fit the spectrum according to a linear frequency scale. It would, therefore, be advantageous to utilize a modeling technique which weighted the error at the lower frequencies more strongly than the error at higher frequency. Such a weighting might take the form of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|W(e^{j\omega})E_q(e^{j\omega})\right|^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|W(e^{j\omega})\right|^2\left|E_q(e^{j\omega})\right|^2=\sum_{n=-\infty}^{\infty}\left|w(n)*e_q(n)\right|^2 \tag{71}$$

where $w(n)$ is the impulse response of the weighting filter, and $W(e^{j\omega})$ is the corresponding frequency response. The weighting filter must be selected such that lower frequencies are weighted more heavily than higher frequencies, as is the case of the auditory frequency scale. To accomplish such a task, a technique which will be referred to as the least-squares weighted error method will be introduced.

In the typical least-squares method, the energy of the equation error given by (8) is minimized. For convenience, (8) is repeated below

$$E_q(z)=A(z)Y(z)-B(z)X(z) \tag{72}$$

If each side of (72) is filtered by a weighting filter $W(z)$,

$$W(z)E_q(z)=A(z)W(z)Y(z)-B(z)W(z)X(z) \tag{73}$$

results. This can be rewritten as

$$E_{qw}(z)=A(z)\hat{Y}(z)-B(z)\hat{X}(z) \tag{74}$$

where

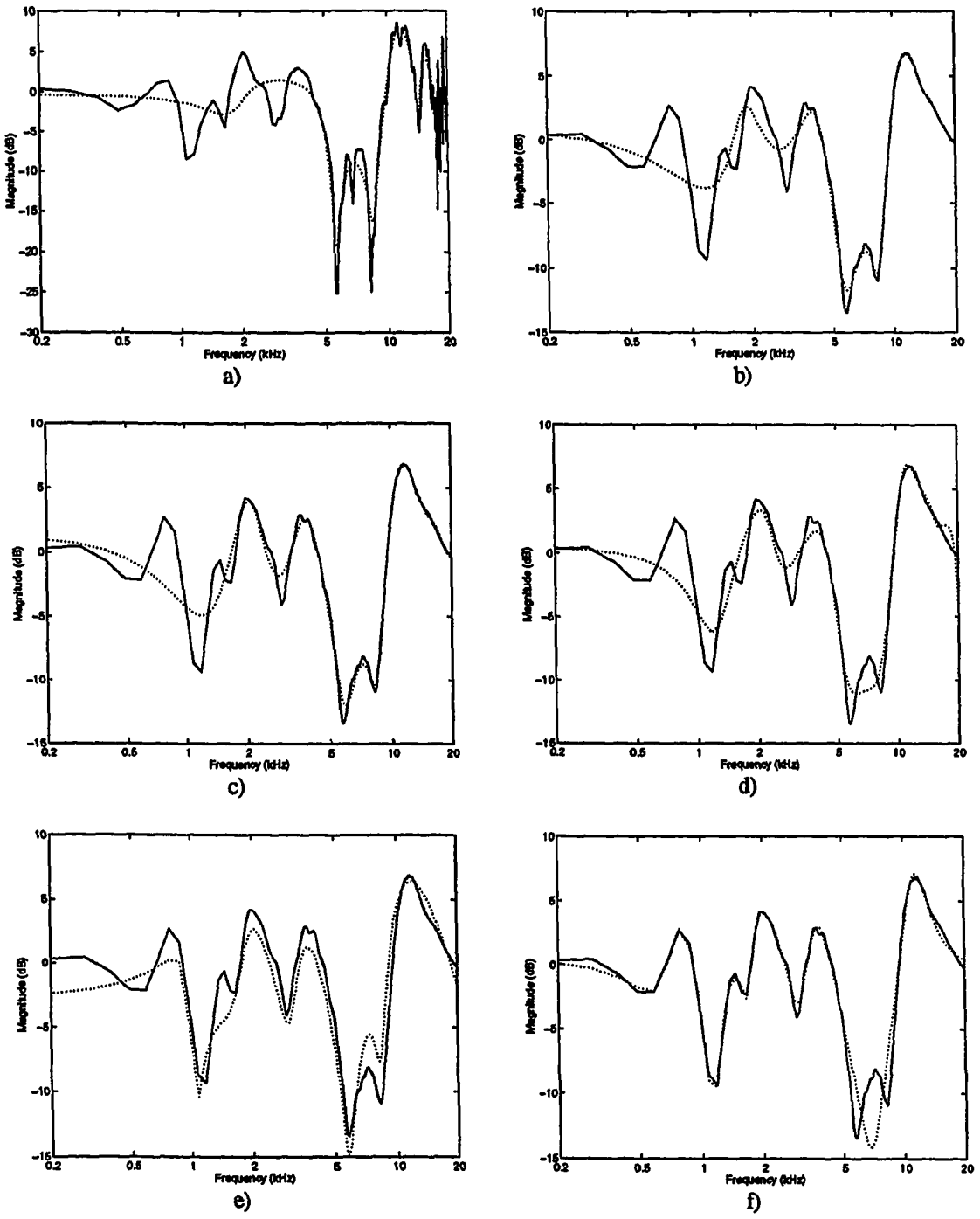


Figure 4-12 Comparison between smoothed spectrum and spectrum of a a) LS(12,12) model b) WLSE(12,12) model c) IP(12,12) model d) WIP(12,12) model e) PZC(12,12) model f) BMR(12,12) model.

$$\hat{Y}(z)=W(z)Y(z) \tag{75}$$

is the output $Y(z)$ prefiltered by $W(z)$,

$$\hat{X}(z)=W(z)X(z) \tag{76}$$

is the input $X(z)$ prefiltered by $W(z)$, and

$$E_{qw}(z)=W(z)E_q(z) \tag{77}$$

is a weighted equation error. One will note that the form of (74) is identical to the form of (72). Therefore, the least-squares method can be applied to solve (74). By prefiltering the excitation and response, a weighted error can be minimized. This technique is illustrated in Figure 4-13.

The weighting function can be selected to have a magnitude-squared response at each frequency f which is inversely proportional to the critical bandwidth at f. Based on the critical bandwidth shown in Figure 4-9, the desired magnitude response of $W(z)$ is computed as shown in Figure 4-14. The actual filter can be designed as a minimum-phase FIR filter.

For the sample DTF shown in Figure 4-12a, the LSWE method was applied. A comparison between a smoothed DTF response and the response of a 12th order model designed using LSWE method is shown in Figure 4-12b. Compared to Figure 4-12a, the LSWE method offers a significant improvement at lower frequencies. In terms of critical band distance, the difference between the LSWE(12,12) model and the smoothed filter prototype

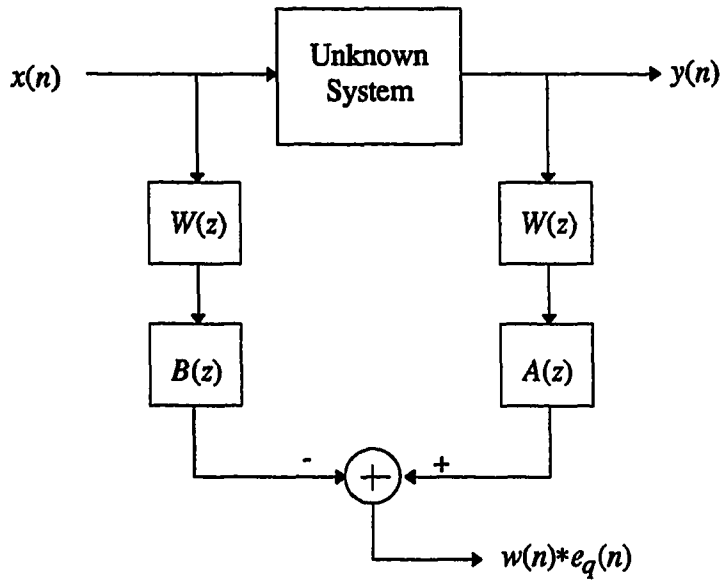


Figure 4-13 Diagram of least-squares weighted error method.

was 8.2 dB. When compared to a critical band distance of 13.1 dB in the case of the standard least-squares method, this improvement is significant.

4.3.6 Iterative prefiltering (IP) method

In a similar manner, the iterative prefiltering method can be applied to the modeling of smoothed DTFs. Unlike the LS method and WLSE method, the iterative prefiltering method attempts to minimize the modeling error energy as previously stated. Figure 4-12c compares the magnitude response of the sample DTF which has been smoothed using a 0.5 smoothing factor with the magnitude response of a 12th order filter designed using iterative prefiltering.

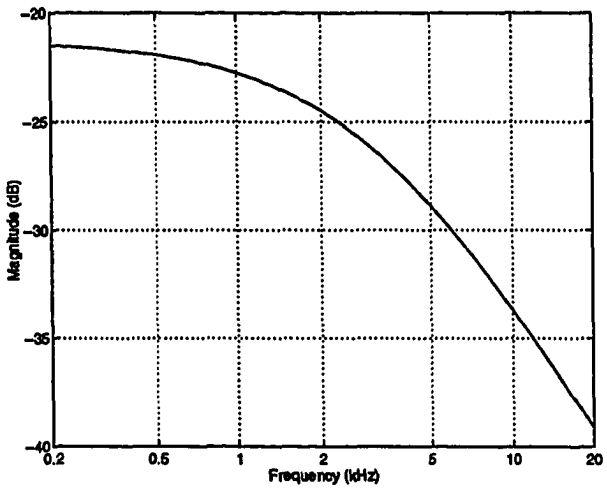


Figure 4-14 Magnitude response of weighting filter.

The initial parameter estimate for the iterative prefiltering method is obtained using the least-squares technique. The iterative prefiltering method does a much better job than the LS method and a slightly better job than the LSWE method. In terms of critical band distance, the IP(12,12) model results in a critical band distance of 7.4 dB.

4.3.7 Weighted iterative prefiltering (WIP) method

It would be beneficial to have the ability to weight the modeling error in the iterative prefiltering method in the same manner in which the least-squares error was weighted in the least-squares weighted error method.

In the iterative prefiltering method described in Chapter 3, the energy of an approximate modeling error was minimized. In (30), this modeling error was shown to be approximated as

$$E_m(z) \cong \frac{A(z)Y(z)}{\hat{A}_1(z)} - \frac{B(z)X(z)}{\hat{A}_1(z)} \tag{78}$$

where $\hat{A}_1(z)$ is the initial denominator parameter estimate. After the ith iteration of the iterative prefiltering method, this modeling error approximation becomes

$$E_m(z) \cong \frac{A(z)Y(z)}{\hat{A}_i(z)} - \frac{B(z)X(z)}{\hat{A}_i(z)} \tag{79}$$

where $\hat{A}_i(z)$ is the denominator parameter estimate obtained after the ith iteration.

If each side of (79) is filtered by a weighting filter $W(z)$, a weighted approximation of the modeling error given by

$$W(z)E_m(z) \cong \mathrm{W(z)}\frac{A(z)Y(z)}{\hat{A}_i(z)} - \mathrm{W(z)}\frac{B(z)X(z)}{\hat{A}_i(z)} \tag{80}$$

will result. This can be then be expressed as

$$W(z)E_m(z) \cong A(z)\hat{Y}_w(z) - B(z)\hat{X}_w(z) \tag{81}$$

where

$$\hat{X}_w(z) = W(z)\frac{X(z)}{\hat{A}_i(z)} \tag{82}$$

is the original input prefiltered by $1/\hat{A}_i(z)$ and then filtered by $W(z)$, and

$$\hat{Y}_w(z)=W(z)\frac{Y(z)}{\hat{A}_i(z)} \tag{83}$$

is the original output prefiltered by $1/\hat{A}_i(z)$ and then by $W(z)$. Figure 4-15 illustrates the resulting *weighted iterative prefiltering* method. The weighting filter is designed to be a minimum-phase, FIR filter with a magnitude response as shown in Figure 4-14.

The weighted iterative filtering method can be applied to directional transfer functions using the following steps:

1. Given a smoothed DTF impulse response, obtain a parameter estimate using the iterative prefiltering method. The least-squares method can be used to generate the initial parameter estimates for the iterative prefiltering method.
2. Prefilter the original input and output according to (82) and (83) using the iterative prefiltering denominator estimate from step 1 and the weighting filter $W(z)$.
3. Apply least-squares method to the prefiltered input and output to generate the WIP estimate.

The magnitude response of a model estimated using the weighted iterative prefiltering method is shown in Figure 4-12d. Compared to the standard iterative prefiltering method, the critical band distance has decreased only slightly from 7.4 dB to 7.0 dB.

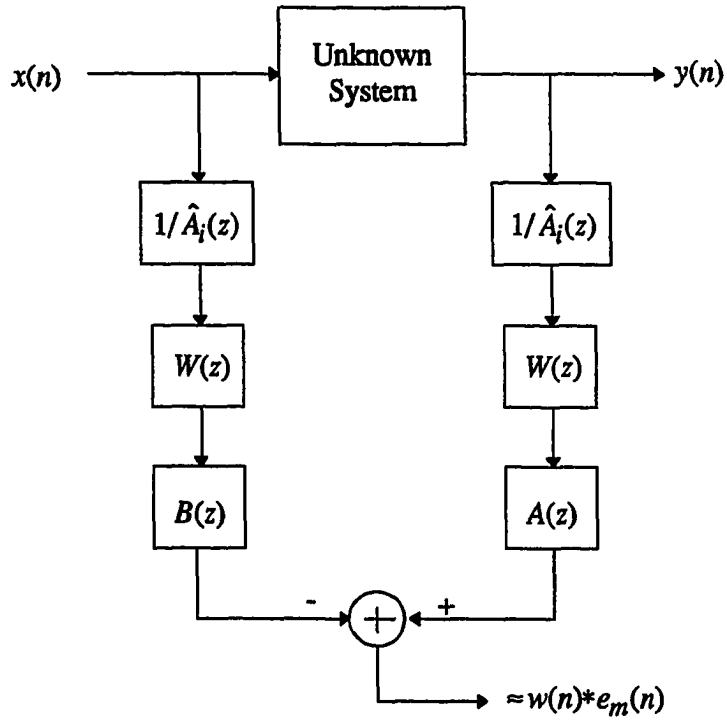


Figure 4-15 Illustration of the weighted iterative prefiltering method.

4.3.8 Pole-zero cancellation method

It is interesting to note the effect of critical band smoothing on the location of poles and zeros of the system models. Given the transfer function $H(z)$ of a model, the system poles and zeros can be obtained by performing a factorization on the numerator and denominator polynomials as indicated by (3)

$$H(z)=\frac{b_0(1-d_1z^{-1})(1-d_2z^{-1})\cdots(1-d_qz^{-1})}{(1-c_1z^{-1})(1-c_2z^{-1})\cdots(1-c_pz^{-1})}. \tag{84}$$

Figure 4-16 shows the location of poles and zeros for 30th order filters designed using the iterative prefiltering method for various critical band smoothing factors. Without critical band smoothing, the poles and zeros seem to be somewhat uniformly distributed about the entire frequency range, i.e., roughly the same concentration of poles and zeros at higher frequencies as there is at lower frequencies. When critical band smoothing is applied, one might expect the poles and zeros to shift toward the lower frequencies since the upper frequencies have been smoothed and are likely to be more easily modeled.

This, unfortunately, is not the case. As the amount of smoothing is increased, the concentration of poles and zeros appears to stay uniformly distributed across the entire frequency range. At the higher frequencies, however, the poles and zeros tend to "pair up". This pairing of poles and zeros corresponds to a "near-cancellation" of numerator and denominator factors in (84).

A possible method to reduce the model order without significantly sacrificing performance would be to remove the closest pole-zero pairs. Since a rational system with real coefficients will only have complex poles and zeros which occur in conjugate pairs, any cancellation of pole-zero pair in the upper half-plane must be accompanied by a cancellation of the corresponding pole-zero pair in the lower half-plane.

