Review of Digital Filter Design and Implementation Methods for 3-D Sound

Jyri Huopaniemi and Matti Karjalainen

*Helsinki University of Technology, Laboratory of Acoustics and Audio
Signal Processing, Otakaari 5A, FIN–02150 Espoo, Finland*

`Jyri.Huopaniemi@hut.fi, Matti.Karjalainen@hut.fi`
`http://www.hut.fi/HUT/Acoustics/`

ABSTRACT

In this paper, we discuss methods for digital filter design with application to 3-D sound. A review of existing filter design methods for binaural and transaural processing is presented. New methods that take into account the non-uniform frequency resolution of the human ear are explored. Listening tests have been performed to determine the subjective preference of different designs.

0 INTRODUCTION

Measurements and models of head-related impulse responses (HRIR) and the corresponding frequency domain transfer functions (head-related transfer function, HRTF) of human subjects or dummy heads are the source of information for the research of spatial hearing and applications of binaural technology (see, e.g., [1] - [5] for fundamentals on these subjects). These transfer functions evaluated at discrete azimuth and elevation angles are sufficient for the synthesis of realistic three-dimensional sound events for headphone or loudspeaker listening. One of the problems in 3-D sound synthesis, however, is the computational load of accurate HRTF approximation. To overcome this, both computationally efficient and perceptually relevant digital models of HRTFs have to be created.

3-D sound system design can be divided into two cases according to the reproduction method: 1) *binaural processing* for headphone listening, and 2) *transaural processing* for loudspeaker (2 speakers) listening [3]. The significant difference in these methods is the crosstalk that is introduced in loudspeaker listening and that has to be canceled in transaural synthesis.

In this work, different methods for digital filter design with application to 3-D sound are discussed. A short overview of HRTF filter design can be

1

found, e.g., in a book by Begault ([1], pp. 158-163). Traditionally, HRTFs have been modeled with finite impulse response (FIR) filters based on minimum-phase reconstruction and windowing in the time domain [6] [7]. Some more advanced techniques have been discussed in [8]. Recursive IIR filter design methods have also been presented [9]-[14] but the field has not been thoroughly explored. We have made comparisons of different HRTF filter design methods and performed listening tests to verify our results. New methods have been explored that take into account the non-uniform frequency resolution of the human ear [15]. These *warped filters* have the property of focusing attention on the lower frequency range where the ear is at its most selective.

In transaural filter design, simplified models of HRTFs have been used to overcome some of the known problems such as a limited listening area, the "sweet spot" [16]. Generally, the use of IIR filters in transaural systems may be motivated due to the recursive nature of cross-talk canceling [13].

This paper is organized as follows. In Chapter 1, general properties of HRTFs such as amplitude and phase features and equalization strategies are overviewed. Filter design issues for binaural and transaural systems are discussed in Chapter 2. Filter implementation issues are considered in Chapter 3. In Chapter 4, listening tests for binaural filter design that were performed during the scope of this study are discussed. Finally, in Chapter 5, conclusions are drawn and directions are given for future work.

1 HRTF MODELING

The synthesis of binaural or transaural signals can be accomplished based on two approaches: the computational, and the empirical approach [17]. The empirical approach uses HRTF data obtained from measurements on dummy heads or real persons. Approximations for HRTFs can also be calculated by analytical means, using computer models that resemble wave propagation and diffraction around a sphere or a replica of a human head [18] [19]. The computational approach is applicable, e.g., in the design of cross-talk canceling filters for transaural processing [16].

1.1 HRTF Properties

HRTFs are the output of a linear and time-invariant system, that is, the diffraction and reflections of the human head, the outer ear, and the torso. Thus the impulse responses can directly be represented as FIR filters. There are often computational constraints that lead to the need of HRTF approximation. This can be carried out using conventional digital filter design techniques. It is, however, necessary to note that the filter design problem is not a straightforward one. We should be able to design arbitrary-shaped mixed-phase filters that meet the set criteria both in the amplitude and phase response. The main

questions of interest that the filter design expert is faced with are now: What is important in HRTF modeling? Are there constraints in the amplitude and phase response and if so, how are they distributed over the frequency range of hearing?

1.2 Amplitude Properties of HRTFs

The major cues of human spatial hearing contained in HRTFs are the interaural time differences (ITD) and the interaural amplitude or level differences (IAD, ILD) between the two ears. The IAD cues have a dominant role in localization in the frequency range above 1.5 kHz [2]. Furthermore, the highly idiosyncratic spectral (amplitude) high-frequency cues of the HRTFs contribute to localization in the median plane and in the cone of confusion, where the interaural cues are ambiguous.

1.3 Phase Properties of HRTFs

The ITD is the major cue of human sound localization at low frequencies, below 1.5 kHz, where the head dimensions are large compared to the wavelength of sound. An attractive property of HRTFs is that they are nearly of minimum phase [20]. The excess phase that is the result of subtracting the original phase response from its minimum-phase counterpart has been found to be approximately linear. This suggests that the excess phase can be separately implemented as an allpass filter or a simple delay line. In the case of binaural synthesis, the interaural time delay (ITD) part of the two HRTFs may be modeled as a separate delay line, and minimum-phase HRTFs may be used for synthesis. Research has been carried out in this area and it can be concluded that minimum-phase reconstruction does not have any perceptual consequences [6] [21]. This information is crucial in the design and implementation of digital filters for 3-D sound.

1.4 Individual Differences

Fully satisfactory binaural or transaural synthesis can only be achieved using individual HRTFs [22] [23]. Generally, it is desired to create a database of HRTFs that would work for a large population of listeners. This can be achieved, e.g., by selecting a typical human subject using subjective listening tests [25]. The subjective HRTF quality of different dummy heads has been found inferior to that of human subjects [24], but in many cases, when real-head HRTF measurements are not at hand, compromises have to be made.

1.5 HRTF Equalization

The sound transmission in an HRTF measurement includes characteristics of many subsystems that are to be compensated in order to achieve the desired re-

sponse. The transfer functions of the driving loudspeaker, the microphone and the ear canal (if the measurement position were inside an open ear canal) may thus have to be equalized. If, however, a more general database of HRTFs is desired, we should consider other equalization strategies like free-field equalization or diffuse-field equalization [2] [3]. Free-field equalization is achieved by frequency-domain division (deconvolution) of the measured HRTF by a reference measured in the same ear from a certain direction (typically chosen as 0° azimuth and 0° elevation). In diffuse-field equalization, a reference spectrum is derived by power-averaging all HRTFs from each ear and taking the square root of this average spectrum. Diffuse-field equalized HRTFs are obtained by deconvolving the original by the diffuse-field reference HRTF of that ear. This leads to the fact that the factors that are not incident-angle dependent, such as the ear canal resonance, are removed.

In many cases further pre-processing of the measured HRTF data is required before filter design. An attractive approach for HRTF smoothing is to apply a variable-size window function to the power spectrum to approximate, for example, the critical-band resolution of the human ear [13] [26]. This smoothing applies only to the magnitude response, so it is assumed that the phase can be calculated by minimum-phase reconstruction.

2 FILTER DESIGN FOR 3-D SOUND

In this chapter, an overview of existing and new filter design methods for binaural and transaural synthesis will be given. An illustration of the differences in binaural and transaural synthesis is shown in Figure 1 (based on [16] and [13]).

In the case of binaural filter design, the HRTF measurement may directly be approximated by various filter design methods provided that proper equalization is carried out. In the example of Fig. 1, the monophonic time-domain signal x_m (short for $x_m(n)$) is filtered with two HRTF filter approximations $H_l(z)$ and $H_r(z)$ to create a single virtual source. Advantages of binaural processing are that the listening facilities and positions are not critical. On the other hand, individual HRTFs must be used and care must be taken in the equalization and placing of headphones in order to obtain an immersive 3-D sound scape.

