IEEE Nordic Signal Processing Symposium (NORSIG96), Helsinki, Finland, Sept. 24–27, 1996

Comparison of Digital Filter Design Methods for 3-D Sound

Jyri Huopaniemi and Matti Karjalainen

Helsinki University of Technology
Laboratory of Acoustics and Audio Signal Processing
Otakaari 5 A, FIN-02150 Espoo, Finland
E-mail: Jyri.Huopaniemi@hut.fi, Matti.Karjalainen@hut.fi

ABSTRACT

In this paper, we will compare filter design methods for binaural and transaural processing. Most of the HRTF filter design techniques use a linear frequency scale approach with additional weighting functions in some cases. The human ear, however, processes the auditory information according to a nonuniform critical band frequency resolution. We have presented new methods for HRTF filter design using auditory criteria [1]. These methods are evaluated and compared with traditional design techniques. Results of the study are applicable to real-time auralization tasks where computational efficiency is at its most important.

1. INTRODUCTION AND MOTIVATION

Real-time digital modeling of human spatial hearing cues is often referred to as 3-D sound spatialization or auralization. The static cues of spatial hearing are contained in head-related transfer functions (HRTF). This transfer function describes the transmission from a point in the free field to a point in a human subject's or dummy head's ear canal [2].

Application areas of 3-D audio technology are versatile ranging from teleconferencing to virtual reality and multimedia [3] [4]. In 3-D audio synthesis, two alternative methods for sound production should be considered; *binaural processing* for headphone and *transaural processing* for loudspeaker listening. These methods require different filter design strategies and are clearly separate problem fields, although the basic data – the HRTFs – are the same.

2. HRTF PROPERTIES

HRTF impulse responses are the output of a linear and time-invariant system, that is, the diffraction and reflections around the human head, the outer ear, and the torso. Thus the impulse responses can directly be represented as finite-impulse response (FIR) filters. There are often computational constraints that lead to the need of HRTF impulse response 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.

An attractive property of HRTFs is that they may be modeled as minimum-phase filters [5] [6]. 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. According to Kistler and Wightman [6], minimum-phase reconstruction does not have any perceptual consequences.

2.1. 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 response. The transfer functions of the driving loudspeaker, the microphone and the ear canal (if the measurement position was 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 [7] [8]. Free-field equalization is achieved by dividing 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 specturm 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 dividing 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 frequency response to approximate, for example, the critical-band resolution of the human ear [9] [10]. This smoothing applies only to the magnitude response, so it is assumed that the phase can be calculated by minimum-phase reconstruction.

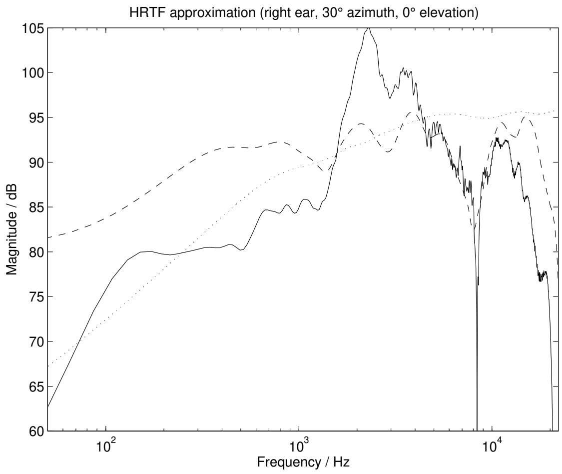

Fig. 1. *HRTF approximation. Solid line: Original KEMAR HRTF [11], Dashed line: Diffuse-field equalized Bark-smoothed KEMAR HRTF [10], Dotted line: Rayleigh scattering function [12].*

2.2. HRTF Modeling

The synthesis of binaural or transaural signals can be accomplished based on two approaches: the computational, and the empirical approach. 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 [12]. The latter approach is applicable, e.g., in the design of cross-talk canceling filters for transaural processing [13] [14]. The difference in the complexity of analytical and empirical HRTF models is illustrated in Fig. 1. The raw measurement (solid line) includes the transfer characteristics of the measurement system. It can be seen that the computational model (dotted line) resembles the measured data, but lacks the idiosyncratic features of the head such as the pinnae.

2.3. 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.

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.* [15] 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.

3. HRTF FILTER DESIGN

3.1. FIR Models

The most straightforward way to approximate HRTF measurements is to use the windowing FIR filter design. A filter of the desired order is obtained by windowing the measured impulse response with, e.g., a rectangular window (frequency sampling method). The use of a rectangular window may be motivated because it is the closest approximation to the original frequency response in the least-squares sense [16].

3.2. IIR Models

A comparison of FIR and IIR filter design methods was presented by Sandvad and Hammershøi [16]. The FIR filters 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 was needed for the same localization accuracy.

In [17], 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. Other IIR approximation models for HRTFs have been presented by Asano *et al.* [18], and Kulkarni and Colburn [19].

3.2.1. Balanced Model Truncation

An attractive technique for HRTF modeling has been proposed by Mackenzie *et al.* [10]. 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:

$$\|F(z) - F_m(z)\|_H \le 2\, trace(\Sigma_2) \qquad (1)$$

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 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.