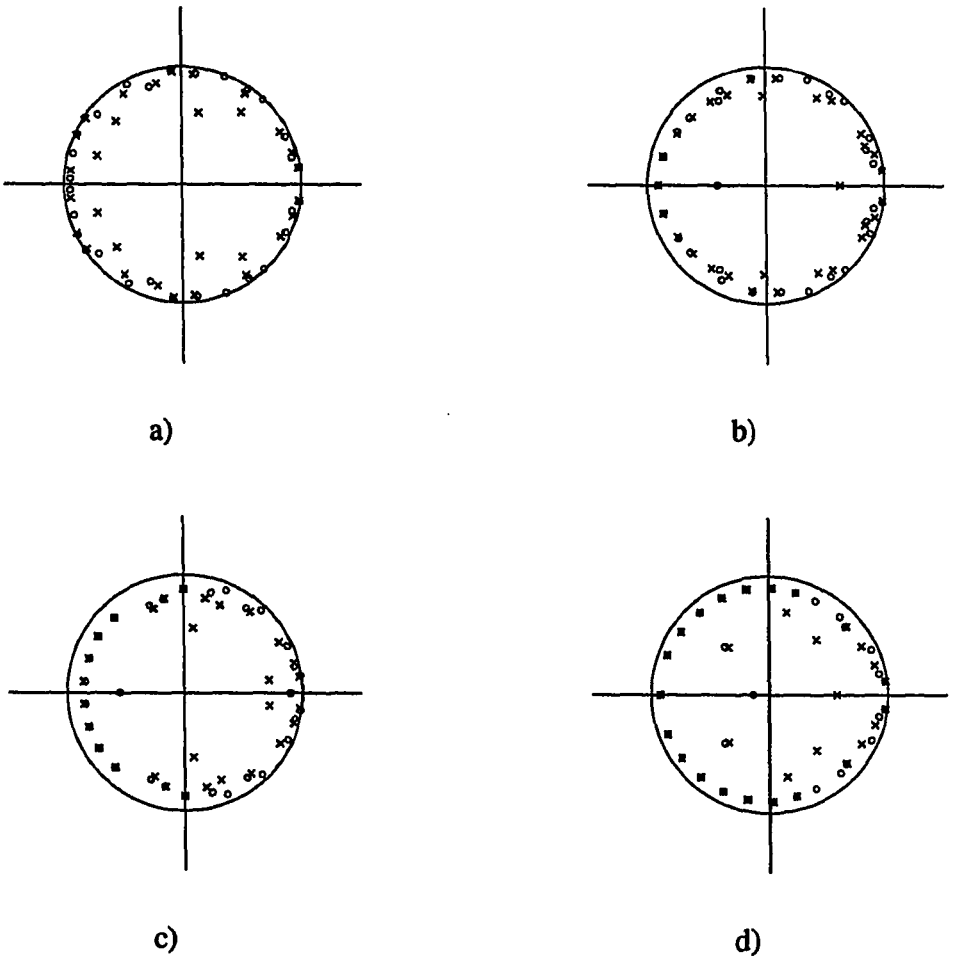


Figure 4-16 Pole-zero maps for models of a sample DTF designed using iterative prefiltering method. Smoothing factors are a) 0.0 b) 0.25 c) 0.5 d) 1.0. The real axis is the horizontal axis. The imaginary axis is the vertical axis.

Given the poles and zeros of a model as determined by (84), the distance between all pole-zero pairs can be computed as

$$D_{ij}=\left|d_j-c_i\right|, \qquad 1\le i\le p,\ 1\le j\le q \tag{85}$$

where d_j is a complex zero, and c_i is a complex pole.

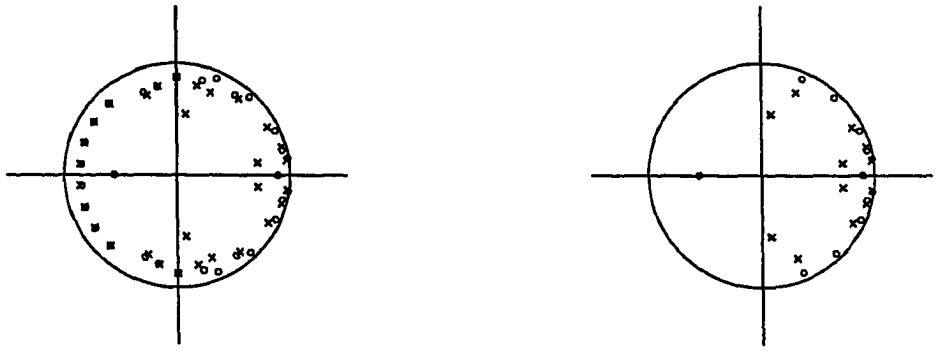
A simple model reduction can be obtained by canceling the pole-zero pairs with the smallest distance. Critical band smoothing followed by pole-zero cancellation results in an overall transfer function which will have a higher concentration of poles and zeros at lower frequencies.

Alternatively, the distance of pole-zeros pairs can be warped such that the distance of pole zero pairs at higher frequencies are contracted. Thus, a linearly warped distance function might be

$$D_{ij}=\frac{\left|d_j-c_i\right|}{\left|\angle c_i\right|}, \qquad 1\le i\le p,\ 1\le j\le q \tag{86}$$

where $\angle c_i$ is the frequency of the ith pole. For this work, however, the distance between pole-zero pairs will be warped according to a critical band scale, i.e., the distance between pole-zero pairs will be divided by the width of a critical band at the frequency of the pole:

$$D_{ij}=\frac{\left|d_j-c_i\right|}{cbw\left(f_s\cdot\frac{\left|\angle c_i\right|}{2\pi}\right)}, \qquad 1\le i\le p,\ 1\le j\le q \tag{87}$$



a) b)

Figure 4-17 Pole-zero plots for a) a IP(30,30) model b) a PZC(12,12) model.

where f_s is the sampling frequency and $cbw(\bullet)$ is a function which returns the critical bandwidth indicated by Figure 4-9 given a frequency f.

Figure 4-17 shows the results of the described pole-zero cancellation method. In (a), a pole-zero plot for a IP(30,30) model is shown based on a sample DTF smoothed using a 0.5 smoothing factor. In (b), a pole-zero plot of the corresponding 12th order model obtained using pole-zero cancellation is shown. It is clear that the closest pole-zero pairs have been removed. Figure 4-12e shows the magnitude response of the model compared to that of the smoothed DTF impulse response.

4.3.9 Balanced model reduction

A similar comparison can be used to study the effects of critical band smoothing on models resulting from balanced model reduction techniques. Figure 4-18 shows pole-zero maps for BMR(20,20) models obtained using the impulse response of a sample DTF for

various critical band smoothing factors. Unlike the least-squares and iterative prefiltering methods, the balanced model reduction technique automatically shifts poles and zeros away from the smoothed frequency region. As the smoothing factor is increased, the poles and zeros of the resulting model become concentrated at the lower frequencies. Thus, a spectral fit is obtained which matches well on an auditory frequency scale as opposed to a linear frequency scale.

Figure 4-12f shows the magnitude response of a BMR(12,12) model obtained using the impulse response of the sample DTF which has been smoothed using a 0.5 smoothing factor. Visually, the spectral fit of the BMR model is considerably better than any of the other techniques shown. In terms of critical band distance, the BMR(12,12) model is only 5.1 dB away from the smoothed impulse response prototype.

It is also interesting to note the effect on the Hankel singular values as the smoothing factor is increased. Figure 4-19 shows Hankel singular values for a sample directional transfer functions for a variety of smoothing factors. In general, the more quickly the singular values drop, the greater the model order reduction that can be achieved. For a smoothing factor of 0.25, the singular values drop significantly more quickly than for the case without smoothing.

As the smoothing factor is increased, however, the rate of decrease of the singular values remains relatively constant. The singular values for smoothing factors of 0.25, 0.5, and 1.0 are essentially the same. One might expect that this would indicate that the fit of the

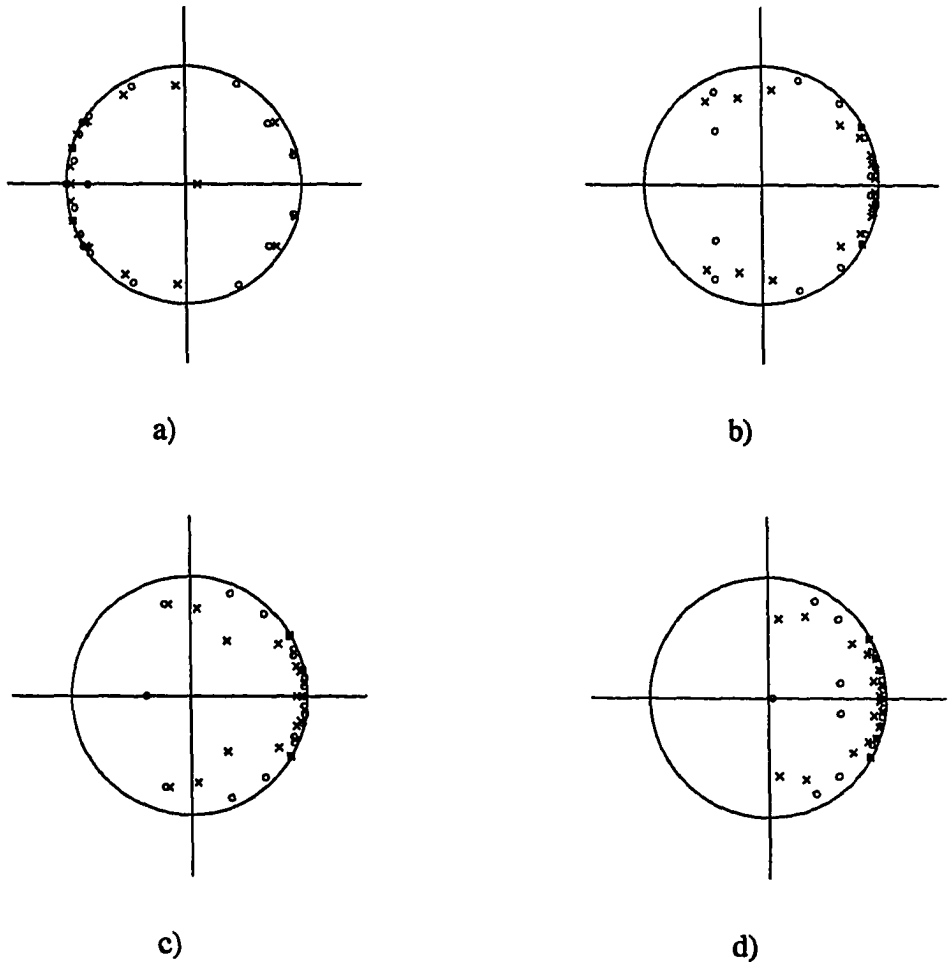


Figure 4-18 Pole-zero maps for models of a sample DTF designed using balanced model reduction method. Smoothing factors are a) 0.0 b) 0.25 c) 0.5 d) 1.0.

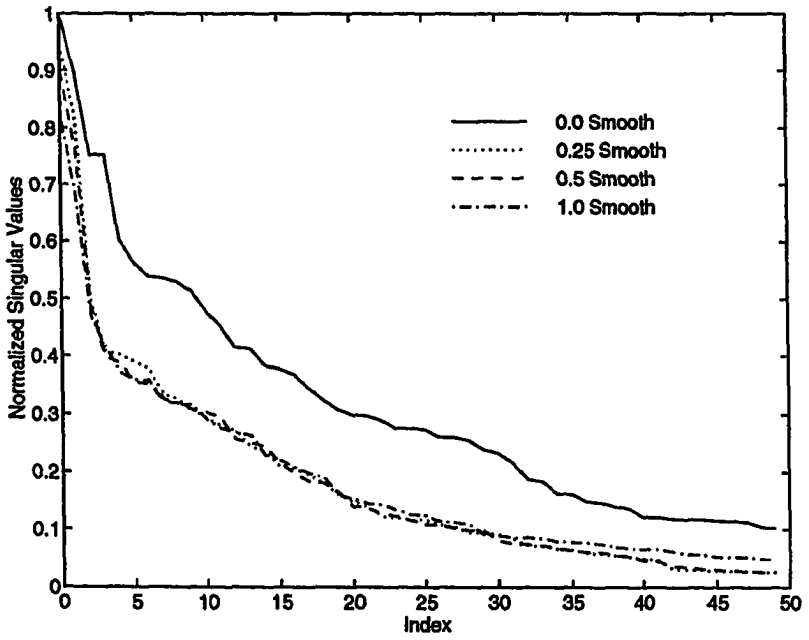


Figure 4-19 Hankel singular values for a variety of smoothing factors.

model would therefore not improve with increasing smoothing factor. Interestingly, this is not the case, as will be shown in Chapter 5.

4.4 Interpolation

In modeling head-related transfer functions or directional transfer functions, it is important to have the ability to interpolate between filters. Because it is not practical to measure head-related transfer functions at every possible location, it becomes necessary to

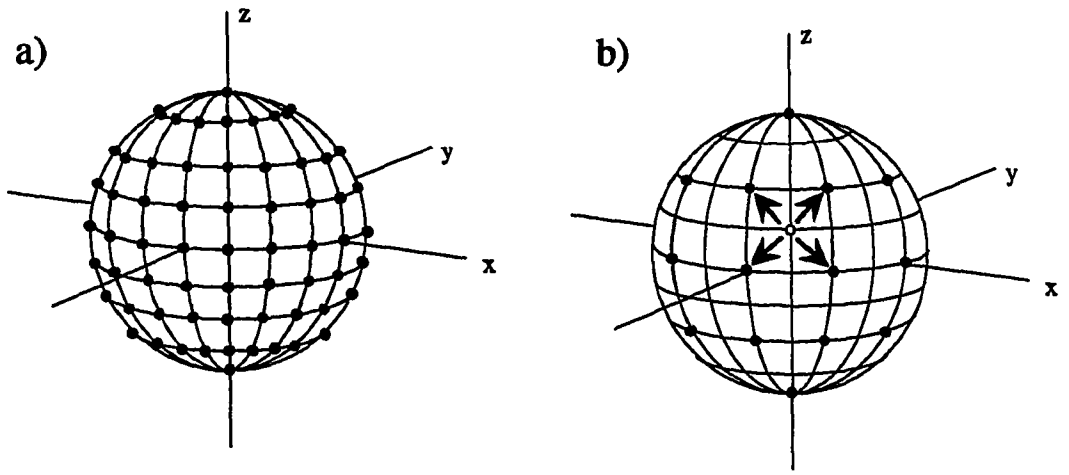


Figure 4-20 a) Uniformly spaced, dense set of filters. b) Interpolation of a uniformly spaced, sparse set of filters.

interpolate for positions not measured, such as illustrated in Figure 4-20. In addition, to store the filter coefficients for all possible locations would require a significant amount of storage. Interpolating between a sparse set of filters could significantly reduced the amount of storage required.

Interpolation between two rational systems, $H_1(z)$ and $H_2(z)$, can most easily be accomplished by interpolating directly between the corresponding impulse responses. For FIR filters, this is identical to interpolation between filter coefficients. For IIR filters, however, interpolation of filters can generally not be achieved by simply interpolating between filter coefficients, since interpolation of the denominator coefficients directly is not guaranteed to result in a stable filter.

An indirect technique for interpolating between two arbitrary rational systems $H_1(z)$ and $H_2(z)$ can be achieved as shown in Figure 4-21a. The interpolation is performed directly between the impulse responses of the two models. The resulting interpolated impulse response is then remodeled. There are obvious disadvantages to such a technique. The computational requirements would be significant because not only would the impulse response of each system have to be computed, but the entire modeling process would have to be performed again. Such a technique could be used to precompute and store coefficients for interpolated models.

A more direct approach is shown in Figure 4-21b. In this approach, the numerator and denominator coefficients are interpolated separately. Interpolation of the numerator coefficients is performed directly. The denominator coefficients, however, are first converted to *reflection coefficients* using a recursive procedure commonly referred to as a *step-down* procedure, as described in [33]. Interpolation is then performed directly on the corresponding reflections coefficients.