In transaural synthesis (see Fig. 1), when loudspeaker listening is desired and signals $\hat{y}_l$ and $\hat{y}_r$ (processed binaural signals) are driven from the speakers, the direction-dependent loudspeaker-to-ear transfer functions $H_i(z)$ and $H_c(z)$ (symmetrical listening position) have to be taken into account in order to obtain a similar effect than in headphone listening. This calls for cross-talk canceling. This can be seen as a cascaded process, where HRTF filters are designed and implemented separately from the cross-talk canceling filters. An-

4

other alternative is to combine these processes and design transaural filters by using, e.g., shuffler structures [16]. In Figure 2, digital filter structures for converting mono- and stereophonic material into binaural and transaural signals are presented.

A list representing research carried out in the field of HRTF approximation by various authors is illustrated in Table 1. The filter order corresponds to the FIR tap size and to the number of poles and zeros in the IIR case (in most cases an equal amount of poles and zeros have been used). One can see from the table that the results from different studies vary considerably from one to another. There are many causes to this. Some of the studies are purely theoretical meaning that the results are formulated in the form of a spectral error measure, or by visual inspection. In some of the references, the authors claim that a certain filter order appeared to be satisfactory in informal listening tests. These cases are marked in the table with a question mark. There have been very few formal listening tests in this field that also give statistically reliable results. Another question is the validity of the HRTF data used in the studies. Whether equalized for free-field conditions or a certain headphone type, whether dummy-head or individual/nonindividual real-head data was used, whether minimum-phase reconstruction was applied, all these aspects may cause the large deviation seen in the results of Table 1.

In the following, methods presented in the literature and their validity are discussed. Furthermore, a framework for auditory-based HRTF filter design is outlined. Comparison of different filter design methods is performed in Chapter 2.5. These results are used in the listening experiment, which is described in Chapter 5.

2.1 Structural Analysis

Interest in functional representations of HRTFs has risen over the past years in search of efficient auralization techniques. These methods resemble the computational head models, but can also be used to approximate real HRTF data. These methods are not directly related to specific filter design issues, but can serve as a basis for, e.g., structural smoothing and preprocessing of the data.

Principal components analysis (PCA) has been used by Kistler and Wightman to approximate minimum-phase HRTFs [6]. In this method the magnitude spectra of the HRTFs were approximated using five principal spectral components of the response. With this method the order of the resulting FIR filters was successfully reduced to 1/3 of the original impulse response with only a slight decrease in localization accuracy.

Chen *et al*. [27] have proposed a feature extraction method, where a complex valued HRTF is represented as a weighted sum of eigentransfer functions

generated using the Karhunen–Loève expansion. The difference compared to the previous PCA model is that a complex HRTF transfer function including magnitude and phase information can be modeled.

2.2 Binaural Filter Design

2.2.1 FIR Models

The most straightforward way to approximate HRTF measurements is to use the frequency sampling FIR filter design [6] [7]. A filter of the desired order is obtained by windowing the measured impulse response with a rectangular window. The use of a rectangular window may be motivated because it is the optimal approximation to the original frequency response in the least-squares sense [9]. The effect of different window functions has been discussed by Sandvad and Hammershøi [9]. They concluded that although rectangular windowing provokes the Gibbs' phenomenon seen as ripple around amplitude response discontinuities, it is still favorable when compared to, e.g., the Hamming window.

Kulkarni and Colburn [8] have proposed the use of weighted least squares (WLS) techniques based on log-magnitude error minimization for finite-impulse response HRTF filter design. They claim that an FIR filter order of 64 is capable of retaining most of spatial information (only an abstract of [8] was available to the present authors at the time of writing).

2.2.2 IIR Models

The earliest HRTF filter design experiments using pole-zero models were carried out by Kendall *et al*. [28]. A comparison of FIR and IIR filter design methods was presented by Sandvad and Hammershøi [9]. The non-minimum-phase FIR filters based on individual HRTF measurements were designed using rectangular windowing. The IIR filters were generated using a modified Yule-Walker algorithm that performs least-squares magnitude response error minimization. The low-order fit was enhanced a posteriori by applying a weighting function and discarding selected pole-zero pairs at high frequencies. Listening tests showed that an FIR of order 72 equivalent to a 1.5 ms impulse response was capable of retaining all of the desired localization information, whereas an IIR filter of order 48 (equal number of poles and zeros) was needed for the same localization accuracy.

In the research carried out by Blommer and Wakefield [10], the error criteria in the ARMA filter design were based on log-magnitude spectrum differences rather than magnitude or magnitude-squared spectrum differences. Furthermore, a new approximation for the log-magnitude error minimization was defined. The theoretical study concluded that it was possible to design low-

order HRTF approximations (the given example used 14 poles and zeros) using the proposed method.

Asano *et al.* have investigated sound localization in the median plane [11]. They derived IIR models of different orders (equal number of poles and zeros) from individual HRTF data. When compared to a reference, a 40th-order pole-zero approximation yielded good results in the localization tests with the exception of increased front-back confusions in frontal incident angles.

Other IIR approximation models for HRTFs have been presented by Ryan and Furlong [29], Jenison [30], and Kulkarni and Colburn [12] (only an abstract was available to the authors at the time of writing).

An attractive technique for HRTF modeling has been proposed by Mackenzie *et al.* [14]. By using *balanced model truncation* (BMT) it is possible to approximate HRTF magnitude and phase response with low order IIR filters (down to order 10). A complex HRTF system transfer function is written as a state-space difference function, which is then represented in balanced matrix form. A truncated state-space realization $F_m(z)$ can be found with a similarity to the original system $F(z)$ which is approximately quantified by the Hankel norm:

$$\left\| F(z) - F_m(z) \right\|_H \leq 2\, trace(\Sigma_2) \tag{1}$$

where Σ_2 is the sum of Hankel singular values of the rejected system after truncation. In our experiments, minimum-phase diffuse-field equalized auditory smoothed HRTFs (based on Kemar measurements by Gardner and Martin [31]) were modeled by 10th order IIR filters created using BMT. The signal-to-error power ratios (SER) were compared to IIR models designed using Prony's method and the Yule-Walker method. The average SER was found to be approximately 10dB better in BMT models. Listening tests based on BMT designs will be carried out in the near future.

2.2.3 Warped Filter Structures

Pscyhoacoustically Valid Frequency Scales and Resolutions

It has been a long tradition in audio technology to plot magnitude responses using the decibel scale for ordinate and a logarithmic frequency scale for abscissa. This was found to describe better the auditory perception than when using linear scales and this is also technically convenient enough.

Digital signal processing (DSP) exhibits an inherent property to express practically everything on a linear frequency scale so that adapting to other scales needs special attention. This is due to the properties of the unit delay as a basic building block which implies uniform time and frequency resolution.

Spectrum analysis through the discrete Fourier transform shows this and, more importantly from the equalization point of view, filter designs follow the same rule unless special effort is taken.

In psychoacoustics it has been shown experimentally that there are yet better scales instead of the linear or logarithmic frequency scales and logarithmic dB-scale. *Loudness* in *sone* units [32] represents the perceived 'intensity' and *loudness level* in *phon* units is a related logarithmic scale. *Pitch*, the perceived 'height' of sound, has several competing scales. The traditional *mel* scale has in many technical fields been replaced by the *Bark* scale (or the critical-band rate scale) [32] although in practice these are very similar (1 Bark ≈ 100 mel). A strong competitor of the Bark scale is the ERB (Equivalent Rectangular Band-width) rate scale [33] that seems to be theoretically better motivated than the Bark scale [34].

Actually we should make difference between frequency resolution functions and pitch scales. Figure 3 shows the four resolution functions discussed above; lin, log, Bark, and ERB resolution in terms of the corresponding Q-value (center frequency divided by bandwidth) as a function of frequency. Linear resolution is plotted for uniform 100 Hz bandwidth and logarithmic resolution for third octave bandwidth. Figure 4 shows the corresponding 'rate' scales vs. log frequency.