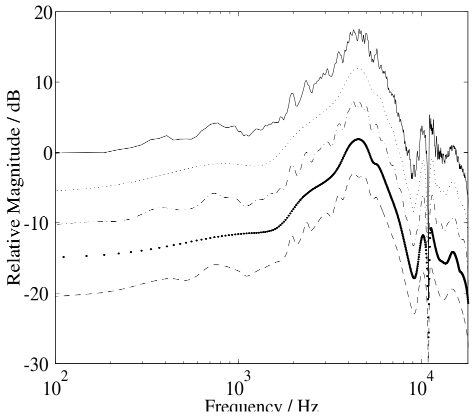

Fig. 2. *Solid line: original measurement, double-dotted line: 90-tap windowing FIR, dash-dot line: 90-tap WFIR, dotted line: Prony IIR design, order 44, dashed line: Warped Prony IIR design, order 44.*

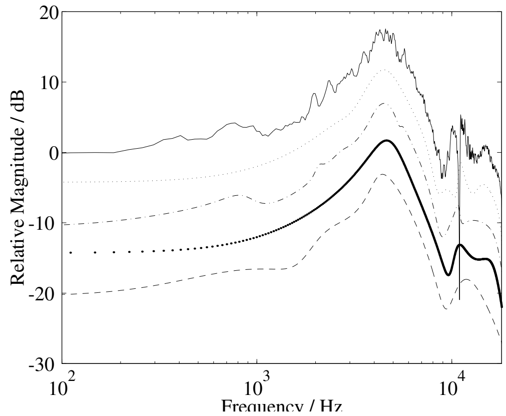

Fig. 3. *Solid line: original measurement, double-dotted line: Prony IIR design, order 22, dash-dot line: Warped Prony IIR design, order 22, dotted line: Prony, order 12, dashed line: Warped Prony, order 12.*

3.2.2. Warped Filter Structures

It is known that the human ear processes the auditory information according to a nonuniform critical band frequency resolution [20]. This suggests that modeling of HRTFs should also be carried out in the same manner. There are two possible approaches to approximate a nonlinear frequency resolution. One possibility is to use weighting functions that allow more error at higher frequencies and demand a better fit at lower frequencies frequencies (e.g., [16], 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.* [9] 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 [1]. 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

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

where λ is the warping coefficient. λ 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 [9]).

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 [20], relatively accurately. For a sampling rate of 44.1 kHz λ = 0.7233 and for 22 kHz λ = 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 [9]. An alternative strategy is presented in [1], where implementation is carried out directly in the warped domain using warped FIR (WFIR) and IIR (WIIR) 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 inside 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.

HRTF measurements were made in an anechoic chamber using the B&K 4100 dummy head. The results were processed according to the following principles: 1) minimum-phase reconstruction and 2) smoothing of the magnitude response data. Frequency warping was performed after the pre-processing. In the warped frequency domain, different FIR and IIR filter design methods were compared. As an example, we used a time-domain IIR design method, Prony's method. In Fig. 2, a modeling result is illustrated. The example HRTF was measured at 0° elevation, 30° azimuth, and the frequency response of the measurement loudspeaker was deconvolved. It can be seen that a warped Prony design easily outperforms a linear Prony design of equivalent order. In this case, the order of the IIR filters was 44 compared to the FIR designs order of 90. Rectangular windowing was used both in the WFIR and FIR designs. The better performance of a WFIR compared to non-warped FIR of the same order is quite noticeable. In Fig. 3, the filter orders have been reduced to 22 and 12. The value of $\lambda = 0.65$ was used, which is slightly lower than for approximative Bark-scale warping.

4. FILTER DESIGN AND 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 specialized hardware (signal processors) for implementation? Are we storing great amounts of HRTF data?

FIR implementation is efficient (N+3 instr. for N taps), and dynamic coefficient interpolation is possible. Designs are usually straightforward (e.g., windowing), but give limited performance especially at low orders.

IIR implementations are slower (2N+3 instr. for order N) 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.

5. CONCLUSIONS

In this paper we have compared existing HRTF filter design and implementation strategies for binaural and transaural processing. In the future, listening tests will be performed on different designs and implementations.

REFERENCES

[1] J. Huopaniemi and M. Karjalainen, "HRTF filter design based on auditory criteria," *Proc. Nordic Acoustical Meeting (NAM'96)*, Helsinki, June 12–14, 1996, pp. 323–330.

[2] 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, May 1995.

[3] D. Begault, *3-D Sound for Virtual Reality and Multimedia*. Associated Press, 1994.

[4] T. Takala, R. Hänninen, V. Välimäki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen, "An integrated system for virtual audio reality," *100th AES Convention*, Copenhagen, Denmark, May 11–14, 1996, preprint no. 4229.

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

[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] J. Blauert, *Spatial Hearing*. MIT Press, 1983.

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

[9] J.-M. Jot, V. Larcher, and O. Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," *98th AES Convention*, Paris, France, Feb. 25–28, 1995, preprint no. 3980.

[10] 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.

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

[12] 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.

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

[14] M. J. Walsh and D. J. Furlong, M. 1995, "Improved spectral stereo head model," *99th Audio Engineering Society Convention*, New York, Oct. 6–9, 1995, preprint no. 4128.

[15] 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.

[16] J. Sandvad and D. Hammershøi, "Binaural auralization. Comparison of FIR and IIR representation of HIRs," *96th AES Convention*, Amsterdam, The Netherlands, Feb. 26–Mar. 1, 1994, preprint no. 3862.

[17] 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.

[18] 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.

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

[20] E. Zwicker and H. Fastl, *Psychoacoustics*. Springer-Verlag, 1990.