For a stable system, all reflection coefficients must have magnitudes less than one. Thus, interpolation directly between reflection coefficients will result in a new set of reflection coefficients which will also have magnitudes less than one, thus yielding a stable system. The resulting interpolated reflection coefficients will then have to be converted back into

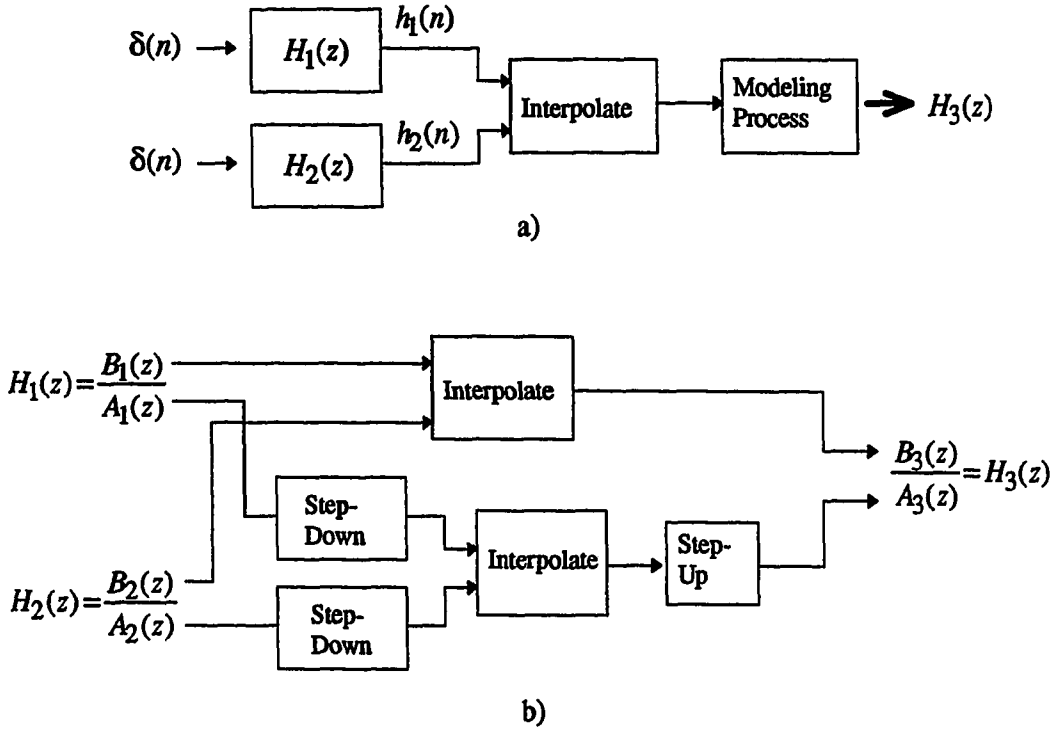


Figure 4-21 a) Indirect interpolation method. b) Direct interpolation method.

denominator coefficients using a *step-up* procedure, as described in [33]. The direct interpolation could be used in a real-time system to interpolate "on-the-fly". Computational requirement for such a process are discussed in the next chapter.

5. RESULTS

This chapter will present results obtained using the modeling techniques discussed in Chapter 4. In the first section, the performance of several of the modeling techniques will be evaluated in terms of critical band distance. Critical band distance measurements will be presented using a range of critical band smoothing factors and model orders. In so doing, the effect that critical band smoothing has on the overall modeling process will be observed. By determining the model order required by each technique to reach a critical band distance threshold, a ranking of these techniques will be made.

A more practical way of ranking the proposed modeling techniques might be to evaluate them based on the amount of computation required for implementation. In Section 5.2, the amount of computation, in number of multiply operations, required for each modeling technique to reach a desired critical band distance threshold will be computed. Calculations will be made for several filter structures, including direct, cascade, and lattice forms. This will allow a comparison of modeling techniques to be made based on implementation complexity.

Section 5.3 will present results from localization tests. Results will demonstrate the extent to which the modeling techniques are successful in preserving localization performance. Data from three subjects will be presented. For each subject, the localization performance of a high-order FIR filter, a low-order filter designed using the iterative prefiltering method, and a low-order filter designed using the balanced model reduction method will be compared.

These results will indicate any noticeable differences between the high-order FIR filter and the two low-order filters in terms of localization performance.

5.1 Critical Band Distance Measurements

5.1.1 Effect of smoothing

A question remains as to how much critical band smoothing should be used in the filter design process. In Chapter 4, two questions were proposed for analyzing the overall effects of critical band smoothing:

1. What is the effect of critical band smoothing on the modeling process?
2. How does critical band smoothing affect the perceptual characteristics of the system?

The first question is concerned with the relationship between the response of the model and the response of the *smoothed* system. In other words, how well does the model fit the smoothed spectrum of the original system, and how does the amount of smoothing affect this fit? The second question is concerned more with the relationship between the response of the original system and the smoothed response of that system, i.e., how well does the smoothed system match the original system? To study the relationships posed by these questions, the critical band distance measurement will be employed.

Ultimately, one is concerned with how well the system model performs in comparison to the original system. In terms of critical band distance, one is primarily concerned with the distance between the model response and the original system response. The critical band

distance measurement illustrates an interesting relationship which exists among the original system response, the smoothed system response, and the system model response.

If one defines the vectors C and C_m to respectively be the spectrum of the original system and the spectrum of the corresponding system model expressed as critical band levels in dB, then the critical band distance between C and C_m is simply the Euclidean distance between C and C_m. This critical band distance can be denoted as $d(C,C_m)$, where $d(x,y)=\|x-y\|$ simply represents the Euclidean distance between a point represented by vector x and a point represented by vector y.

Similarly, let the vector C_s denote the smoothed spectrum of the original system expressed in critical band levels for some unspecified smoothing factor. The critical band distance between the smoothed system response and the response of the system model is then $d(C_s,C_m)$. Likewise, the critical band distance between the original system response and the smoothed system response is then $d(C,C_s)$.

An elementary relationship, the *triangle inequality*, can be used to mathematically relate the three quantities $d(C,C_m)$, $d(C_s,C_m)$, and $d(C,C_s)$. This relationship, given by

$$d(C,C_m) \leq d(C,C_s)+d(C_s,C_m) \tag{88}$$

is illustrated in Figure 5-1.

As an example of this inequality, the balanced model reduction method was used to compute the parameters of BMR(12,12) models for directional transfer functions of one subject for 90 randomly selected positions not limited to the horizontal plane. The DTF

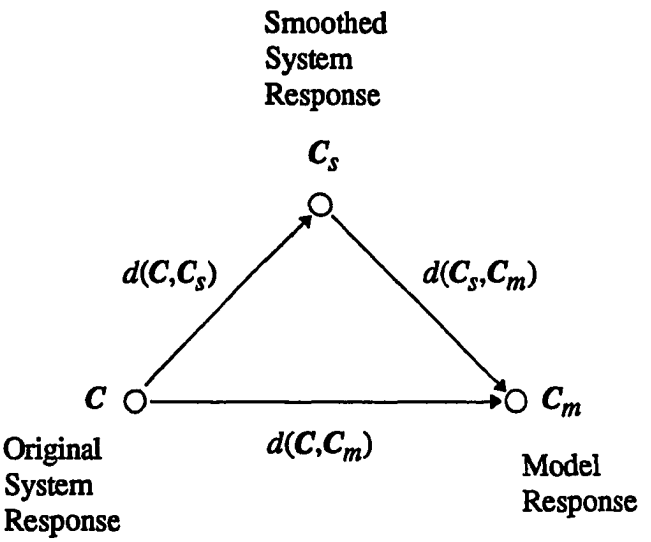


Figure 5-1 Relationship of critical band distances.

impulse response was smoothed prior to application of the modeling technique using a variety of smoothing factors, ranging from 0.0 to 0.5 using increments of 0.1. For each smoothing factor, the distances $d(C,C_m)$, $d(C_s,C_m)$, and $d(C,C_s)$ were computed and averaged over all 90 positions. These distances are plotted in Figure 5-2

As the amount of smoothing is increased, the system model forms an increasingly better fit of the smoothed system. Thus, $d(C_s,C_m)$ is decreasing for an increasing smoothing factor. The distance between the original system and the smoothed system, however, will generally be increasing. Thus, $d(C,C_s)$ increases with increasing smoothing factor. The overall critical band distance between the original system response and the model response

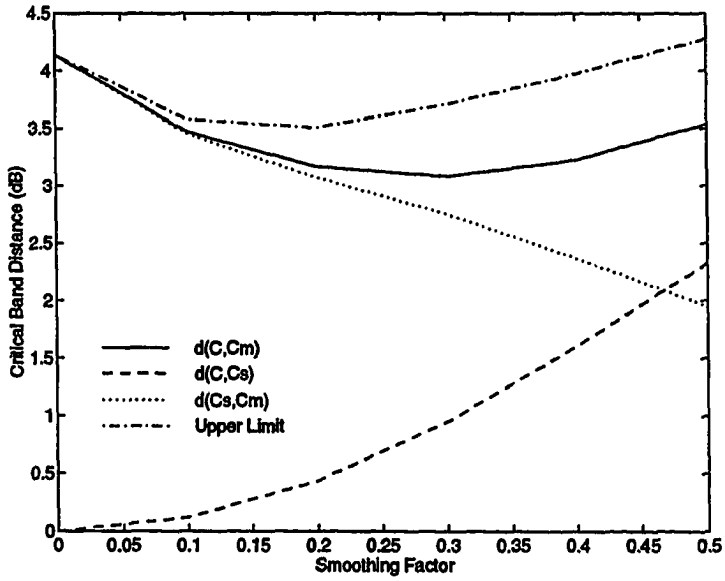


Figure 5-2 Critical band distances $d(C,C_m)$, $d(C_s,C_m)$, and $d(C,C_s)$ versus smoothing factor for 12th order models designed using balanced model reduction method.

$d(C,C_m)$ will therefore be upper bounded by the sum of $d(C,C_s)$ and $d(C_s,C_m)$, as indicated by the dashed-dotted line. The actual $d(C,C_m)$ is indicated in the figure by the solid line.

The optimal amount of smoothing in this situation can be selected as the minimum point on the solid curve. This point gives the critical band smoothing factor which results in minimum critical band distance. In this example, the actual critical band distance $d(C,C_m)$ achieved a minimum distance for a smoothing factor of 0.3.

It is interesting to note the fact that the quantity $d(C,C_s)$ is independent of the modeling technique employed. According to Figure 5-2, a small smoothing factor results in

low rate of increase in $d(C,C_s)$, as would be expected. At a smoothing factor of 0.2, the critical band distance $d(C,C_s)$ is only 0.5 dB. As the smoothing factor is increased, however, the critical band distance between the original filter specification and the smoothed filter specification increases at a greater rate. Because $d(C,C_s)$ is independent of the modeling technique applied, a comparison of modeling techniques will proceed by examining $d(C_s,C_m)$ and $d(C,C_m)$ for several modeling techniques presented in Chapter 4 as the smoothing factor is varied.

5.1.2 Comparison of modeling techniques

To compare the effect of smoothing on the modeling techniques, the directional transfer functions of one subject for the same 90 positions previously used were modeled using 25 coefficient models. For IIR filters, a 25 coefficient model will refer to a 12th order numerator and a 12th order denominator. For an FIR filter, a 25 coefficient model will refer to a 24th order numerator and a 0th order denominator. Critical band smoothing was applied prior to the modeling process. For each position, the smoothing factor was varied from 0.0 to 0.5 using increments of 0.1.

Figure 5-3 shows critical band distance $d(C_s,C_m)$ for each smoothing factor averaged over all positions. Average critical band distance measurements are shown for the LS, LSWE, IP, WIP, PZC, and BMR models. For the PZC model, an IP(30,30) system was reduced by simple pole-zero cancellation of the nearest pole-zero pairs. For comparison a 24th order FIR filter was computed by simple truncation of the impulse response. The LS, FIR, LSWE, and

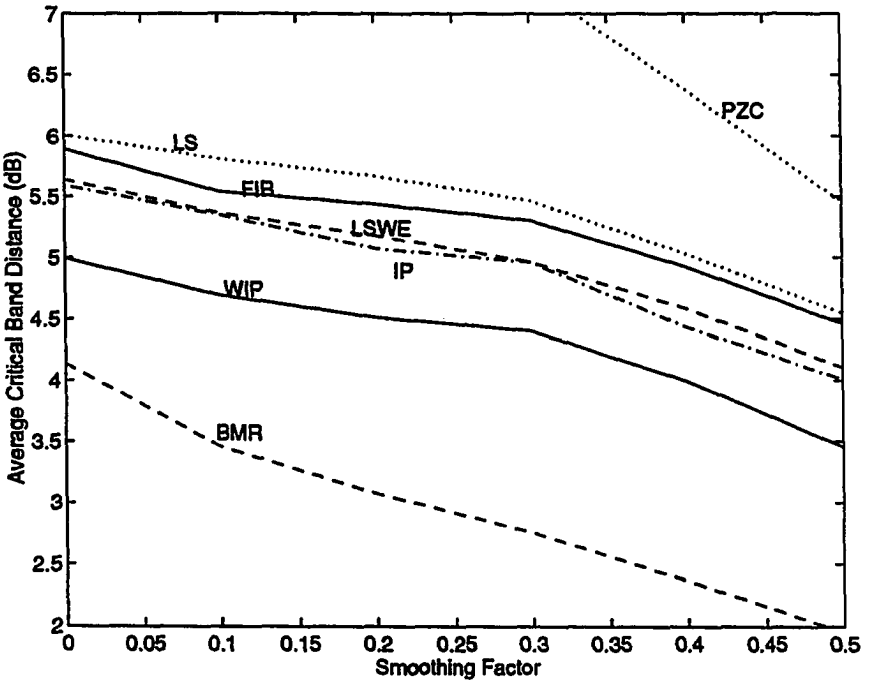


Figure 5-3 Average critical band distance $d(\boldsymbol{C}_s,\boldsymbol{C}_m)$ between smoothed filter specification and response of model for 25 coefficient models and varying smoothing factor.

IP models all seem to perform similarly. In fact, these methods are all within 0.5 dB of one another.

The WIP model shows some improvement. It is roughly 0.5 dB better than the IP model of the same order. The performance of the PZC model is somewhat disappointing. For low smoothing factors the average critical band distance is not even within the limits of the graph. For high smoothing factors, however, it appears as though the performance of the PZC method would be comparable to that of other techniques. It is clear that the balanced model reduction method performs considerably better than the other methods shown.

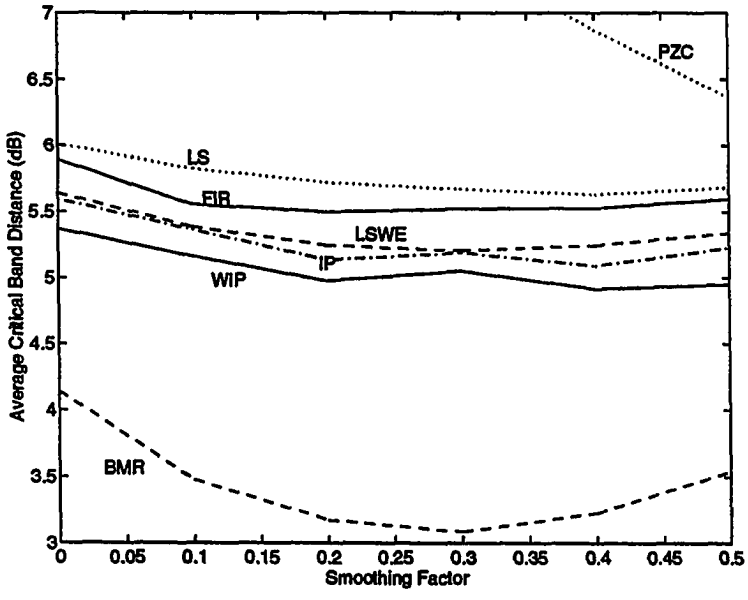


Figure 5-4 Average critical band distance $d(C,C_m)$ between original filter response and response of model for 25 coefficient models and varying smoothing factor.

Similarly Figure 5-4 shows the critical band distance $d(C,C_m)$ versus smoothing factor averaged over all 90 positions. With the exception of the BMR model, critical band smoothing does not seem to have a significant effect on the overall critical band distance. For the range of smoothing factors shown, the difference, in terms of average critical band distance, between the FIR, LS, LSWE, and IP models is less than 0.5 dB, although the IP model has a smaller critical band distance at every point. Again, the BMR model shows superior performance. The BMR model at this order has as much as a 1.5 dB decrease in

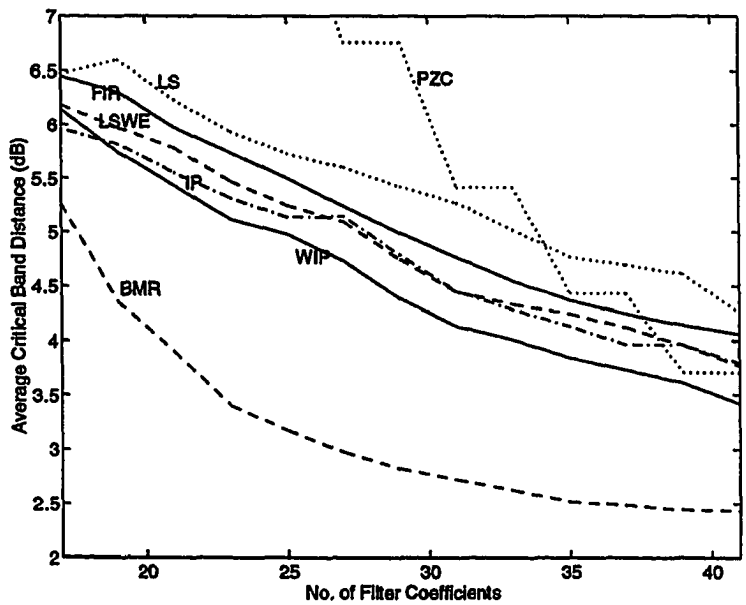


Figure 5-5 Critical band distance $d(\boldsymbol{C},\boldsymbol{C}_m)$ between original filter specification and response of the model for a variety of modeling techniques and fixed smoothing factor.

critical band distance compared to the WIP model and as much as 2.5 dB compared to the FIR model.