As can be seen from Figures 3 and 4, the log and the ERB resolution functions are relatively close to each other. The Bark resolution is similar above 500 Hz. The constant bandwidth resolution function related to the linear frequency scale is generally not acceptable when characterizing responses from the auditory point of view. This is unfortunate since DSP methods, including filter design methods, work inherently on a linear scale. Based on the above theoretical discussion we may draw the conclusion that both the design of equalizer filters and the characterization of equalized responses are best represented on the ERB scales, the logarithmic and the Bark scales being useful approximations, and the linear scale being inferior. The question whether the monaural psychoacoustic prinicples apply in the same way to binaural hearing is a valid one. Thus, care should be taken when using (monaural) psychoacoustical measures in binaural design.

2.2.4 Warped Filters

The non-linear frequency resolution of human hearing suggests that modeling of HRTFs should also be carried out in the same manner. There are two possible approaches to approximate a non-linear frequency resolution. One possibility is to use weighting functions that allow more error at higher frequencies and demand a better fit at lower frequencies (e.g., [9], applied to HRTFs). The other possibility is to use a non-linear frequency resolution in the

filter design. This is often referred to as *frequency warping*.

Approximations of HRTFs using auditory criteria have not been extensively studied. Jot *et al*. [13] have proposed a method where the HRTFs are preprocessed using auditory smoothing and the IIR filter design using a standard Yulewalk algorithm is carried out in the warped frequency domain. A framework for warped HRTF filter design has been established by the authors [15]. The fundaments of warped filters are studied in the following.

Frequency scale warping is in principle applicable to any design or estimation technique. The most popular warping method is to use the bilinear conformal mapping. The bilinear warping is realized by substituting unit delays with first-order allpass sections

$$z^{-1} \Leftarrow D_1(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}} \tag{2}$$

where λ is the warping coefficient. This means that the frequency-warped version of a filter can be implemented by such a simple replacement technique. It is easy to show that the inverse warping can be achieved with a similar substitution but using $-\lambda$ instead of λ (this was used in [13]). In Figure 5, the effect of warping using the first-order allpass structure with different values of λ is illustrated.

The usefulness of frequency warping in our case comes from the fact that, given a target transfer function $H(z)$, we may find a lower order warped filter $H_w(D_1(z))$ that is a good approximation of $H(z)$. $H_w(D_1(z))$ should be designed in a warped frequency domain so that using allpass delays $D_1(z)$ instead of unit delays maps the design to a desired transfer function in the ordinary frequency domain. For an appropriate value of λ, the bilinear warping can fit the psychoacoustic Bark scale, based on the critical band concept [35], relatively accurately. For a sampling rate of 48 kHz $\lambda = 0.7313$ and for 22 kHz $\lambda = 0.6288$.

The transfer function expressions of warped filters may be expanded (dewarped) to yield equivalent IIR filters of traditional form, such as direct form II filters. Such implementations have been reported in the literature [13]. An alternative strategy is presented in [15], where implementation is carried out directly in the warped domain using warped FIR (WFIR) and IIR (WIIR) structures. The WFIR and WIIR structures are depicted in Figures 6 and 7. For more details, the reader is referred to an article by Karjalainen *et al*. [36] on the realization of warped filter structures.

The first advantage of warped forms over traditional filters is that in many cases the warping by allpass sections results in filters less critical from the point of view of computational precision needed. Another desirable feature found in WFIR structures is that for variable filters the coefficients are not in-

9

side recursive loops so that transients due to changing coefficients are effectively minimized. This feature may be attractive, e.g., in dynamic interpolation of HRTFs, where nonrecursive structures have been found to perform better.

Low-order approximations of HRTFs have been formulated by Huopaniemi and Karjalainen [15] using the proposed WFIR and WIIR approximations methods. In this work, one of the goals was to compare warped filter designs to conventional ones. Details, results and discussion will be presented in Chapters 4 and 5.

2.3 TRANSAURAL FILTER DESIGN

The theory of transaural stereo processing (crosstalk-compensated binaural information presented over a pair of loudspeakers) was first formulated by Schroeder and Atal over 30 years ago [37]. They described the use of a crosstalk cancellation filter for converting binaural recordings made in concert halls for loudspeaker listening. Their impressions of listening to transaurally reproduced dummy-head recordings were "nothing less than amazing". However, they observed the limitations of the listening area, the "sweet spot", which still remains a problem in transaural reproduction.

Damaske studied the transaural reproduction issues and formulated the theory further in the TRADIS project (True Reproduction of All Directional Information by Stereophony) [38]. He conducted studies on image quality deterioration as a function of listener placement. The transaural theory was refined and to some extent revitalized by works of Cooper and Bauck (Cooper and Bauck 1989). They created a concept of *spectral stereo*, which originally applied simplified head models for transaural processing. These techniques have been further developed by, for example, Kotorynski [39], and MacCabe and Furlong [40] to include improved head models and more sophisticated signal processing techniques. Recently, transaural processing systems for virtual acoustic source generation have been presented by Nelson *et al*. [41].

There are basically two alternatives in transaural filter design. One alternative is to use HRTF filter approximations in cascade with cross-talk canceling filters resulting in a total of four filters per a binaural signal. Another and a more attractive solution is the use of lattice structures, originally formulated by Cooper and Bauck [16]. The lattice structures illustrated in Fig. 2 can reduce the number of needed filters to two for monaural and binaural source material in the case of symmetrical listening position.

3 IMPLEMENTATION ISSUES

Computational efficiency is desirable in real-time auralization systems. To compare different filter design and implementation strategies, one should pay attention particularly to the following viewpoints:

1) Is the system dynamic, i.e., do we need HRTF interpolation?
2) Are we using minimum-phase HRTF approximations?
3) Are we using specialized hardware (signal processors) for implementation?
4) Are we storing great amounts of HRTF data?

The following benchmarks have been calculated for a Texas Instruments TMS320C3x floating point signal processor, but are practically similar in other processors as well. FIR implementation is efficient (N+3 instructions for N taps), and dynamic coefficient interpolation is possible. Designs are usually straightforward (e.g., frequency sampling), but give limited performance especially at low orders.

IIR implementations are slower ($2N+3$ instructions for order N in direct form II implementation) if dynamic synthesis is required (cross-fading, transient elimination often doubles the computation). Pole-zero models are suited for arbitrary-shaped magnitude-response designs, thus low-order designs are possible.

The efficiency of warped vs. non-warped filters depends on the processor that is used. For Motorola DSP56000 series signal processors a WFIR takes three instructions per tap instead of one for an FIR. For WIIR filters four instructions are needed instead of two for an IIR filter. In custom design chips the warped structures may be optimized so that the overhead due to complexity can be minimized. The warped structures may also be expanded, "dewarped", into direct form filters, which will lead to the same computational demands as with normal IIR filters.

4 EXPERIMENTAL STUDIES

One goal of this project was to compare different filter design methods using both theoretical and experimental means. In the following, details are presented about the HRTF filter design examples and the listening experiments that were conducted.

4.1 HRTF Filter Design

The HRTFs used in the filter design examples and listening tests were measured from a Cortex MK2 dummy head in an anechoic chamber. The Cortex dummy head was equipped with Brüel&Kjaer 4190 microphones (blocked ear canal version). The transducer used in the measurements was a four-inch Audax AT100M0 loudspeaker mounted in a plastic ball. A random-phase flat amplitude spectrum pseudorandom noise signal was used as the excitation sequence. Data were played and recorded using an Apple Macintosh host com-

puter and the QuickSig signal processing environment [42]. A National Instruments NI-2300B DSP card with Texas Instruments TMS320C30 signal processor and high-quality 16-bit AD/DA converters were used for both HRTF measurements and listening experiments.

The HRTF data were post-processed for headphone listening experiments in the following way. A compensation measurement was made to account for the measurement system by placing a microphone at the dummy head position with the head absent (similarly as in [4], p. 301). Headphone transfer functions for the Sennheiser HD580 headphone were measured on the dummy head. A 300-tap FIR inverse filter for each ear was designed using least-squares approximation. The HRTF data was then convolved by the compensation response and the headphone correction filter for both ears.

The minimum-phase reconstruction was carried out using windowing in the cepstral domain (as implemented in the Matlab Signal Processsing Toolbox rceps.m function [43]). The cross-correlation method [6] was used to find the ITD for each incident angle. The ITD was inserted as a delay line. We used three different minimum-phase HRTF approximations: Windowed FIR design (rectangular window), time-domain IIR design (Prony's method, implemented in Matlab [43]), and a warped IIR design (warped Prony's method, warping coefficient $\lambda = 0.65$). In Table 2, the processed filter lengths for different filter types are illustrated.

The example HRTFs were measured at 0° and 135° azimuth (0° elevation) positions. The magnitude responses of the HRTFs can be seen in Figures 8-13. It can be seen that a warped Prony design easily outperforms a linear Prony design of equivalent order. The better fit at low frequencies when comparing WIIR approximation to windowed FIR can also be observed. The value of $\lambda = 0.65$ was used, which is slightly lower than for approximative Bark-scale warping.

4.2 Listening Experiment

In order to verify the theoretical filter design results we carried out headphone listening experiments. The goal in the study was to determine the subjective thresholds of filter order using different design methods when compared to a reference HRTF (similarly as in [9]). A total of 8 test subjects participated in the listening experiment, 6 male and 2 female with ages ranging between 21 and 35. The hearing of all test subjects was tested using standard audiometry. None of the subjects had reportable hearing loss that could effect the test results. It should be pointed out here that the experiment was done using *non-individualized HRTFs* (measured on a dummy head) that were equalized for a specific headphone type (Sennheiser HD580).

4.2.1 Test Method

In the listening experiment we used an adaptive TAFC (Two Alternatives Forced Choice) bracketing method. The method is to great detail described in [44] and widely used in, e.g., audiometric tests. In each trial, two test sequences were presented with a 0.5 s interval between the samples. The first test signal was always the reference signal, and the second signal varied according to adaptation. Each test type was repeated three times and only the last two were accounted for in the data analysis.

The test persons were given written and oral instructions. They were also familiarized with a test sequence that demonstrated both distinguishable and undistinguishable test signal pairs.

4.2.2 Test Stimuli

A total of four different stimuli were first processed for a pilot study, pink noise, male and female speech, and a music sample. All samples were digitally copied and processed from the Music for Archimedes CD[1]. In the final experiment, however, only the pink noise sample was used. This was due to the fact that remarkable differences in different filter designs could clearly be heard only using wide-band test signals. A pink noise sample with a length of one second (50 ms onset and offset ramps) was used in the final experiment. The level of the stimuli was adjusted so that the peak A-weighted SPL did not exceed 70 dB at any point. This has been done in order to avoid level adaptation and the acoustical reflex (Stapedius reflex).

4.2.3 Test Procedure

The test person was seated in a semi-anechoic chamber (anechoic chamber with a hard cardboard floor). The test stimuli were presented over headphones. A computer keyboard was placed in front of the test person. Each test person was individually familiarized and instructed to respond in the following way: "press 1 if the signals are the same", "press 2 if the signals are different", "press Space if you want to repeat the signal pair".

As a total, three different filter approximations for two apparent source positions were used. Each alternative was repeated three times. The results of the listening tests were gathered automatically by a program written for the QuickSig environment [42]. The result data were transferred into Matlab, where analysis was performed.

---

[1] Music for Archimedes, CD B&O 101 (1992)

4.3 Results

In Figure 14, the results of the listening test are presented. This figure illustrates the distribution of just noticeable difference (JND) thresholds calculated across two tested azimuth angles, 0° and 135° with three filter types, FIR, IIR, and WIIR. The median value as well as the lower and upper quartile (25% and 75% levels) values are shown.

The results show that the distribution of results in the listening panel was relatively small, although not as well defined as a pilot study indicated. This may be a consequence of using an inhomogeneous listener panel. Some of subjects were experienced analytic listeners while some did not have prior experience in a listening panel. A longer training prior to final experiments could have made the test results more systematic [45].

From Fig. 14 it can be seen that the WIIR performance from the filter order point of view is superior when compared to FIR and IIR designs. From the computational point of view, however, (see section 3) the FIR and WIIR implementations appear to be approximately equal in performance. The warped IIR designs, however, easily outperforms a conventional IIR design. A useful criterion to select filter order values could be the upper quartile (75%) or even higher level of subject reactions. Using the 75% quartile results, one concludes in the following statements.

For non-individualized (dummy-head) HRTFs equalized to a specific headphone, the filter orders where 75% of the population stated no difference when compared to the reference were approximately:

- Order 40 for a frequency-sampling FIR design
- Order 25 for a time-domain IIR design (Prony's method)
- Order 20 for a warped IIR design (Prony's method, $\lambda = 0.65$)

In comparison to the results presented in the literature, some comments can be made. The empirical study by Sandvad and Hammershøi [9] resulted in orders 72 for a FIR and 48 for an IIR filter. The difference compared to the results may be caused by the fact that Sandvad and Hammershøi used individual HRTFs and headphone calibration, and both speech and pink noise samples. However, the estimated detection probabilities (maximum likelihood estimation was used as a statistical model) for the given results were approximately 0.08, higher than in our conclusions. Moreover, the filter orders used in that study were relatively sparse (24, 36, 72 and 128 taps for FIR, 10, 20, 30 and 48 taps for IIR using pink noise).

4.4 Filter Design Errors as Spectral Distance Measure

There is a need to have a simple numerical measure of filter design quality that is meaningful also from the perceptual point of view. We experimentally derived a spectral (magnitude) distance measure in the following way.

- The equalized impulse response is first FFT transformed to power spectrum, resampled (by interpolation) uniformly on a logarithmic frequency scale, smoothed with about 0.2 octave resolution (this resolution value was specified somewhat arbitrarily to be not too far from the ERB resolution, see section 3.5), and converted to dB scale.

- The difference of the spectrum to be analyzed and a reference spectrum is computed for the passband region of approximation. The reference spectrum may be simply the average level of the spectrum to be analyzed or some other reference. In our case it was the reference in our listening experiments, described in section 4.2.

- A root-mean-square value of the difference spectrum is computed and this is used as an objective spectral distance measure to characterize the perceptual difference between the magnitude responses or a deviation from a reference response. Notice that the values of the spectral distance measure used in our study are not calibrated to be compared directly with any perceptual difference measures.

Figure 15 plots the spectral distance measures as functions of filter order for the three HRTF filter types used in our study: FIR, IIR, and WIIR. The reference for distance computation was the highest-order response, in order to make the results compatible to the setup used in our listening experiments.

From the spectral distance measure of Fig. 15 it can be seen that from the auditory point of view the WIIR filters have the best performance. It is, however, a valid question, why the FIR filters performed so well in the listening tests without having any weighting or frequency warping applied. Further work in this is to be conducted.

5 CONCLUSIONS

In this paper we have reviewed existing HRTF filter design and implementation strategies for binaural and transaural processing. In theoretical and empirical studies three different filter design methods were compared: FIR design based on the frequency sampling method (rectangular windowing), time-domain IIR design (Prony's method), and a warped IIR design. Frequency

warping enables the design of filters using a non-linear frequency scale, similar to the function of the human ear. The results of the listening tests showed that warped IIR structures outperform conventional IIR filters. The performance of FIR filters was equal to that of the warped IIR filters.

In the future, more listening tests will be made both on non-individualized and individualized HRTFs in order to get a more thorough view of filter performance.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Begault, 3-D Sound for Virtual Reality and Multimedia (Academic Press, 1994).

[2] J. Blauert, *Spatial Hearing. The psychophysics of human sound localization* (Revised edition, MIT Press, Cambridge, Massachusetts, 1997).

[3] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, pp. 171–218 (1992).

[4] H. Møller, M. Sørensen, D. Hammershøi, and C. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300–321 (1995 May).

[5] G. Kendall, "3-D sound primer: Directional hearing and stereo reproduction," *Computer Music Journal*, vol. 19, no. 4, pp. 23–46 (Winter 1995).

[6] D. Kistler and F. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, no. 3, pp. 1637–1647 (1992).

[7] D. Begault, "Challenger to the successful implementation of 3-D sound," *J. Audio Eng. Soc.*, vol. 39, no. 11, pp. 864–870 (1991 Nov.).