Up until this point, the model order has been fixed while the smoothing factor was varied. To study the relative performance of the modeling techniques discussed as the model order is varied, a comparison was made using a fixed critical band smoothing factor. Figure 5-5 illustrates the relative performance in terms of the critical band distance $d(\boldsymbol{C},\boldsymbol{C}_m)$ of various techniques over a range of model orders using the same set of 90 directional transfer functions for a smoothing factor of 0.3.

It is apparent that the balanced model reduction achieves the overall best performance of the models shown. At lower orders the rate of decrease in critical band distance is greater than for other models. The balanced model reduction method has its greatest advantage at around 27 coefficients for this smoothing factor.

The PZC model performs poorly at low orders, but performs well at high model orders. Since the PZC model in this case was obtained through pole-zero cancellation of an IP(30,30) model, it is believed that the performance of the PZC model will approach that of an IP(30,30) as the model order is increased. The order of the original system from which the PZC model is obtained is likely to have a significant effect on the critical band distance of the PZC model and could be one reason that the PZC shows poor performance. In this work, all PZC models are obtained from IP(30,30) models. The effect of the initial system order on the average critical band distance will not be studied.

To study the combined effect of smoothing factor and model order, the average critical band distance between the original filter specification and the response of the model was computed over a range of smoothing factors and model orders. Results for various techniques are tabulated in Appendix C.

To summarize these results, the optimal model order required to meet a given average critical band distance was computed for each technique. The amount of smoothing which minimized the critical band distance at this order was also found. Table 5-1 lists the optimal (model order, smoothing factor) pair required to meet a desired average critical band distance. For instance, to obtain a set of filters using the balance model reduction technique which results in an average critical band distance of 4 dB, one would need to use a smoothing factor

of 0.4 and a model order of 10. If a given critical band distance threshold is outside of the range of the tables in Appendix C, the model order is listed as an upper bound order or a lower bound order and the smoothing factor is left unspecified.

For the balanced model reduction method, a contour plot showing constant critical band distance contours is shown in Figure 5-6. For the plot, the horizontal axis indicates smoothing factor and the vertical axis indicates model order. Two major observations can be made from the data.

Firstly, the optimal smoothing factor and model order required to reach a desired average critical band distance can be selected as the minimum model order on the corresponding critical band distance contour curve. Secondly, the optimal amount of smoothing increases as the desired critical band distance increases. Thus, to meet a 2.5 dB critical band distance constraint, a model order of 16 with a smoothing factor of 0.2 could be selected. However, to meet a 4 dB average critical band distance constraint, a model order of 10 with a smoothing factor of 0.3 would be required.

5.2 Computation

A number of filter structures can be used to implement a given a rational system model with transfer function

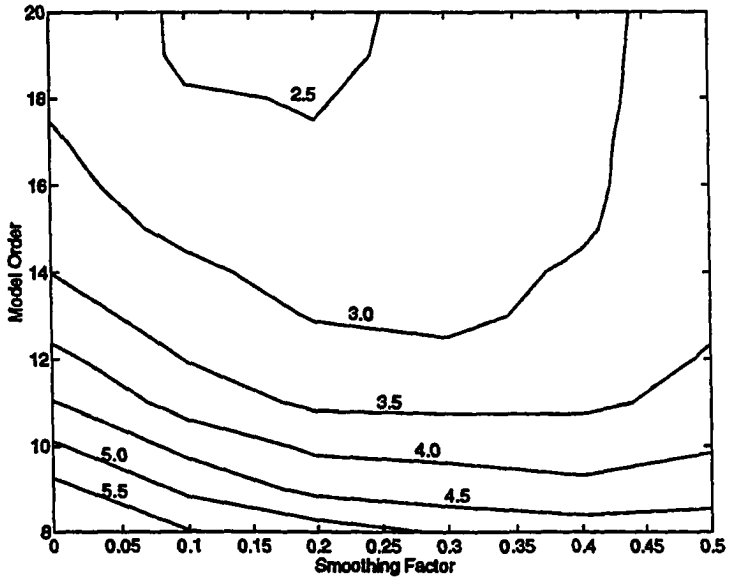


Figure 5-6 Contour plot showing constant critical band distance contours for balanced model reduction method.

Table 5-1 Optimal (model order, smoothing factor) for various modeling techniques.

		FIR	LS	LSWE	IP	WIP	PZC	BMR
Average Critical Band Distance	3 dB	(>40,?)	(>20,?)	(>20,?)	(>20,?)	(>20,?)	(>20,?)	(13,0.3)
	4 dB	(>40,?)	(>20,?)	(19,0.1)	(18,0.2)	(17,0.3)	(17,0.4)	(10.0.4)
	5 dB	(28,0.1)	(16,0.4)	(13,0.3)	(13,0.4)	(12,0.2)	(15,0.0)	(8,0.4)
	6 dB	(20,0.2)	(10,0.4)	(9,0.3)	(8,0.3)	(8,0.4)	(13,0.3)	(<8,?)

$$H(z)=\frac{B(z)}{A(z)}=\frac{b_0+b_1z^{-1}+b_2z^{-2}+\cdots+b_qz^{-q}}{1+a_1z^{-1}+a_2z^{-2}+\cdots+a_pz^{-p}}. \tag{89}$$

The work presented here will focus on systems in which the numerator and denominator polynomials are of equal order, i.e., $p=q$.

Perhaps the simplest approach would be a *direct* form. Two examples of direct forms are shown in Figure 5-7a and Figure 5-7b. Direct forms are among the easiest structures to implement because the numerator and denominator coefficients in (89) can be used directly. The structure shown in Figure 5-7a is commonly referred to as direct form I. For an Nth order numerator and an Nth order denominator, the direct form I structure requires $2N$ delays and $2N+1$ multiplications. A similar structure, shown in Figure 5-7b, is the direct form II structure. This form requires the same number of multiplications, but only requires N delays.

Direct forms are simple to construct and have minimal computation and storage requirements. They are, however, sensitive to parameter variation, an important consideration in finite-precision implementation of IIR filters. For this reason, two other common structures are presented. Because the sensitivity of an IIR filter generally increases with filter order, a simple approach to lowering the filter sensitivity would be to construct the filter as a series of cascaded low-order systems. Typically, second-order systems are used, referred to as *biquads*.

Figure 5-7c shows a *cascaded* filter structure based on second-order system. Each second order system is implemented in direct form. A direct form second-order system requires 5 multiplications. Therefore, for an Nth order system, where N is even, the total

number of multiplications for the cascaded form is 2.5N, and the computational requirements are not significantly greater than those of the direct form implementation. Thus, for a slight increase in computation, a decrease in sensitivity can be achieved using cascaded systems. As an added savings, the gain of each second-order section can be accumulated into a single gain constant. Each unity-gain, second-order section would therefore require only 4 multiplications. For an Nth order system, where N is even, the total number of multiplications would be only $2N$+1. In a practical implementation, however, additional overhead may be significant.

A third common structure is the *lattice* form. In the lattice structure, neither the numerator nor the denominator polynomial coefficients are used directly. Instead, the denominator coefficients (a_i's) are converted into reflection coefficients using a recursive step-down procedure, as discussed in Section 4.4. The lattice filter structure is shown in Figure 5-7d. Here, the k_i's are the reflection coefficients of the all-pole system formed from the denominator polynomial and the c_i's are found as the solution to

$$b_m = c_m - \sum_{m=0}^{q} c_i a_{i-m}^{(i)} \tag{90}$$

where the $a_i^{(j)}$'s are the autoregressive coefficients obtained during the jth step of the recursive step-down solution for the reflection coefficients [18].

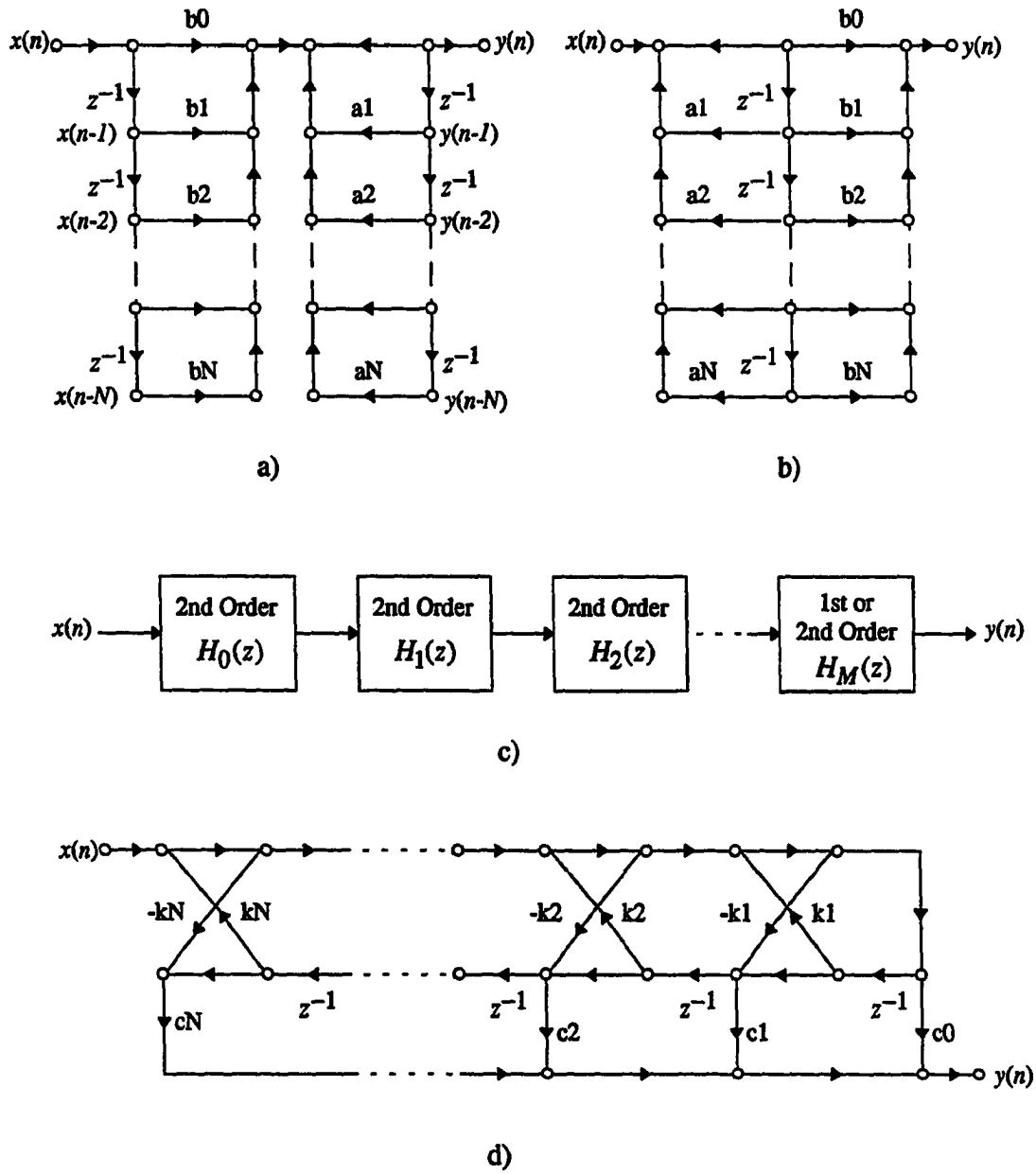


Figure 5-7 Filter structures. a) Direct form I. b) Direct form II. c) Cascade. d) Lattice.

Table 5-2 Number of multiplications required to meet 5 dB critical band distance.

	FIR	LS	LSWE	IP	WIP	PZC	BMR
Direct	29	33	27	27	25	31	17
Cascade	-	33	27	27	25	31	17
Lattice	-	49	40	40	37	46	25

The lattice structure has lower sensitivity than either the direct or cascaded forms, but it also has increased computational requirements. For an Nth order filter with an Nth order numerator and an Nth order denominator, the lattice filter requires $3N+1$ multiplications. The lattice structure does, however, lend itself well to interpolation since the reflection coefficient used in its construction can be interpolated directly without risk of instability.

To summarize the computational requirements for directional transfer function system models using the three filter structures described, the total computation required to meet a critical band distance threshold of 5 dB was computed for a number of modeling techniques. Using the minimum model orders tabulated in Table 5-1, the number of multiplications for each of the three filter structures was computed. Table 5-2 lists the results. The computational effort required for other critical band distance thresholds can be found from the data in Table 5-1. These data are not tabulated.

Compared to a simple FIR filter, it does not appear that any of the techniques shown result in a significant savings in computation with the exception of the WIP and BMR models. The LS and PZC methods actually result in increased computation. The LSWE and IP

methods result in a minor reduction in computation for the direct and cascade structures, but result in a considerable increase in computation using the lattice structure.

The WIP and BMR models show the most promise for computational reduction for this critical band distance. Compared to the FIR filter, these models result in a 41% reduction in number of multiplications for the same critical band distance when implemented in direct form and a 14% reduction in number of multiplications when implemented in lattice form.

5.3 Localization Tests

5.3.1 Objective

To study the effect of these modeling techniques on the localization performance, a series of localization tests was performed. The objective of the localization tests was to compare the performance of the low-order models to the performance of a high-order FIR filter. The FIR filter was selected as a 100 coefficient filter obtained by means of truncation of the minimum-phase DTF impulse responses. Based on informal listening tests, two low-order models, an IP(15,15) model and a BMR(10,10) model, were selected for comparison. Both models were obtained using a smoothing factor of 0.2. In terms of critical band distance, the IP(15,15) and the BMR(10,10) models have similar performance (3.9 dB and 4.5 dB respectively).

5.3.2 Synthesis

A real-time system was developed to test the operation of the filters designed in the previous chapter. A general-purpose, PC-based signal processing board (Spectrum

TMS320C30 System Board) was used to implement the filters. The BMR(10,10) and IP(15,15) models were implemented in direct form on a Texas Instruments TMS320C30 floating-point processor. Output signals were generated using on-board 16-bit stereo D/A converters, programmed for a 16 kHz sampling rate.

For localization tests, input was generated from a circular internal noise buffer, although a 16 bit on-board A/D converter could be utilized as well. Filter coefficients, along with interaural time delays, for tested positions were computed offline and stored in on-board memory. Delay values for minimum-phase DTFs were computed as the delay which resulted in maximum correlation with the original impulse response. Control of the system was maintained via a graphical user interface running under the Microsoft Windows operating environment.

5.3.3 Procedure

Three young adults acted as subjects. All had normal hearing with no history of hearing problems. All had had previous experience in localization experiments.

The stimulus consisted of a 250 ms burst of white, Guassian noise which was then bandpass filtered from 200 Hz to 15 kHz using a 2048-tap linear-phase FIR filter.

A random set of 144 positions was selected from a spherical shell of 505 possible positions. For each subject tested, four runs of the data set were made for each model resulting in a total of 576 trials per subject per model. The ordering of the positions in the data set was scrambled for each run.

At each position tested, the stimulus was filtered using the experimental device described above and presented through a pair of Sennheiser HD-430 headphones. Filters were modeled using individualized minimum-phase directional transfer functions computed from the head-related transfer functions measured for each subject. Additional headphone corrections were not made.

The recording of listener feedback was similar to that of [3]. The apparent location of the source, as perceived by the listeners was relayed verbally. Listeners would respond by calling the apparent azimuth and elevation angles according the spherical coordinate system specified in Figure 2-1. A certain amount of error in estimating the angle of the auditory event was expected for each subject. This error, however, was not taken to be an issue of great concern because it was assumed that the judgment error would statistically be essentially the same for all models tested. Distance judgments were neither recorded nor requested.