[8] A. Kulkarni and H. S. Colburn, "Efficient finite-impulse-resonse filter models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pt. 2, pp. 3278 (1995).

[9] J. Sandvad and D. Hammershøi, "Binaural auralization. Comparison of FIR and IIR representation of HIRs," Presented at the 96th AES Convention, Amsterdam, The Netherlands, 1994 Feb. 26–Mar. 1, preprint 3862.

[10] M. A. Blommer and G. H. Wakefield, "On the design of pole-zero approximations using a logarithmic error measure," *IEEE Trans. Signal processing*, vol. 42, no. 11, pp. 3245–3248 (1994).

[11] F. Asano, Y. Suzuki, and T. Sone, T, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Am.*, vol. 88, no. 1, pp. 159–168 (1990).

[12] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response filter models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pt. 2, pp. 3278 (1995).

[13] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," Presented at the 98th AES Convention, Paris, France, 1995 Feb. 25–28, preprint 3980.

[14] J. Mackenzie, J. Huopaniemi, V. Välimäki, and I. Kale, "Low-order modelling of head-related transfer functions using balanced model truncation," Accepted for publication in: *IEEE Signal Proc. Letters* (1996).

[15] J. Huopaniemi and M. Karjalainen, "Comparison of digital filter design methods for 3-D sound," *Proc. IEEE Nordic Sig. Proc. Symp. (NORSIG'96)*, Espoo, pp. 131–134 (Sept. 25–28, 1996).

[16] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol. 37, no. 1/2, pp. 3–19 (1989).

[17] M. J. Walsh and D. J. Furlong, M. 1995, "Improved spectral stereo head model," Presented at the 99th AES Convention, New York, 1995 Oct. 6–9, preprint 4128.

[18] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).

[19] D. H. Cooper and J. L. Bauck, "Corrections to L. Schwarz, 'Zur Theorie der Beugung einer evenen Schallwelle and der Kugel,' Akust. Z., vol. 8, pp. 91-117 (1943)," *J. Acoust. Soc. Am.*, vol. 80, no. 6, pp. 1793–1802 (1986).

[20] S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.*, vol. 61, no. 6, pp. 1567–1576 (1977).

[21] A. Kulkarni, S. K. Isabelle and H. S. Colburn, "On the minimum-phase approximation of head-related transfer functions," *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, 1995 October 15–18.

[22] E. M Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, T, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, no. 1, pp. 111–123 (1993 July).

[23] H. Møller, M. Sørensen, C. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469 (1996 June).

[24] H. Møller, C. Jensen, D. Hammershøi, and M. Sørensen, "Using a typical human subject for binaural recording," Presented at the 100th AES Convention, Copenhagen, Denmark, 1996 May 11–14, preprint 4157.

[25] H. Møller, Personal communication, Nordic Acoustical Meeting (NAM'96), Helsinki, Finland, 1996 June 12–14.

[26] J. Köring and A. Schmitz, "Simplified cancellation of cross-talk for playback of head-

related recordings in a two-speaker system," *Acustica*, vol. 79, pp. 221–232, 1993.

[27] J. Chen, B. Van Veen, and K. E. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 439–452 (1995).

[28] G. S. Kendall, and W. L. Martens, "Simulating the cues of spatial hearing in natural environments," *Proc. 1984 Int. Comp. Music Conf. (ICMC'84)*, Paris, pp. 111–125 (1984).

[29] C. Ryan and D. Furlong, "Effects of headphone placement on headphone equalization for binaural reproduction," Presented at the 98th AES Convention, Paris, 1995 Feb. 25–28, preprint 4009.

[30] R. L. Jenison, "A spherical basis function neural network for pole-zero modeling of head-related transfer functions," *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, 1995 October 15–18.

[31] B. Gardner and K. Martin, "HRTF measurements of a KEMAR", *J. Acoust. Soc. Am.*, vol. 97, no. 6, pp. 3907–3908 (1995).

[32] E. Zwicker and H. Fastl, *Psychoacoustics* (Springer-Verlag, 1990).

[33] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "Auditory filter shapes at low center frequencies," *J. Acoust. Soc. Am.*, vol. 88, pp. 132–140 (1980).

[34] B. C. J. Moore, R. W. Peters, and B. R. Glasberg, "A revision of Zwicker's loudness model," *Acta Acustica*, vol. 82, pp. 335–345 (1996).

[35] J. O. Smith, and J. Abel, "The Bark bilinear transfrom," *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk Mountain House, New Paltz, New York, 1995 October 15–18.

[36] M. Karjalainen, A. Härmä, and U. K. Laine, "Realizable warped IIR filters and their properties," To be published: *Proc. IEEE ICASSP'97*, Munich, Germany, 1997 April.

[37] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," *IEEE Conv. Rec.*, pt. 7, pp. 150–155 (1963).

[38] P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction," *J. Acoust. Soc. Am.*, vol. 50, pt. 2, pp. 1109–1115 (1971 Oct.).

[39] K. Kotorynski, "Digital binaural/stereo conversion and crosstalk cancelling," Presented at the 89th AES Convention, Los Angeles, 1990 Sept. 21–25, preprint 2949.

[40] C. J. MacCabe and D. Furlong, "Spectral stereo surround sound pan-pot," Presented at the 90th AES Convention, Paris, 1991 Feb. 19–22, preprint 3067.

[41] P. A. Nelson, F. Orduña-Bustamante, and D. Engler, "Experiments on a system for the synthesis of virtual acoustic sources," *J. Audio Eng. Soc.*, vol. 44, no. 11, pp. 990–1007 (1996 November).

[42] M. Karjalainen, "DSP software integration by object-oriented programming – a case study of QuickSig," *IEEE ASSP Magazine*, pp. 21–31 (1990 April).

[43] MathWorks, Inc., *MATLAB, Signal Processing Toolbox*.

[44] ISO 8253-1. *Acoustics - Audiometric test methods - Part 1: Basic pure tone air and bone conduction threshold audiometry*, pp. 587–599 (1989).

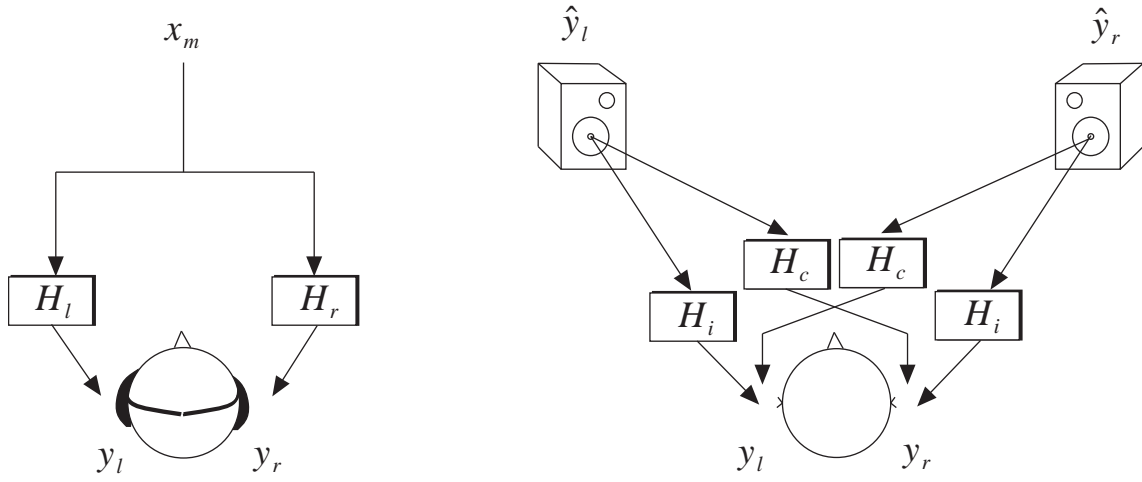[45] S. Bech, "Training of subjects for auditory experiments," *Acta Acustica*, vol. 1, pp. 89–99 (1993 June/August).