5.3.4 Analysis

The data resulting from the above described testing procedure consisted of 576 intended source positions and 576 recorded responses for each subject for each of three modeling cases. Results are shown in Figure 5-8 through Figure 5-13. To analyze these data, the spherical coordinates of azimuth and elevation angles (assuming a radius of one) were first converted into Cartesian coordinates. In the new Cartesian coordinate system the origin was at the center of the subject's head. The x, y, and z axis were as labeled in Figure 2-1. Thus, the positive x axis was to listener's immediate right, the positive y axis was to the immediate

front of the listener, and the positive z axis extended directly above the listener. These Cartesian coordinates were then expressed as angles (in degrees) such that

$$\begin{aligned} x_\theta &= \sin^{-1}(x) \\ y_\theta &= \sin^{-1}(y) \\ z_\theta &= \sin^{-1}(z) \end{aligned} \tag{91}$$

where $\sin^{-1}(\bullet)$ returns the angle θ in degrees such that $-90 \le \theta \le 90$.

In this new coordinate system, x_θ indicates how far to the left or to the right of the origin a position is, albeit in degrees as opposed to units of distance. Similarly, the coordinates y_θ and z_θ indicate respectively how far in front or back and how far up or down a position is.

Each coordinate of the intended source position is then plotted versus the corresponding coordinates of the perceived or reported source position. Plotting all data points results in three scatter plots. One indicates the "left-right" correlation of the data, one indicates the "front-back" correlation, and one indicates the "up-down" correlation. Such scatter plots are shown in Figure 5-8 through Figure 5-13. In the new coordinate system, a front-back reversal will not appear in the left-right scatter plot, but will appear in the front-back plot. Similarly, an up-down reversal will not appear in the left-right plot but will appear in the up-down scatter plot.

For each subject, nine scatter plots are shown, 3 scatter plots for each model tested. For the FIR model, Figures 5-8, 5-10, and 5-12 show the scatter plots respectively for

subjects A, B, and C. Similarly, for the IP and BMR models, Figures 5-9, 5-11, and 5-13 show the scatter plots respectively for subjects A, B, and C.

Visual inspection indicates the localization performance of the two low-order models is nearly as good as the performance of the high-order FIR model for the three subjects shown. Little difference can be seen in the scatter plots of subjects A and C among the high-order FIR filter and the two low-order models. For subject B, however, who shows superior localization performance compared to the other two subjects for the high-order FIR filter, the differences are a bit more apparent. Left-right and up-down correlation plots look essentially the same, but the slight degradation in the low-order models is apparent in the front-back scatter plot.

Correlation and average angle of error (in degrees) were also computed for the data shown in each scatter plot. In terms of correlation and average error (in degrees), the three models are also in high agreement. Correlation and average error data are listed in Table 5-3. Data are the essentially the same for all three models with the following exceptions. In subject B, the front-back correlation and corresponding average angle of error were respectively 0.90 and 14°. For the IP(15,15) model the correlation was only 0.81 and the average angle of error was 17°. This was not a significant increase in average angle of error but was a noticeable decrease in correlation. Similarly, the BMR(10,10) model resulted in a front-back correlation of 0.83 and an average angle of error also of 17°.

Table 5-3 Correlation, average angle of error, and reversal results for localization data.

		A			B			C		
		L-R	F-B	U-D	L-R	F-B	U-D	L-R	F-B	L-R
Correlation	FIR	0.88	0.81	0.86	0.95	0.90	0.91	0.82	0.62	0.80
	IP	0.88	0.77	0.80	0.95	0.81	0.88	0.85	0.66	0.81
	BMR	0.85	0.82	0.85	0.95	0.83	0.89	0.83	0.62	0.82
Average Angle of Error	FIR	17	20	16	9	14	12	19	26	25
	IP	18	21	19	10	17	13	17	24	24
	BMR	19	20	17	10	17	13	18	26	22
Reversals (%)	FIR		11			7			29	
	IP		12			12			24	
	BMR		10			14			28	

The number of front-back reversals was also computed as listed in Table 5-3. Again, subjects A and C demonstrated little differences among models. In subject B, however, there was a significant increase in the percentage of reversals, increasing from 7% for the FIR model to 12% for the IP(15,15) model and to 14% for the BMR(10,10) model. These observations agree with those obtained using visual inspection.

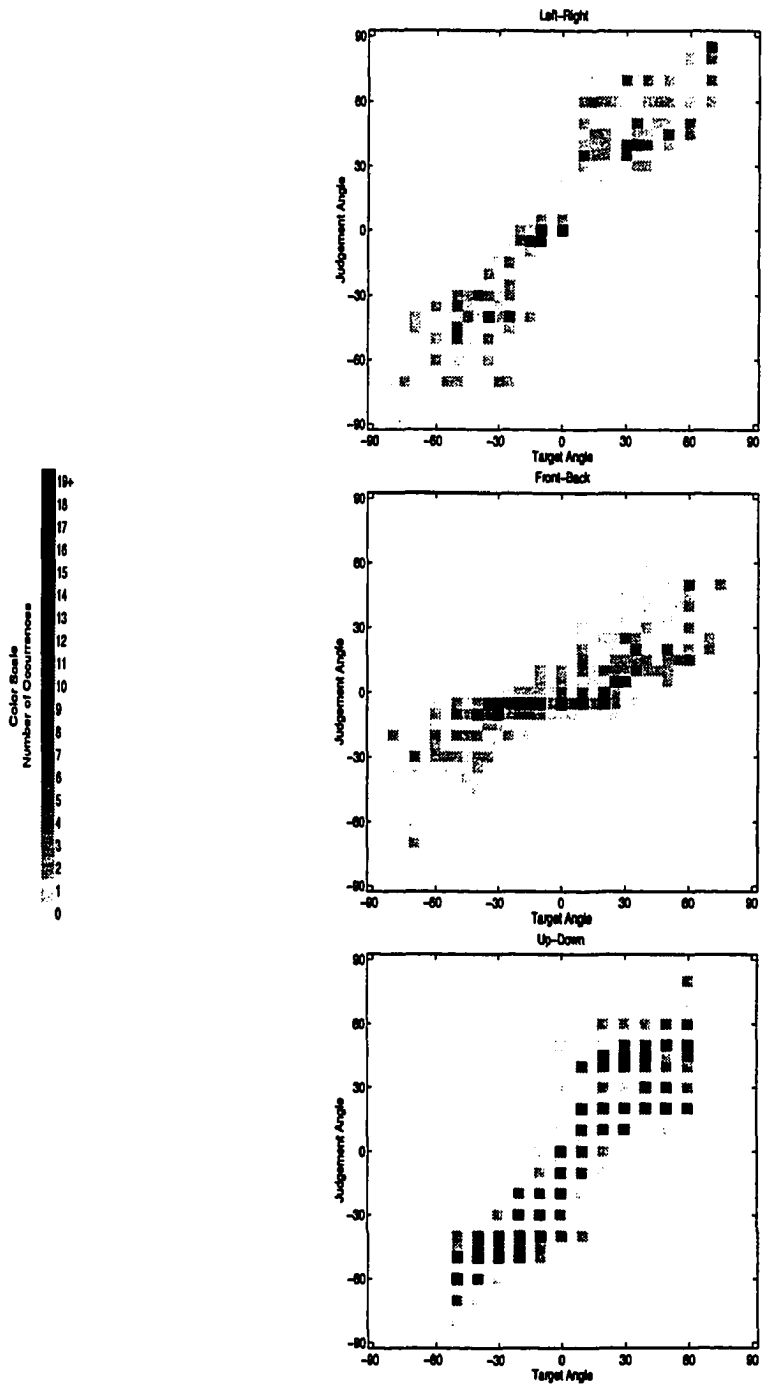


Figure 5-8 Subject A scatter plots for FIR(99,0) filter.

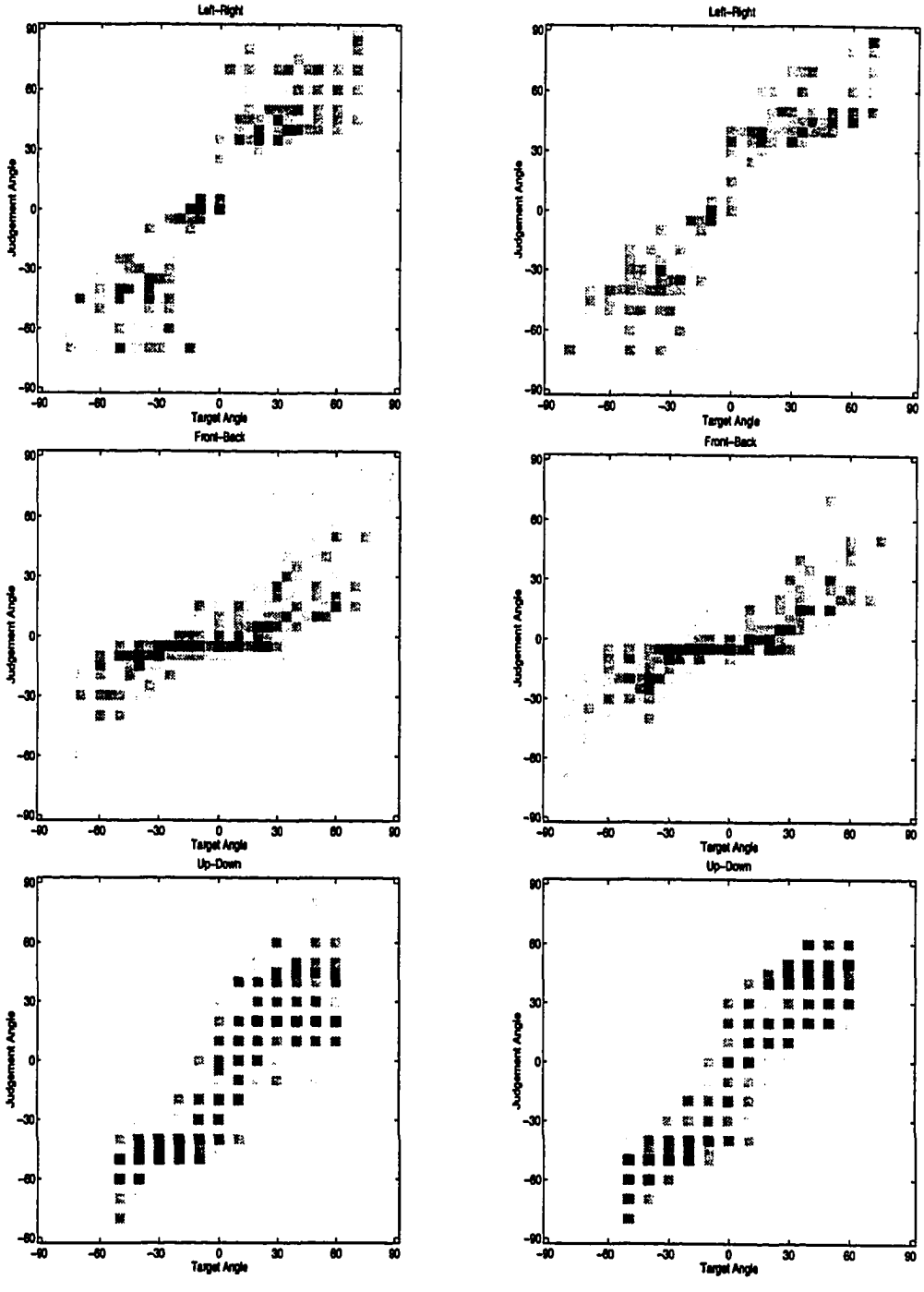


Figure 5-9 Subject A scatter plots for IP(15,15) filter (left column) and for BMR(10,10) filter (right column).

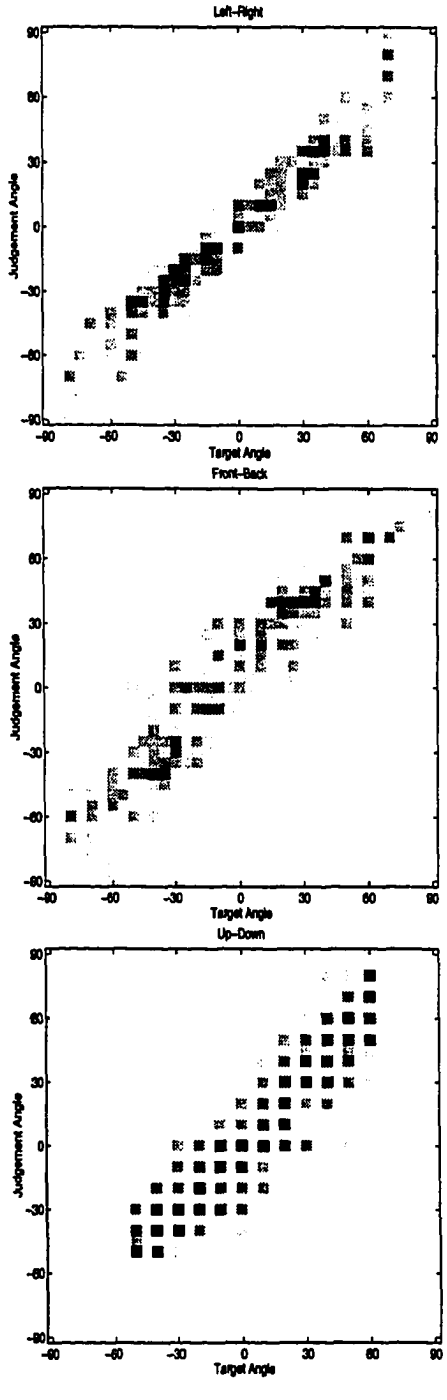


Figure 5-10 Subject B scatter plots for FIR(99,0) filter.

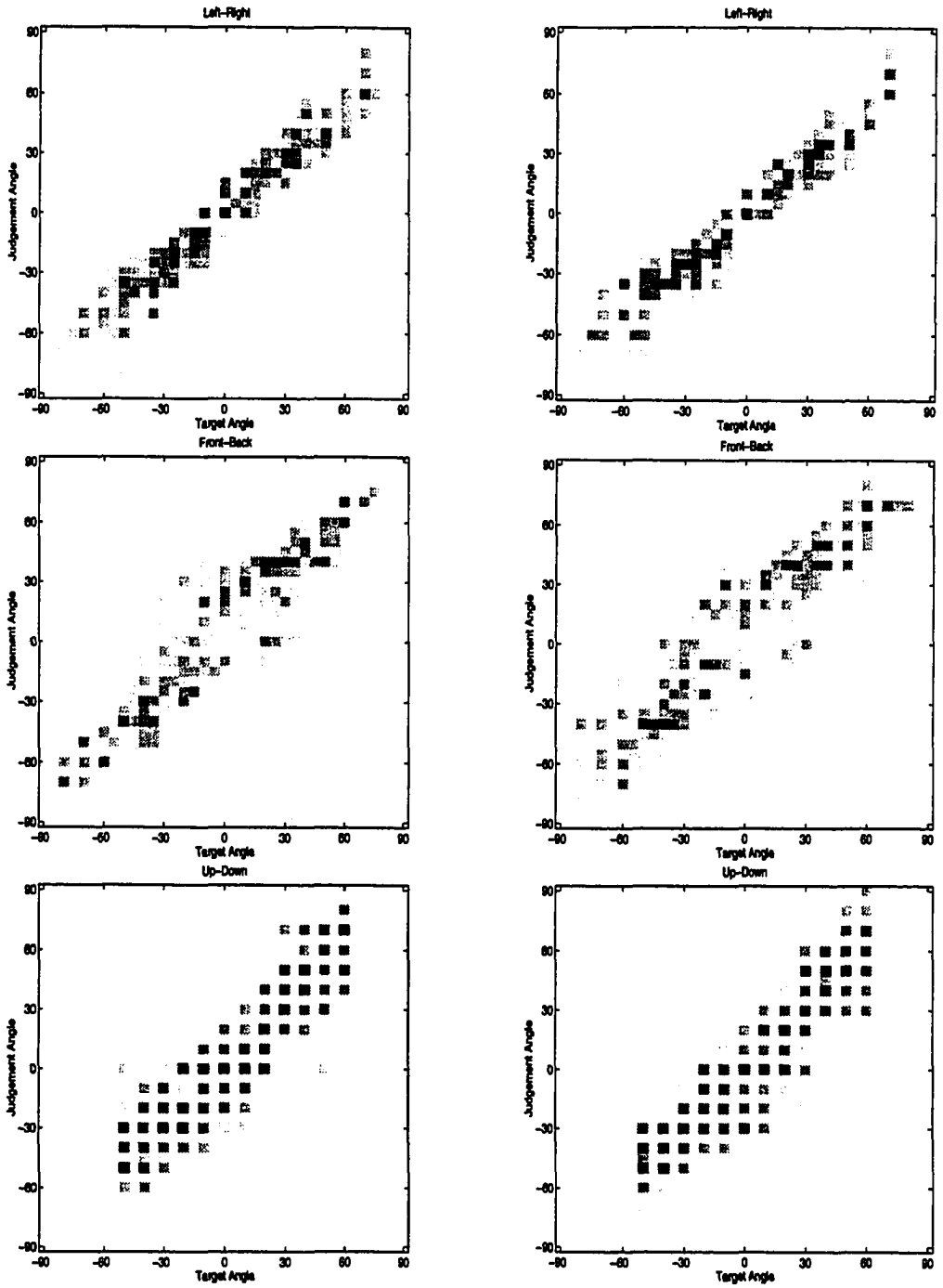


Figure 5-11 Subject B scatter plots for IP(15,15) filter (left column) and for BMR(10,10) filter (right column).