Fig. 1. *Transfer functions for binaural and transaural processing.*
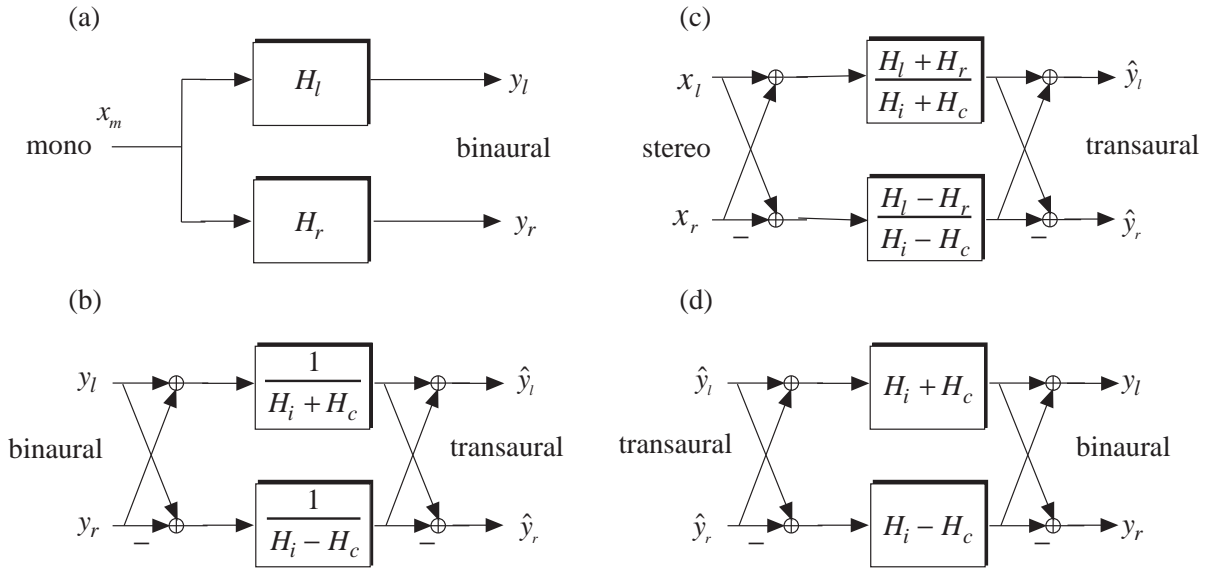
Fig. 2. *Shuffler structures for 3-D sound filter conversions. a) monophonic to binaural, b) binaural to transaural, c) stereophonic to transaural, and d) transaural to binaural conversion (based on [13].*
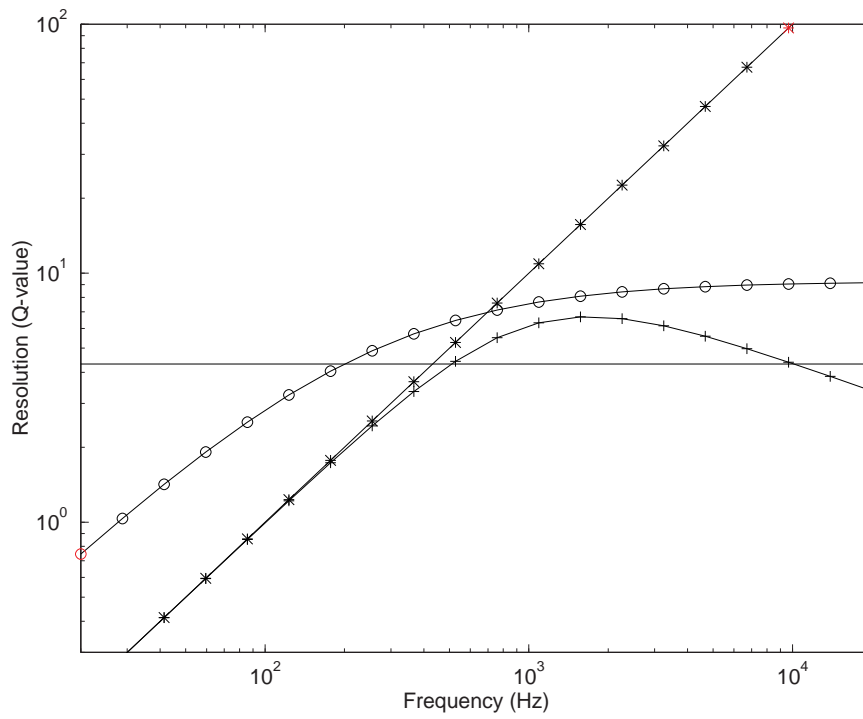
Fig. 3. *Frequency resolution (Q-value) curves as functions of frequency: Solid line = third-octave (constant Q); o-o = ERB resolution; +-+= Bark (critical band) scale; *-* = constant 100 Hz bandwidth (linear resolution).*
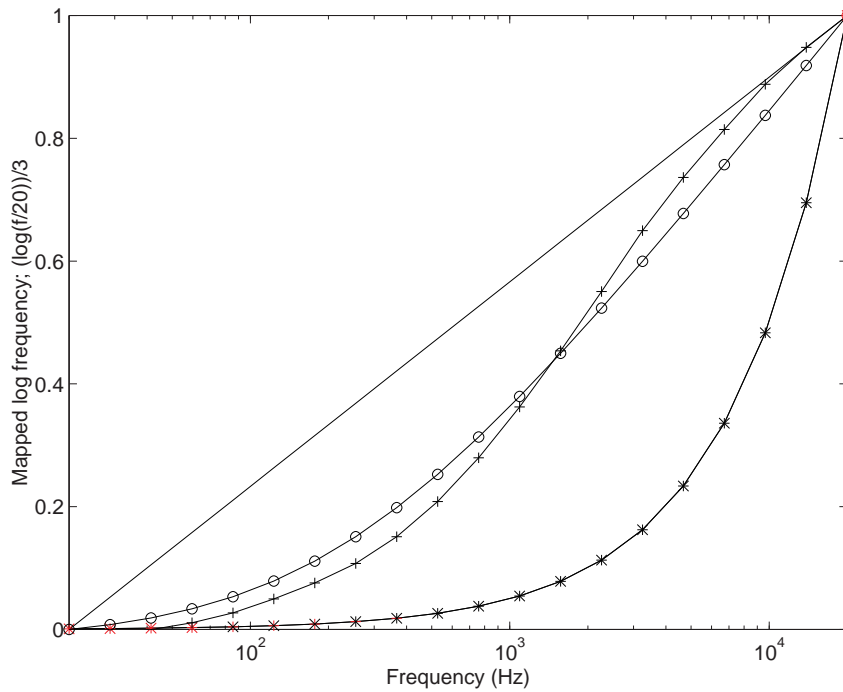
Fig. 4. *Mapped frequency scales as functions of frequency: Solid line = logarithmic (cf. constant Q); o-o = ERB rate scale; +-+= Bark (critical band) scale; *-* = linear frequency scale.*
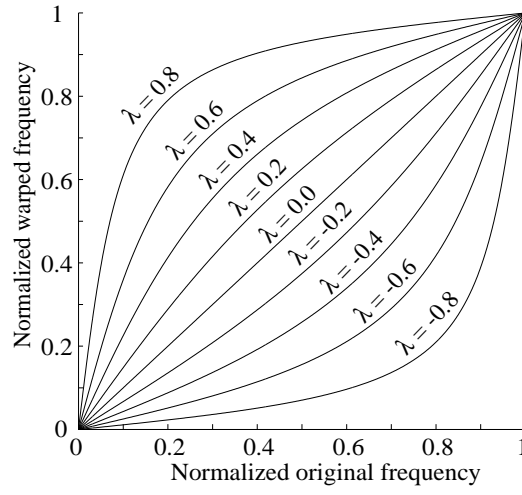
Fig. 5. *Frequency warping characteristics of the first-order allpass section for different values of the warping parameter λ. Frequencies are normalized to the Nyquist rate.*
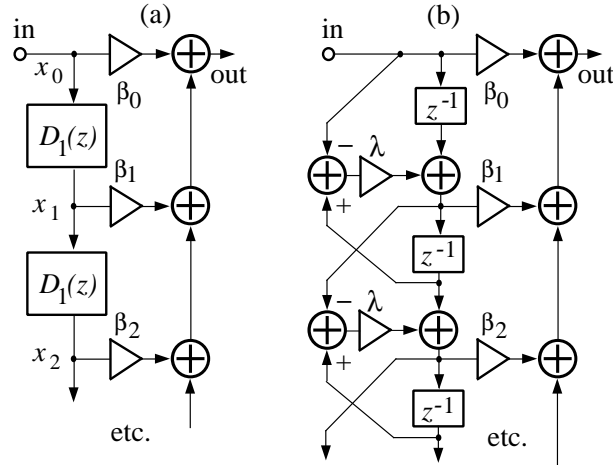

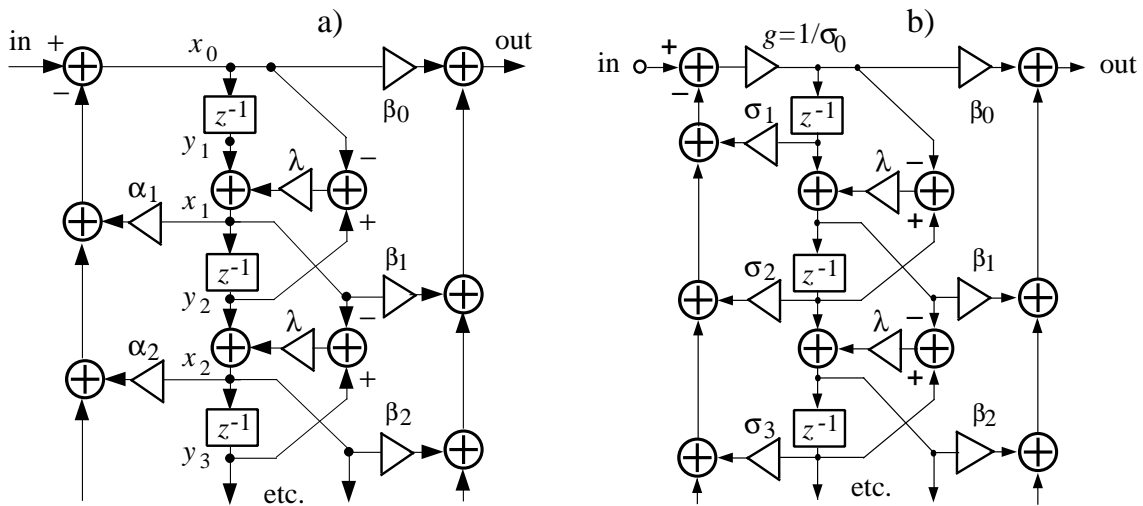**Fig. 6.** *a) The principle of warped FIR filter and b) practical implementation.*
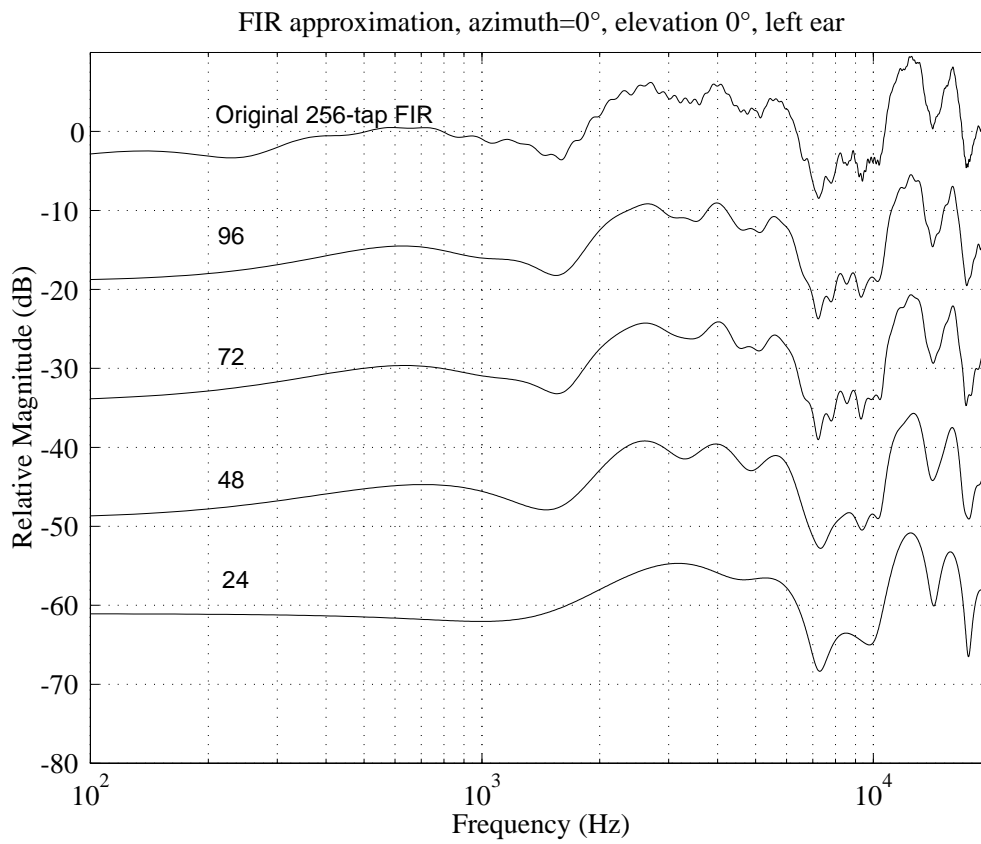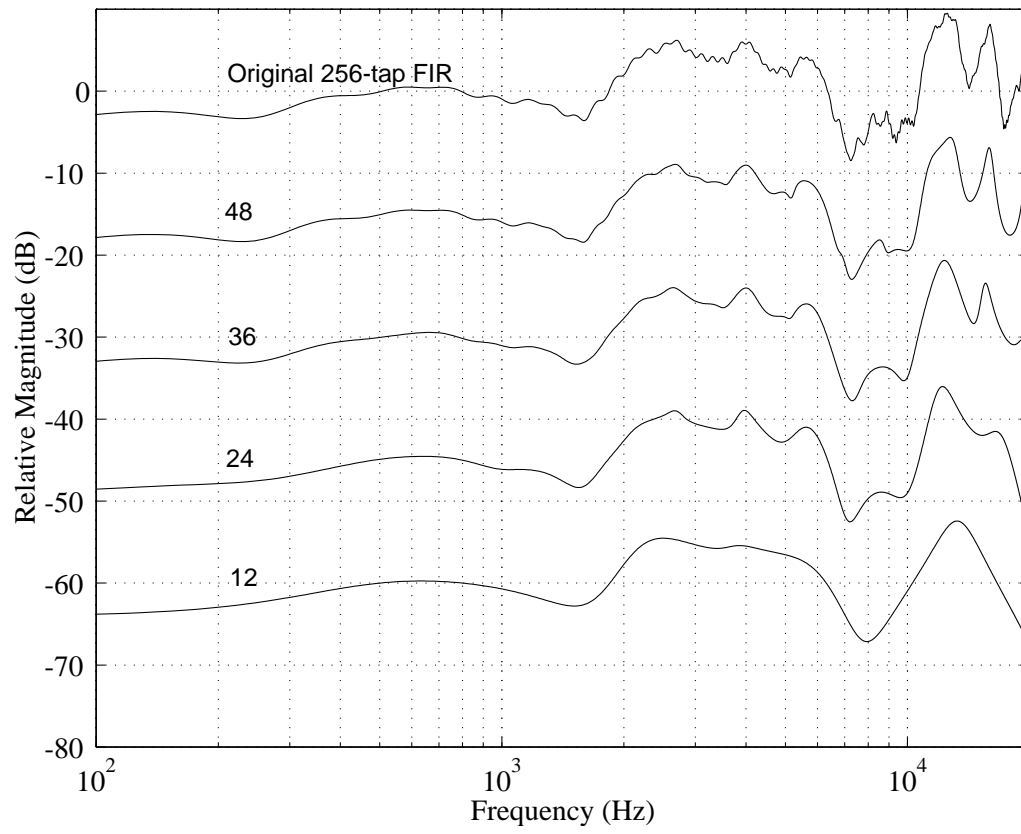