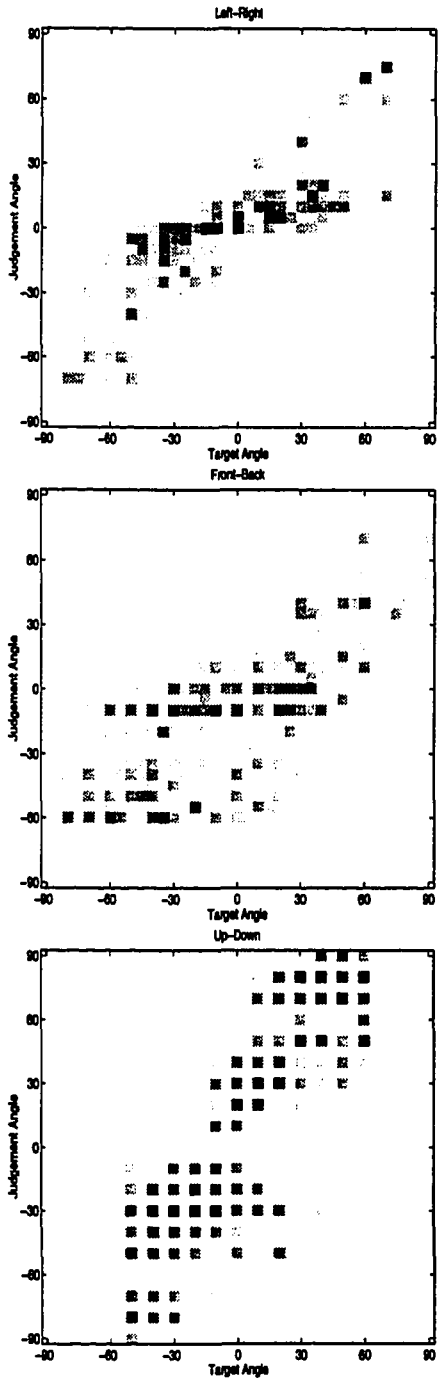


Figure 5-12 Subject C scatter plots for FIR(99,0) filter.

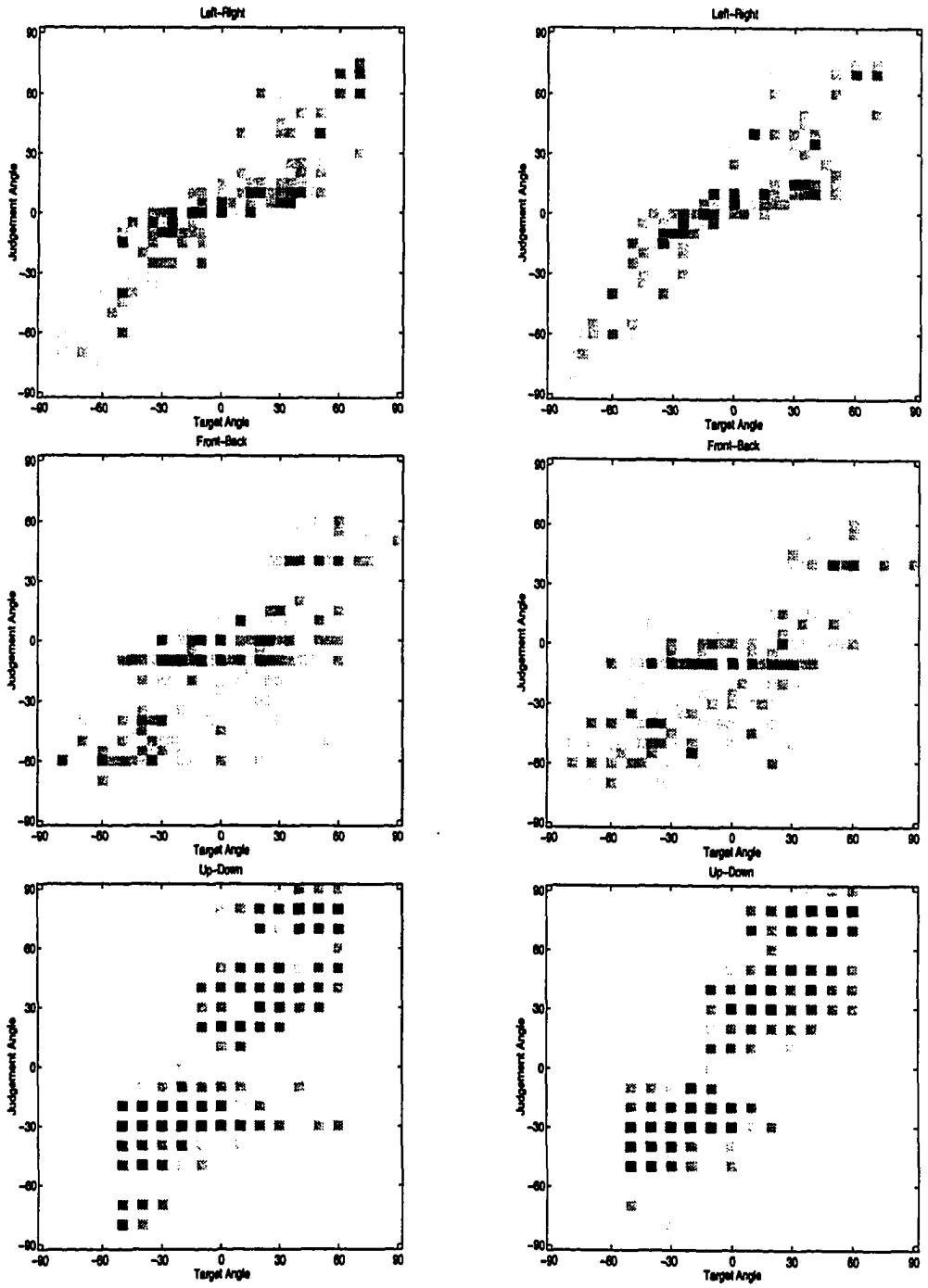


Figure 5-13 Subject C scatter plots for IP(15,15) filter (left column) and for BMR(10,10) filter (right column).

6. CONCLUSIONS

6.1 Discussion

Conventional modeling techniques, including the least-squares method, the extended Prony's method, the iterative prefiltering method, and the balanced model reduction method, were presented for estimating the parameters of the system model. A technique for estimating the order, based on minimum eigenvalues of a covariance matrix, of the system was also presented. Based on the minimum eigenvalue model order estimation technique, the head-related transfer functions were shown to be primarily autoregressive systems. However, when the diffuse field response was extracted, the resulting directional transfer functions were shown to be primarily autoregressive moving-average (pole-zero) systems. For this reason, modeling of the DTFs focused on systems with equal order numerator and denominator polynomials.

It was shown that application of conventional modeling techniques provided for a poor match between the original system response and the response of the model evaluated on an auditory frequency scale. To improve this performance, a technique referred to as critical band smoothing was presented, which roughly attempted to emulate the process of hearing which resulted in masking and critical band phenomena. A parameter of the critical band smoothing process, the smoothing factor, expressed as a fraction of a critical band, was used to regulate the amount of smoothing that was applied.

For objective evaluation of modeling techniques, the critical band distance measurement was introduced. The critical band distance was essentially the Euclidean distance of two power spectrums expressed as critical band levels in dB. The advantage of a critical band scale was that the bandwidth of each band increased with increasing frequency, thus weighting each spectrum according to an auditory frequency scale as opposed to a linear frequency scale.

If the response of the directional transfer function was smoothed prior to the system modeling process, it was shown that a better model fit of the smoothed system response could be obtained in terms of critical band distance. A critical band distance relationship was presented which related the original system response to the smoothed system response and to the response of the resulting system model. It was demonstrated that, as the smoothing factor was increased, the critical band distance between the original system response and the smoothed system response also increased. On the other hand, the critical band distance between the smoothed system response and the response of the model decreased as the smoothing factor was increased. Since, for small smoothing factors, the rate of decrease of the critical band distance between the smoothed response and the model response was greater than the rate of increase of the critical band distance between the original system response and the smoothed response, the critical band distance between original system response and the model response was generally decreasing for small smoothing factors.

The effect of critical band smoothing varied with modeling technique. For the FIR and LS models, critical band smoothing had little effect. For the BMR models, critical band smoothing had a significant effect. In general, it was observed that low-order models required

a higher smoothing factor than higher models for achieving maximum performance in terms of the critical band distance between the original system response and the model response.

For a fixed smoothing factor, increasing the model order generally lead to a linearly decreasing critical band distance for most modeling techniques presented. For the balanced model reduction method, however, the critical band distance decreased at a high rate for low model orders and decreased at a low rate for high model orders. Thus, the balanced model reduction method had its greatest advantage at low model orders.

To study the effect of modeling on localization performance, a series of localization experiments was performed using individualized transfer functions.. Three subjects, with normal hearing and previous experience in localization experiments, were selected for the tests. A comparison was made between a 100 coefficient FIR filter, a 15th order filter designed using the iterative prefiltering method, and a 10th order model designed using the balanced model reduction method. Results indicated that all models provided essentially the same performance with the following exceptions. Visual inspection of scatter plots revealed a modest degradation in front-back correlation for the low-order models. In terms of correlation and average angle of error, results were also similar for all models tested. For one subject, however, who showed a 0.90 front-back correlation, the 10th order iterative prefiltering method yielded a 0.81 front-back correlation and the 10th order balanced model reduction filter yielded a 0.83 front-back correlation. The results also indicated an increase in the number of front-back reversals, apparent in the same subject.

Overall, the results indicated that low-order modeling of the directional transfer functions is possible without significantly sacrificing perceptual performance. The primary

advantage of such low-order models was that implementations based these models could thus be constructed with reduced computation and storage requirements. One modeling technique in particular, based on critical band smoothing followed by balanced model reduction, proved to have better performance in terms of minimum critical band distance than other techniques explored. The perceptual validity of such a technique was also reinforced by the results of sound localization experiments.

6.2 Future Work

Regarding the use of the low-order modeling techniques discussed in this dissertation, a number of areas remain which are deserving of further investigation. Three such areas are briefly described below.

- **Localization experiments.** To fully assess the performance of the presented modeling techniques, a more thorough series of localization experiments would need to be performed. The localization experiments presented in this dissertation merely demonstrated the feasibility of low-order models. A more thorough evaluation of these modeling techniques would require large amounts of data collected for a large number of subjects, using a variety of model orders and smoothing factors. Results from such tests would reveal insight regarding the effects of smoothing factor on localization performance and possibly lead to a better understanding of the relationship between critical band distance and localization performance.
- **Implementation issues.** Implementation issues associated with such low-order models is also a topic of concern which needs further study. The IIR filters used in the

localization tests described in Section 5.3 were implemented using direct filter structures and single-precision floating-point coefficients. For increased filter orders, however, the high sensitivity of the direct form structure may make direct form structures impractical. The use of cascade and lattice structures could thus be employed. Informal experiments using cascade structures have not indicated sensitivity problems using the modeling techniques discussed. The cascade structure requires moderate computational effort but does, however, make interpolation difficult. Lattice structures are easier to interpolate but require the highest computational effort. Further investigation into filtering techniques which provide for ease of interpolation without significantly increasing computational effort is required.

- **Application to other areas.** The modeling techniques discussed were primarily intended for applications of virtual acoustics. However, there is potential for use of these models in other audio applications where computation and/or storage requirements are of concern. Such areas might include real-time speech coding and wideband audio compression.

ACKNOWLEDGMENTS

The author wishes to express appreciation to Dr. Tenkasi. V. Ramabadran and Dr. Satish Udpa for their support, guidance, and encouragement throughout all stages of this work.

Very special thanks are to extended to Dr. Frederic Wightman, Dr. Doris Kistler, and the rest of the staff at the Hearing Development Research Laboratory at the University of Wisconsin-Madison, not only for providing the needed laboratory resources to perform sound localization experiments, but also for their valuable contribution of ideas and insightful discussions.

The author is grateful for support from Rockwell International Corporation, without which such a project could not even have been considered.

BIBLIOGRAPHY

[1] Elizabeth M. Wenzel, Frederic L. Wightman, and Scott H. Foster, "A Virtual Display System for Conveying Three-Dimensional Acoustic Information," *Proceedings of the Human Factors Society - 32nd Annual Meeting*, 1988, pp. 86-90.

[2] Frederic L. Wightman and Doris J. Kistler, "Headphone Simulation of Free-Field Listening. I: Stimulus Synthesis," *J. Acoust. Soc. Am.*, vol. 85, no. 2, Feb. 1989, pp. 858-867.

[3] Frederic L. Wightman and Doris J. Kistler, "Headphone Simulation of Free-Field Listening. II: Psychophysical Validation," *J. Acoust. Soc. Am.*, vol. 85, no. 2, Feb. 1989, pp. 868-878.

[4] Gordon Graff, "Virtual Acoustics Puts Sound in its Place," Science, vol. 256, 1 May 1992, pp. 616-617.

[5] Srinivas Ramanthan, P. Venkat Rangan, and Harrick M. Vin, "Integrating Virtual Reality, Teleconferencing, and Entertainment Into Multimedia Home Computers," *IEEE Trans. Consumer Electronics*, vol. 38, no. 2, May 1992, pp. 86-90.

[6] T.J. Doll, J.M. Gerth, W.R. Engelman, and D.J. Folds, "Development of Simulated Directional Audio for Cockpit Applications," *USAF Report No. AAMRL-TR-86-014*, 1986.

[7] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, Cambridge, MA: MIT Press, 1983.

[8] G. Plenge, "On the Differences Between Localization and Lateralization," *J. Acoust. Soc. Am.*, vol. 56, no. 3, Sept. 1974, pp. 944-951.

[9] D. W. Batteau, R. L. Plante, R. H. Spencer, and W. E. Lyle, *Localization of Sound: Part 5*, (Report No. TP3109, Part 5), China Lake, CA, U. S. Naval Ordinance Test Station.

[10] E.A.G. Shaw, "Transformation of Sound Pressure Level From the Free Field to the Eardrum in the Horizontal Plane," *J. Acoust. Soc. Am.*, vol. 56, no. 6, Sept. 1974, pp. 1848-1861.

[11] Fred Wightman and Doris Kistler, "The Importance of Head Movements for Localizing virtual auditory objects," *Proceedings of the 1994 International Conference on Auditory Display*, Santa Fe, NM, November 1994..

[12] Scott H. Foster, Elizabeth M. Wenzel, and R. Micheal Taylor, "Real Time Synthesis of Complex Acoustic Environments."

[13] Durand R. Begault, "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems," *J. Audio Eng. Soc.*, vol. 40, no. 11, Nov. 1993, pp. 895-904.

[14] S. Meghrgardt and V. Mellert, "Transformation Characteristics of the External Human Ear," *J. Acoust. Soc. Am.*, vol. 161, no. 6, June 1977, pp. 1848-1861.

[15] M.R. Schroeder, "Synthesis of Low Peak-Factor Signals and Binary Sequences with Low Autocorrelation," *IEEE Trans. Information Theory*, vol. IT-16, Jan. 1970, pp. 85-89.

[16] Durand R. Begault, "Challenges to the Successful Implementation of 3-D Sound," *J. Audio Eng. Soc.*, vol. 39, no. 11, Nov. 1991, pp. 864-870.

[17] Jae S. Lim and Alan V. Oppenheim, *Advanced Topics in Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1988.