Fig. 7. *The unrealizable direct form of warped IIR filter and b) the realizable modified implementation.*

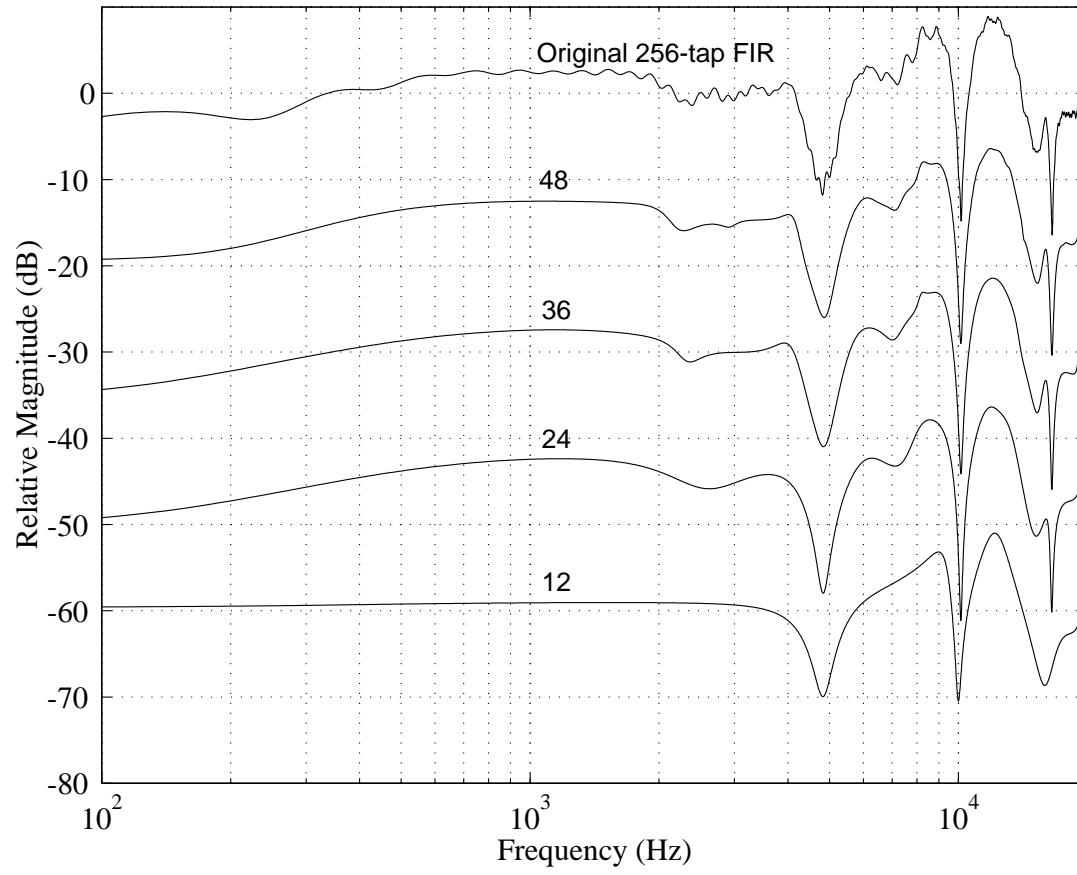**Figs. 8-10.** *FIR, IIR and WIIR approximation of Cortex HRTFs, 0° azim.*

FIR approximation, azimuth=0°, elevation 0°, left ear

WIIR approximation, azimuth=0°, elevation 0°, left ear, lambda=0.65

Original 256-tap FIR

48

36

24

12

Relative Magnitude (dB)

Frequency (Hz)

24

IIR approximation, azimuth=135°, elevation 0°, left ear

Listening test results: azimuth angles 0°, 135°
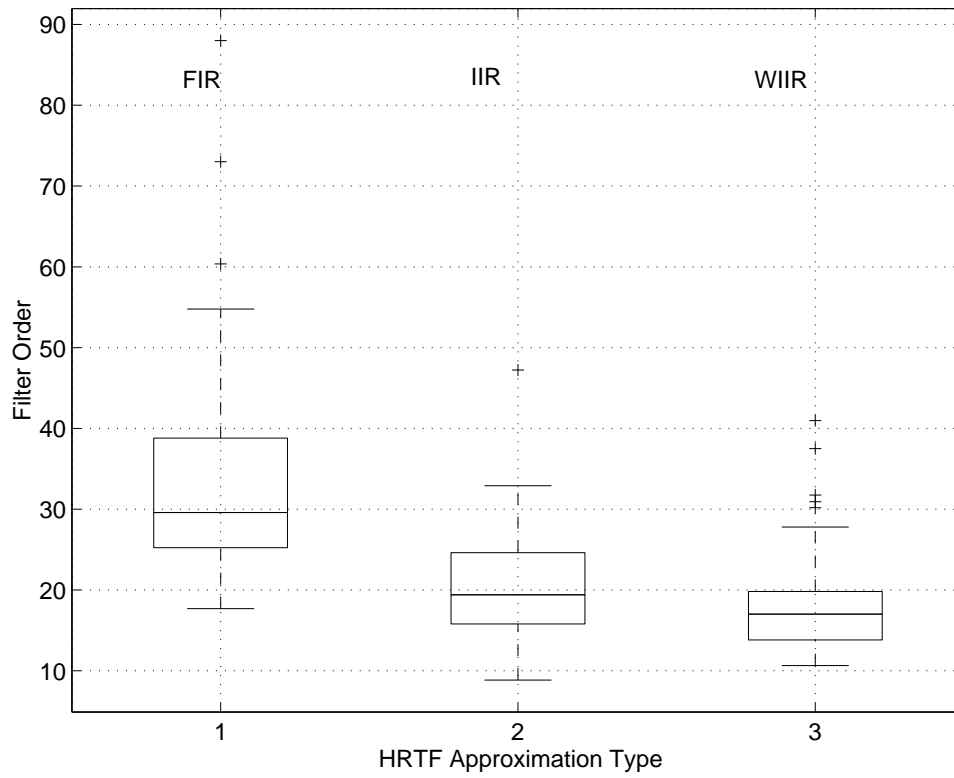
Research Group	Design type	Filter Order	Study
Begault, 1991 [7]	binaural / FIR	81-512	empirical
Sandvad and Hammershoi, 1994 [9]	binaural / FIR	72	empirical
Kulkarni and Colburn, 1995 [8]	binaural / FIR	64	empirical?
Asano *et al*., 1990 [11]	binaural / IIR	>40	empirical
Sandvad and Hammershoi, 1994 [9]	binaural / IIR	48	empirical
Blommer and Wakefield, 1994 [10]	binaural / IIR	14	theoretical
Jot *et al*., 1995 [13]	binaural / IIR	10-20	empirical?
Ryan and Furlong, 1995 [29]	binaural / IIR	24	empirical?
Kulkarni and Colburn, 1995 [12]	binaural / IIR	6	empirical?
Kulkarni and Colburn, 1995 [12]	binaural / IIR	25 (all-pole)	empirical?
Mackenzie *et al*., 1996 [14]	binaural / IIR	10	theoretical
MacCabe and Furlong, 1991 [40]	transaural / IIR	20	empirical?
Kotorynski, 1995 [39]	transaural / FIR	64	empirical?
Kotorynski, 1995 [39]	transaural / IIR	32	empirical?

Table 1. *HRTF filter design data from the literature.*

FIR (rect. windowing)	IIR (Prony's method)	Warped IIR (λ=0.65)
256 (reference)	128	128
128	64	64
96	48	48
88	44	44
80	40	40
72	36	36
64	32	32
60	30	30
56	28	28
52	26	26
48	24	24
44	22	22
40	20	20
36	18	18
32	16	16
28	14	14
24	12	12
20	10	10
16	8	8

Table 2. *HRTF filter types and orders used in the listening experiment.*