[18] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1989.

[19] Lennart Ljung, *System Identification: Theory for the User*, Englewood Cliffs, NJ: Prentice Hall, 1987.

[20] Thomas W. Parsons, *Voice and Speech Processing*, New York: McGraw Hill, 1987.

[21] Simon Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 1991.

[22] T. W. Parks and C. S. Burrus, *Digital Filter Design*, New York: John Wiley, 1987.

[23] K. Steiglitz and L. E. McBride, "A Technique for the Identification of Linear Systems," *IEEE Trans. Automatic Control*, vol. AC-10, Oct. 1965, pp. 461-464.

[24] K. Steiglitz, "On the Simultaneous Estimation of Poles and Zeros in Speech Analysis," *Trans. Acoustics, Speech, and Signal Processing*, vol. 25, June 1977, pp. 229-234.

[25] Lars Pernebo and Leonard M. Silverman, "Model Reduction via Balanced State Space Representations," *IEEE Trans. Automatic Control*, vol. AC-27, no. 2, April 1982, pp. 382-387.

[26] A.M. Davidson, "Balanced Systems and Model Reduction," *Electronics Letters*, vol. 22, no. 10, May 1986, pp. 531-532.

[27] Bartlomiej Beliczynski, Izzet Kale, and Gerald D. Cain, "Approximation of FIR by IIR Digital Filters: An Approach Based on Balanced Model Reduction," *IEEE Trans. Signal Processing*, vol. 40, no. 3, March 1992, pp. 532-541.

[28] B.-S. Chen, S.-C. Peng, and B.-W. Chiou, "IIR Filter Design via Optimal Hankel-Norm Approximation," *IEE Proceedings, Part G*, vol. 139, no. 5, Oct. 1992, pp. 586-590.

[29] B.-S. Chen, S.-C. Peng, and B.-W. Chiou, "Minimum Sensitivity IIR Filter Design Using Principal Component Approach," *IEE Proceedings, Part G*, vol. 138, no. 4, August 1991, pp. 474-482.

[30] Venkatappa Sreeram and Panos Agathoklis, "Design of Linear-Phase IIR Filters via Impulse-Response Grammians," *IEEE Trans. Signal Processing*, vol. 40, no. 2, Feb. 1992, pp. 389-394.

[31] Chi-Tsong Chen, *Linear System Theory and Design*, New York: Holt, Rinehart, and Winston, Inc., 1984.

[32] Alle-Jan Van Der Veen, Ed F. Deprettere, and A. Lee Swindlehurst, "Subspace-Based Signal Analysis Using Singular Value Decomposition," Proceedings of the IEEE, vol. 81, no. 9, Sept. 1993, pp. 1277-1308.

[33] Steven M. Kay, *Modern Spectral Estimation*, Englewood Cliffs, NJ: Prentice Hall, 1988.

[34] S. Lawrence Maple, *Digital Spectral Analysis with Applications*, Englewood Cliffs, NJ: Prentice Hall, 1987.

[35] Gang Liang, D. Mitchell Wilkes, and James A. Cadzow, "ARMA Model Order Estimation Based on the Eigenvalues of the Covariance Matrix," *IEEE Trans. Signal Processing*, vol. 41, no. 10, October 1993, pp. 3003-3009.

[36] James A. Cadzow and Otis M. Solomon, Jr., "Algebraic Approach to System Identification," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no. 3, June 1986, pp. 462-469.

[37] Pavel Zahorik, Fred Wightman, and Doris Kistler, "Sound Localization in Varying Virtual Acoustic Environments," *Proceedings of the 1994 International Conference on Auditory Display*, Santa Fe, NM, November 1994..

[38] Yoichi Haneda, Shoji Makino, and Yutaka Kaneda, "Common Acoustical Poles Independent of Sound Directions and Modeling of Head-Related Transfer Functions," *J. Acoust. Soc. Jpn.*, vol. 15, no. 4, 1994, pp. 277-279.

[39] D. A. Bies and C. H. Hansen, *Engineering Noise Control*, London: Unwin Hyman, 1988.

[40] John Makhoul and Lynn Cosell, "LPCW: An LPC Vocoder with Linear Predictive Spectral Warping," Presented to the ICASSP-76.

[41] Jouni Koljonen and Matti Karjalainen, "Use of Computational Psychoacoustical Models in Speech Processing: Coding and Objective Performance Evaluation," Presented to the ICASSP-84.

APPENDIX A

Minimum Eigenvalue Model Order Estimation

Noise-free case

Assume that a rational system $H(z)$ given by (2) outputs a response $y(n)$ when excited by an input $x(n)$. In the minimum eigenvalue model order estimation technique discussed in Section 3.3, a covariance matrix $\boldsymbol{R}_{p,q}$ is formed from the excitation and response data as in (55) for some undetermined model orders (p, q). It was stated in Section 3.3 that if the true order of $H(z)$ is (n_p,n_q) and if (p, q) is selected such that $p \geq n_p$ and $q \geq n_q$, then $\boldsymbol{R}_{p,q}$ will have at least one zero eigenvalue. Thus, in this situation (noise-free case) the model order can be selected as the lowest order (p, q) at which the minimum eigenvalue drops to zero.

Noise-Contaminated Input and Noise-Contaminated Output

Now consider the case in which both the system excitation $x(n)$ and the system response $y(n)$ are contaminated with a zero mean, white noise with variance σ^2. This situation can equivalently be viewed as the data matrix $\boldsymbol{D}_{p,q}$ in (54) being contaminated with a noise matrix $\Gamma_{p,q}$ in which the columns of $\Gamma_{p,q}$ are samples of a zero mean, white noise process with variance σ^2. The contaminated data matrix $\hat{\boldsymbol{D}}_{p,q}$ can thus be expressed as

$$\begin{aligned}\hat{D}_{pq} &= \left[\hat{Y}_p \vdots -\hat{X}_p\right] \\ &= D_{pq} + \Gamma_{pq}\end{aligned} \tag{A1}$$

where D_{pq} is the noise-free data matrix given by (54), $\hat{Y}_p$ is the noise-contaminated response matrix, $\hat{X}_q$ is the noise-contaminated excitation matrix, and Γ_{pq} is given by

$$\Gamma_{pq} = \begin{bmatrix} \gamma_{y(0)} & \gamma_{y(-1)} & \cdots & \gamma_{y(-p)} & \gamma_{x(0)} & \gamma_{x(-1)} & \cdots & \gamma_{x(-q)} \\ \gamma_{y(1)} & \gamma_{y(0)} & \cdots & \gamma_{y(1-p)} & \gamma_{x(1)} & \gamma_{x(0)} & \cdots & \gamma_{x(1-q)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{y(N-1)} & \gamma_{y(N-2)} & \cdots & \gamma_{y(N-p-1)} & \gamma_{x(N-1)} & \gamma_{x(N-2)} & \cdots & \gamma_{x(N-q-1)} \end{bmatrix} \tag{A2}$$

For large N, the corresponding covariance matrix can be approximated as

$$\begin{aligned}\hat{R}_{pq} &= \hat{D}_{pq}^{\mathrm{T}} \hat{D}_{pq} \\ &= \left(D_{pq} + \Gamma_{pq}\right)^{\mathrm{T}} \left(D_{pq} + \Gamma_{pq}\right) \\ &\approx R_{pq} + \sigma^2 NI\end{aligned} \tag{A3}$$

since the elements of Γ_{pq} are assumed to be uncorrelated with the elements of the data matrix D_{pq}.

By substituting (56) into (A3) and utilizing the fact that $Q^{\mathrm{T}}Q = I$, it can be shown that $\hat{R}_{pq}$ can be decomposed into the form

$$\hat{R}_{pq} \approx Q\left(\Lambda + \sigma^2 NI\right)Q^{\mathrm{T}} \tag{A4}$$

in which Q and Λ are respectively the eigenvector and eigenvalue matrices given in (56). From (A4), it is apparent that in the presence of noise the eigenvalues of R_{pq} are uniformly incremented by an amount roughly equal to $\sigma^2 N$. Thus, $\hat{R}_{pq}$ will no longer have a zero eigenvalue but will have a minimum eigenvalue approximately equal to $\sigma^2 N$.

Noise-Free Input and Noise-Contaminated Output

Now consider the more applicable case in which the system excitation $x(n)$ is known exactly, while the observed system response $\hat{y}(n)$ is a noise-contaminated response. Such a situation may occur when the system excitation is known to be an impulse, i.e., $x(n)=\delta(n)$ (discrete-time impulse). As before, it is assumed that the modeling error $e(n)$ is a zero mean, white noise process with variance σ^2.

Here the noise-contaminated data matrix can be expressed as

$$\begin{aligned}\hat{D}_{pq} &= \left[\hat{Y}_p \vdots -X_p\right] \\ &= D_{pq} + \Gamma_{pq}\end{aligned} \tag{A5}$$

where

$$\Gamma_{pq} = \begin{bmatrix} \gamma_{y(0)} & \gamma_{y(-1)} & \cdots & \gamma_{y(-p)} & 0 & 0 & \cdots & 0 \\ \gamma_{y(1)} & \gamma_{y(0)} & \cdots & \gamma_{y(1-p)} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{y(N-1)} & \gamma_{y(N-2)} & \cdots & \gamma_{y(N-p-1)} & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{A6}$$

Let the corresponding covariance matrix again be denoted $\hat{R}_{pq}$. In this case, for large N the covariance matrix $\hat{R}_{pq}$ can be approximated as

$$\begin{aligned}\hat{\boldsymbol{R}}_{pq} &= \hat{\boldsymbol{D}}_{pq}^{\mathrm{T}}\hat{\boldsymbol{D}}_{pq} \\ &= \left(\hat{\boldsymbol{D}}_{pq}+\Gamma_{pq}\right)^{\mathrm{T}}\left(\hat{\boldsymbol{D}}_{pq}+\Gamma_{pq}\right) \\ &\approx \boldsymbol{R}_{pq}+\sigma^2 N\hat{\boldsymbol{I}}_{pq}\end{aligned} \tag{A7}$$

where $\boldsymbol{R}_{pq}$ is the noise-free covariance matrix and $\hat{\boldsymbol{I}}_{pq}$ is a $(p+1+q+1)\times(p+1+q+1)$ matrix given by

$$\hat{\boldsymbol{I}}_{pq}=\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{bmatrix} \tag{A8}$$

in which $\boldsymbol{I}$ is a $(p+1)\times(p+1)$ identity matrix. It is apparent that $\hat{\boldsymbol{I}}_{pq}$ is only a "partial" identity matrix since all of the diagonal elements are not "1". Such a matrix results from the assumption that the system excitation is noise-free.

Because $\hat{\boldsymbol{I}}_{pq}$ is not a true identity matrix, it becomes difficult to predict the specific increments in the eigenvalues of $\hat{\boldsymbol{R}}_{pq}$. Although, an eigenvalue decomposition similar to (56) can be performed on $\hat{\boldsymbol{R}}_{pq}$ given by (A7), the special form of (A4) cannot be preserved.

One can, however, determine an upper bound expression for the increment in eigenvalues. Decompose the noise-contaminated covariance matrix into the form

$$\hat{\boldsymbol{R}}_{pq} = \hat{\boldsymbol{Q}}\hat{\Lambda}\hat{\boldsymbol{Q}}^{\mathrm{T}} \tag{A9}$$

In this case, the both the eigenvalues and eigenvectors have been altered compared to the noise-free case. Now rewrite (A9) as

$$\hat{\Lambda}=\hat{Q}^{T}\hat{R}_{p,q}\hat{Q} \tag{A10}$$

to get an expression for $\hat{\Lambda}$. Using the approximation given by (A3), one can expand (A10) to yield

$$\begin{aligned}\hat{\Lambda}&=\hat{Q}^{T}\left(R+\sigma^{2}N\hat{I}_{p,q}\right)\hat{Q}\\&=\hat{Q}^{T}R\hat{Q}+\sigma^{2}N\hat{Q}^{T}\hat{I}_{p,q}\hat{Q}\end{aligned} \tag{A11}$$

To determine the increment in eigenvalues, i.e., $\hat{\Lambda}-\Lambda$, define a transformation matrix P which transforms the noise-free eigenvector matrix Q into the noise-contaminated eigenvector matrix $\hat{Q}$. The desired transformation is given by

$$\hat{Q}=QP \tag{A12}$$

which directly implies $P=Q^{T}\hat{Q}$. It is important to note the fact that

$$\begin{aligned}P^{T}P&=\left(Q^{T}\hat{Q}\right)^{T}\left(Q^{T}\hat{Q}\right)\\&=\hat{Q}^{T}QQ^{T}\hat{Q}\\&=I\end{aligned} \tag{A13}$$

where I is the identity matrix.

Substituting (A12) into (A11) yields

$$\begin{aligned}\hat{\Lambda}&=(QP)^{T}R(QP)+\sigma^{2}N\hat{Q}^{T}\hat{I}_{p,q}\hat{Q}\\&=P^{T}(Q^{T}RQ)P+\sigma^{2}N\hat{Q}^{T}\hat{I}_{p,q}\hat{Q}\\&=P^{T}\Lambda P+\sigma^{2}N\hat{Q}^{T}\hat{I}_{p,q}\hat{Q}\end{aligned} \tag{A14}$$

Note that although neither of the matrix expressions on the right side of (A14) necessarily represents a diagonal matrix, the summation of the two must be diagonal since the eigenvalue matrix $\hat{\Lambda}$ on the left side *is* diagonal.

One can now derive an upper bound expression for the increment in eigenvalues due to the presence of noise. Equating the diagonal parts of (A14) and applying (A13) results in

$$\begin{aligned}\hat{\Lambda}&=diag(P^{T}\Lambda P)+diag(\sigma^{2}N\hat{Q}^{T}\hat{I}_{p,q}\hat{Q})\\&=\Lambda+\sigma^{2}N\cdot diag(\hat{Q}^{T}\hat{I}_{p,q}\hat{Q})\\&=\Lambda+\sigma^{2}N\Psi\end{aligned} \tag{A15}$$

where $diag(\bullet)$ represents the extraction of the main diagonal, and the diagonal matrix Ψ contains the "partial inner product" of each column of $\hat{Q}$. That is, the diagonal elements of the matrix Ψ can be expressed and bounded as

$$\begin{aligned}\psi_i&=\sum_{j=0}^{p}\hat{q}_{ij}^{2}\\&\le\sum_{j=0}^{p+q+1}\hat{q}_{ij}^{2}=\hat{q}_i^{T}\hat{q}_i=1\end{aligned} \tag{A16}$$

where Ψ_i is the ith diagonal element of Ψ, $\hat{q}_i$ is the ith eigenvector of $\hat{Q}$, and $\hat{q}_{ij}$ is the jth element of $\hat{q}_i$. Substituting (A16) into (A15) leads to the upper bound expression for eigenvalue increment given by

$$\hat{\Lambda} - \Lambda \leq \sigma^2 NI \tag{A17}$$

where I is the identity matrix. This expression relates the modeling error variance to the minimum eigenvalue of the corresponding covariance matrix.

APPENDIX B

Design of IIR Filters Using Model Reduction Techniques

The state-space model familiar to control systems theory is the basis for many model reduction techniques [32,33]. An introduction to state-space models is briefly presented, and the formulation of balanced model reduction techniques follows.

State-space realizations

The following state-space model can be used to describe any discrete-time, linear, time-invariant system:

$$\begin{aligned} x(k+1) &= Ax(k)+Bu(k) \\ y(k) &= Cx(k)+Du(k) \end{aligned} \tag{B1}$$

where $x(k)$ is the r-dimensional state vector, A is an $r\times r$ matrix, B and C^T are $r\times 1$ vectors, and D is a scalar. These equations are termed the *state-equations* of the system. The dimension r is called the *order of the system.* The state equations dictate a mapping from the input $u(k)$ and the current state $x(k)$ to the next state $x(k+1)$, and to the output $y(n)$. The transfer function for such a representation is given by

$$H(z)=\frac{Y(z)}{U(z)}=C(zI-A)^{-1}B+D \tag{B2}$$

and the corresponding impulse response is given by

$$\begin{aligned} h &= \left[\cdots\ 0\ h_0\ h_1\ h_2\ h_3\ \cdots\right]^T \\ &= \left[\cdots\ 0\ D\ CB\ CAB\ CA^2B\ \cdots\right]^T \end{aligned} \tag{B3}$$

The state-space realization problem becomes one of determining a state-space model representation given the transfer function of the system. There exists an infinite number of state-space realizations for a given transfer function. For any realization, one can define an observability matrix and a controllability matrix as [31]:

$$S_O = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{r-1} \\ \vdots \end{bmatrix} \qquad S_C = \left[B\ AB\ A^2B\ \cdots\ A^{r-1}B\ \cdots\right] \tag{B4}$$

The concept of controllability describes the ability of an input to cause a change in the state of the system. Similarly, the concept of observability describes the possibility of estimating the state of a system based on observations of the output. Two related expressions, the controllability grammian matrix and the observability grammian matrix [27,29], can be defined as

$$\begin{aligned} W_C &= S_C S_C^T \\ W_O &= S_O^T S_O \end{aligned} \tag{B5}$$

A system is considered to be *controllable* if its controllability matrix has rank r, or, equivalently, if its controllability grammian matrix is nonsingular [31]. Likewise, a system is

considered to be *observable* if its observability matrix is of rank r or its observability grammian matrix is nonsingular [31]. A realization of a system which is both observable and controllable is called a *minimal* realization [31]. A minimal realization has a system order less than or equal to any other realization. Note, however, that minimal realizations are not unique. In applications considered here, one will only be concerned with obtaining minimal system realizations, since minimal realizations will lead to implementations with the lowest computation.

One will also be concerned with finding realizations which are *balanced.* A system is said to be balanced if its controllability and observability grammians are equal [25]. Balanced realizations are useful in model reduction applications because balanced systems can be reduced by truncation of states. One way to obtain a balanced realization is through singular value decomposition of the *Hankel matrix.*

Singular value decomposition of the Hankel matrix

The Hankel matrix is defined [32] as

$$\Phi(H) = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & \ddots & \\ h_3 & \ddots & & \\ \vdots & & & \end{bmatrix} \tag{B6}$$

where h_n is the impulse response of the system. Thus, given the impulse response of a system, or equivalently, the coefficients of a high-order FIR filter, the Hankel matrix can be constructed. Note that the Hankel matrix is upper-diagonal and constant along its

antidiagonals. The rank of the Hankel matrix is equal to the minimal order of the system. Given an *N*th order FIR filter, the Hankel matrix will be finitely dimensioned.

Singular value decomposition (SVD) is a very powerful tool for the analysis of signals and systems. Using singular value decomposition, an $N \times N$ Hankel matrix can be decomposed into the following form [32]

$$\Phi(H) = U \Sigma V^{T} \tag{B7}$$

where U and V are unitary matrices, and Σ is a diagonal matrix:

$$\Sigma = diag(\sigma_1, \sigma_2, \sigma_3, \cdots, \sigma_r, \sigma_{r+1}, \cdots, \sigma_N) \tag{B8}$$

in which, typically, the diagonal entries are arranged in decreasing order, i.e., $\sigma_1 > \sigma_2 > \ldots > \sigma_r > \sigma_{r+1} > \ldots > \sigma_N$. The diagonal entries of Σ are called the *Hankel singular values*. For a rank r Hankel matrix, the singular values $\sigma_1 \ldots \sigma_r$ will be nonzero, while the singular values $\sigma_{r+1} \ldots \sigma_N$ will equal zero. An important result of this decomposition is the fact that the Hankel matrix can be expressed as a linear combination of rank-1 matrices

$$\Phi(H) = \sum_{i=1}^{N} \sigma_i (u_i v_i^T) = \sum_{i=1}^{r} \sigma_i (u_i v_i^T) \tag{B9}$$

where u_i represents the ith column of U and v_i represents the ith column of V. For a rank-r Hankel matrix, only the first r terms in the summations will contribute to $\Phi(H)$. It becomes clear that the first r columns of U span the columns space of $\Phi(H)$. These vectors are called

the *left singular vectors*. Similarly, the first r columns of V span the row space of $\Phi(H)$ and are called the *right singular vectors*.

In the presence of noise the Hankel matrix will become full rank. The addition of white, Guassian noise, will cause an increase in the singular values proportional to the noise variance [32]. This will make it more difficult to estimate the order of the system, because all diagonal elements will be nonzero.

The question remains as to how a balanced system is realized using the Hankel matrix. It has been shown that the Hankel matrix can be formed as the product of the observability matrix and the controllability matrix

$$\Phi(H) = S_o S_c \tag{B10}$$

This can easily be verified by substituting (B4) and (B6) into (B10). If a system transformation is performed on system (A, B, C, D) using a nonsingular transformation matrix T, the Hankel matrix of the transformed system remains unchanged. This can be seen from (B10)

$$\begin{aligned} \Phi(H) &= S_o S_c \\ &= S_o T T^{-1} S_c \\ &= S_o' S_c' \end{aligned} \tag{B11}$$

where

$$S_o' = S_o T \tag{B12}$$

is the observability matrix of the transformed system and

$$S_c' = T^{-1} S_c \tag{B13}$$

is the controllability matrix of the transformed system.

To obtain a balanced realization, one must determine a transformation matrix T which transforms an unbalanced system into a balanced one. If one first performs a singular value decomposition of the Hankel matrix, and then sets the following equalities

$$\begin{aligned} S_o' &= U\Sigma^{1/2} \\ S_c' &= \Sigma^{1/2} V^{\mathrm{T}} \end{aligned} \tag{B14}$$

a balanced realization will result. This can be seen by computing the controllability and observability grammian matrices as

$$\begin{aligned} W_c &= S_c' S_c'^{\mathrm{T}} = (U\Sigma^{1/2})(U\Sigma^{1/2})^{\mathrm{T}} = \Sigma UU^{\mathrm{T}} = \Sigma \\ W_o &= S_o' S_o'^{\mathrm{T}} = (\Sigma^{1/2} V^{\mathrm{T}})(\Sigma^{1/2} V^{\mathrm{T}})^{\mathrm{T}} = \Sigma^{1/2} VV^{\mathrm{T}} \Sigma^{1/2} = \Sigma \end{aligned} \tag{B15}$$

The resulting controllability and observability Grammians are equal. Comparing (B12) and (B13) leads to the following equations

$$T^{-1} S_c = \Sigma^{1/2} V^{\mathrm{T}} \tag{B16}$$

$$S_o T = U\Sigma^{1/2} \tag{B17}$$

Solving (B16) for T yields

$$\begin{aligned} T^{-1}S_c &= \Sigma^{1/2}V^{\mathrm{T}} \\ S_c &= T\Sigma^{1/2}V^{\mathrm{T}} \\ S_c(\Sigma^{1/2}V^{\mathrm{T}})^{-1} &= T \\ S_o^{-1}S_oS_c(\Sigma^{1/2}V^{\mathrm{T}})^{-1} &= T \\ S_o^{-1}\Phi(H)(\Sigma^{1/2}V^{\mathrm{T}})^{-1} &= T \\ S_o^{-1}(U\Sigma V^{\mathrm{T}})(\Sigma^{1/2}V^{\mathrm{T}})^{-1} &= T \\ S_o^{-1}(U\Sigma^{1/2})(\Sigma^{1/2}V^{\mathrm{T}})(\Sigma^{1/2}V^{\mathrm{T}})^{-1} &= T \\ S_o^{-1}U\Sigma^{1/2} &= T \end{aligned} \qquad \text{(B18)}$$

Similarly, solving (B17) yields

$$\begin{aligned} S_oT &= U\Sigma^{1/2} \\ T &= S_o^{-1}U\Sigma^{1/2} \end{aligned} \qquad \text{(B19)}$$

The desired transformation matrix T is $S_o^{-1}U\Sigma^{1/2}$. Thus, by performing a state transformation based on singular value decomposition of the Hankel matrix a balanced realization can be found. As will become apparent in the next subsection, a balanced realization is useful because it provides for a convenient method of model reduction by simply truncating the least-significant states.

Model reduction

In model reduction, one wishes to find a model of order $q < r$ such that the qth-order model lies closest to the original model according to some distance measurement. Define the Frobenius norm of an $M \times N$ matrix X as follows [32]

$$\|X\|_F = \left[\sum_{i=1}^{M}\sum_{j=1}^{N}|x_{ij}|^2\right]^{1/2} \tag{B20}$$

Also define the Frobenius distance between two $M \times N$ matrices X and Y as

$$d_F(X,Y) = \left[\sum_{i=1}^{M}\sum_{j=1}^{N}|x_{ij}-y_{ij}|^2\right]^{1/2} = \|X-Y\|_F \tag{B21}$$

It can be seen that the Frobenius distance between two matrices is the same as the Frobenius norm of the difference of the two matrices. Now consider the following rank q approximation to the Hankel matrix

$$\hat{\Phi}(H) = \sum_{i=1}^{q} \sigma_i (u_i v_i^T) \tag{B22}$$

It has been shown [32] that this approximation is the optimal approximation to $\Phi(H)$ in the minimum Frobenius distance sense. Using a balanced state-space realization (A,B,C,D), partition Σ according to the q largest singular values:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \tag{B23}$$

where

$$\begin{aligned} \Sigma_1 &= \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_q) \\ \Sigma_2 &= \mathrm{diag}(\sigma_{q+1}, \sigma_2, \cdots, \sigma_n) \end{aligned} \tag{B24}$$

and partition the state-space matrices accordingly:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad C = [C_1 \; C_2] \tag{B25}$$

If the system (A, B, C, D) is asymptotically stable and balanced, the truncated system (A_{11}, B_1, C_1, D) will be asymptotically stable, controllable, and observable [25]. The transfer function corresponding to the system (A_{11}, B_1, C_1, D) can be found by using (B2). Thus, using balanced model reduction an IIR approximation to an FIR filter can been found. The method just described is a rather generic balanced model reduction procedure. Variations of this procedure can be found in the literature [25-30].

APPENDIX C

Table C-1 Average critical band distance, FIR filtering method.

		Smoothing Factor					
		0.0	0.1	0.2	0.3	0.4	0.5
Model Order	16	6.9	6.5	6.5	6.5	6.5	6.5
	18	6.7	6.4	6.3	6.3	6.3	6.3
	20	6.4	6.1	6.0	6.0	6.0	6.0
	22	6.1	5.8	5.7	5.8	5.8	5.8
	24	5.9	5.6	5.5	5.5	5.5	5.6
	26	5.6	5.3	5.2	5.3	5.3	5.4
	28	5.3	5.0	5.0	5.0	5.1	5.2
	30	5.1	4.8	4.8	4.8	4.9	5.0
	32	4.9	4.6	4.5	4.6	4.7	4.8
	34	4.7	4.4	4.4	4.4	4.5	4.6
	36	4.6	4.2	4.2	4.3	4.4	4.5
	38	4.4	4.1	4.1	4.2	4.3	4.4
	40	4.3	4.1	4.1	4.1	4.2	4.4

Table C-2 Average critical band distance, least-squares method.

		Smoothing Factor					
		0.0	0.1	0.2	0.3	0.4	0.5
Model Order	8	7.0	6.6	6.5	6.4	6.5	6.5
	9	6.7	6.6	6.6	6.5	6.4	6.3
	10	6.6	6.4	6.2	6.1	6.0	6.0
	11	6.3	6.1	5.9	5.8	5.7	5.7
	12	6.0	5.8	5.7	5.7	5.6	5.7
	13	5.9	5.7	5.6	5.6	5.6	5.7
	14	5.7	5.4	5.4	5.3	5.4	5.4
	15	5.5	5.4	5.3	5.1	5.1	5.1
	16	5.3	5.2	5.0	5.0	4.9	5.0
	17	5.2	4.9	4.8	4.8	4.8	4.9
	18	5.0	4.8	4.7	4.6	4.7	4.8
	19	4.9	4.7	4.6	4.4	4.5	4.6
	20	4.7	4.4	4.3	4.1	4.2	4.5

Table C-3 Average critical band distance, least-squares weighted error method.

		Smoothing Factor					
		0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
	8	6.6	6.3	6.2	6.2	6.2	6.2
	9	6.5	6.2	6.0	5.9	6.0	6.0
	10	6.2	5.9	5.8	5.7	5.8	5.9
	11	5.8	5.5	5.5	5.5	5.5	5.6
	12	5.6	5.4	5.2	5.2	5.2	5.3
Model	**13**	5.5	5.2	5.1	5.0	5.0	5.1
Order	**14**	5.1	4.8	4.7	4.7	4.8	4.9
	15	4.8	4.5	4.4	4.5	4.6	4.8
	16	4.6	4.4	4.3	4.4	4.5	4.7
	17	4.4	4.3	4.2	4.3	4.4	4.6
	18	4.3	4.2	4.1	4.2	4.3	4.5
	19	4.2	4.0	4.0	4.0	4.1	4.4
	20	3.9	3.8	3.8	3.8	4.0	4.2

Table C-4 Average critical band distance, iterative prefiltering method.

		Smoothing Factor					
		0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
	8	6.4	6.0	6.0	5.9	5.9	5.9
	9	6.3	6.0	5.8	5.7	5.8	5.8
	10	6.1	5.7	5.6	5.6	5.6	5.7
	11	5.8	5.5	5.3	5.3	5.3	5.3
	12	5.6	5.4	5.1	5.2	5.1	5.2
Model	**13**	5.4	5.3	5.1	5.0	4.9	5.0
Order	**14**	5.3	5.0	4.8	4.7	4.7	4.8
	15	5.2	4.7	4.5	4.3	4.4	4.6
	16	4.8	4.4	4.3	4.3	4.4	4.6
	17	4.6	4.2	4.1	4.2	4.3	4.5
	18	4.4	4.1	4.0	4.1	4.1	4.4
	19	4.2	4.0	4.0	3.9	4.1	4.2
	20	4.2	3.9	3.8	3.7	3.9	4.2

Table C-5 Average critical band distance, weighted iterative prefiltering method.

		Smoothing Factor					
		0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
	8	6.6	6.3	6.1	6.0	5.9	5.9
	9	6.2	6.0	5.7	5.6	5.6	5.7
	10	5.9	5.6	5.4	5.3	5.3	5.4
	11	5.6	5.3	5.1	5.1	5.1	5.2
	12	5.4	5.2	5.0	5.1	4.9	4.9
Model	**13**	5.2	4.9	4.7	4.6	4.6	4.7
Order	**14**	5.1	4.6	4.4	4.3	4.3	4.5
	15	4.8	4.3	4.1	4.1	4.2	4.4
	16	4.4	4.0	4.0	4.0	4.1	4.4
	17	4.3	3.9	3.8	3.8	4.0	4.2
	18	4.1	3.8	3.7	3.8	3.9	4.1
	19	3.9	3.8	3.6	3.6	3.8	4.1
	20	3.9	3.5	3.4	3.5	3.6	3.9

Table C-6 Average critical band distance, pole-zero cancellation method.

		Smoothing Factor					
		0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
	8	11.7	12.5	13.1	12.1	11.7	11.0
	9	9.2	10.0	10.4	9.5	8.5	8.2
	10	9.2	10.0	10.4	9.5	8.5	8.2
	11	7.2	7.9	8.3	7.4	6.9	6.4
	12	7.2	7.9	8.3	7.4	6.9	6.4
Model	**13**	5.8	6.5	6.8	6.0	5.6	5.1
Order	**14**	5.8	6.5	6.8	6.0	5.6	5.1
	15	5.0	5.1	5.4	5.0	4.6	4.2
	16	5.0	5.1	5.4	5.0	4.6	4.2
	17	4.2	4.3	4.4	4.1	3.8	3.8
	18	4.2	4.3	4.4	4.1	3.8	3.8
	19	3.7	3.7	3.7	3.6	3.4	3.7
	20	3.7	3.7	3.7	3.6	3.4	3.7

Table C-7 Average critical band distance, balanced model reduction method.

		Smoothing Factor					
		0.0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
	8	6.4	5.5	5.3	4.9	4.8	4.8
	9	5.7	4.9	4.4	4.2	4.1	4.3
	10	5.1	4.3	3.9	3.9	3.8	4.0
	11	4.5	3.8	3.4	3.4	3.4	3.6
	12	4.1	3.5	3.2	3.1	3.2	3.5
Model	**13**	3.8	3.2	3.0	2.9	3.1	3.4
Order	**14**	3.5	3.1	2.8	2.9	3.1	3.4
	15	3.2	2.9	2.7	2.8	2.9	3.3
	16	3.2	2.7	2.6	2.7	2.9	3.3
	17	3.1	2.6	2.5	2.7	2.9	3.3
	18	2.9	2.5	2.5	2.6	2.9	3.3
	19	2.8	2.4	2.4	2.6	2.9	3.2
	20	2.8	2.4	2.4	2.6	2.8	3.2