STRUCTURED IIR MODELS FOR HRTF INTERPOLATION

Marcelo Queiroz, Gustavo H. M. de Sousa**

Computer Science Department, University of São Paulo, Brazil
{mqz,gustavo}@ime.usp.br

ABSTRACT

In this article we propose a structured IIR model for HRTF filters that can be applied to multi-user binaural soundscape rendering of virtual sound sources moving continuously in 3D space. A motivation for using low-order IIR models in HRTF interpolation is to trade some reasonable amount of soundscape fidelity for computational performance. Two strategies for interpolation of these IIR models are discussed, one using filter coefficients and the other using pole-zero models. Numerical experiments show that low-order structured IIR models provide reasonably good fitting of the original filters, allow for a larger number of simultaneous simulations than FIR equivalents, and may be used in an interpolation context without severely degrading the quality of the soundscape approximations.

1. INTRODUCTION

The goal of this paper is to present a structured IIR (Infinite Impulse Response) model for representing HRTF (Head-Related Transfer Function) filters of a finite HRTF database that is suitable for HRTF interpolation. This model presupposes an auralization context where virtual sound sources are binaurally rendered to a user wearing headphones, and the user or the virtual sound sources (or both) are moving continuously in space. The interpolation of the static HRTF filters aims at creating the perception of the spatial trajectories of the virtual sound sources, while ensuring that filter parameters change continuously.

Most previous work on HRTF interpolation for simulating moving virtual sound sources focuses on linear interpolation using FIR (Finite Impulse Response) models, which can be performed in time domain (HRIR or Head-Related Impulse Response interpolation) [10, 1, 13, 12] or frequency domain (using IPTF or Inter-Positional Transfer Functions [4, 9]), with similar results, due to linearity of the Fourier and z transforms. Other related strategies are Magnitude Frequency Response interpolation [5, 8] and SFRS or Spatial Frequency Response Surface interpolation [3], which are based on magnitude information only. Linear interpolation of phase information is intrinsically complicated, as argued in [8], because of phase periodicity and implied uncer-

tainty (a wrapped-around phase value φ in a transfer function might correspond to any value $\varphi + k \cdot 2\pi$ for integer k).

The idea behind representing HRTF filters as IIR filters [7] is of course to reduce the computational effort required for each simulation, which in turn opens up the possibility of performing more simultaneous auralizations for different users at the same time. Possible applications are sound installations where users walk around wearing headphones and position sensors, in such a way that each user receives an individualized soundscape according to its position and head orientation.

It would be correctly argued that low-order IIR models (i.e. with few poles and zeros) are not able to capture all the details of HRTFs (such as pinna cues), implying a loss of individualized audio fidelity. While this is certainly true, even such simplified models are a large improvement on simpler ITD+ILD (Interaural Time and Level Difference) models that would usually be employed in the aforementioned multi-user application context. For instance, the simplified HRTF models would allow the differentiation of sounds arriving directly from the front, from the back or from above the user [7], even though such directions usually share the same values for ITD and ILD.

The interpolation of IIR models of the same order may be accomplished by interpolating the coefficients of the filter equation. Although this procedure is a straightforward generalization from linear interpolation of FIR filters, there appear to be no previous discussions in the literature on the errors produced by this interpolation method in the context of IIR models for HRTF filters.

A completely different approach to HRTF interpolation can be obtained by using a z plane pole-zero representation of IIR filters, and manipulating the set of poles and zeros to change the filters from one configuration to the other [11, 12]. In [11] the manipulations of pole-zero configurations involve gradient searches where the target (interpolated) HRTF is known; this is a context of limited practical interest. The interpolation strategy defined in [12] depends on a given association between poles and zeros of different filters, and proceeds by interpolating the positions of associated poles and zeros to obtain intermediate pole-zero configurations.

The main difficulty in putting this idea into practice lies in defining pole-zero associations between the given data-

base filters. These associations tell the interpolation algorithm which pole (or zero) from filter A corresponds to a given pole (or zero) in filter B, thereby allowing the specification of intermediate configurations. Unfortunately there is no foolproof rule to create this kind of associations, although some ad hoc strategies [12] have been proposed.

We propose a new strategy for IIR interpolation, namely the use of *structured IIR filters*. The specific structure imposed on the IIR representations is based on a division of the frequency range in N frequency bands, and it requires that each frequency band contains exactly one pole and one zero. Each pole and each zero of a filter becomes thus immediately associated with the corresponding pole and zero of another filter via their common frequency bands. We show that HRTF filters may be approximated by structured IIR filters with errors that are comparable to those obtained by Kalman filter models [7]. We further show that structured IIR filters can be interpolated using filter coefficient interpolation and pole-zero interpolation with better results than the corresponding interpolation of Kalman filters.

The structure of this paper is as follows. Section 2 presents two methods to construct alternative IIR representations for each HRTF in the database, the Kalman method and the structured IIR method. Section 3 discusses two IIR interpolation strategies, filter coefficient interpolation and pole-zero interpolation, and section 4 presents and discusses numerical experiments regarding the fitness of the IIR models and of the interpolation methods. Finally, section 5 presents some conclusions and further work.

2. IIR MODELS FOR HRTFS

HRTF filters are usually obtained by direct measurements of the corresponding HRIR using contact microphones inside the ears of a person or a head doll [2]. Such a representation in the form of an impulse response is equivalent to the list of coefficients of a FIR filter. As the number of coefficients in this representation increases (with a fixed sample rate), the better the reproduction of subtleties that torso, head and pinna imprints in the sound signal as it enters the auditory canal.

While highly individualized HRTF filters may provide wonderful results in terms of auralization, they are computationally expensive to run on a multi-user setting, not to mention the difficulties in measuring a different HRTF database for each individual user. A reasonable alternative is the use of non-individualized HRTF datasets, from with users can extract useful directional information [14]. But even when adopting non-individualized HRTF databases such as CIPIC [2], no less than 200 filter coefficients are needed for each HRTF.

IIR (or recursive) filters are able to represent transfer functions with high level of detail and fewer coefficients [7] due to the feedback terms, which correspond to polynomial terms in the denominator of the transfer function. In the sequel two techniques for obtaining IIR approximations to FIR filters are presented.

2.1. The Kalman Method

One method for obtaining an IIR filter approximation [6, 7] is based on observing incoming values $x_0, x_1, x_2, \ldots$ and outcoming values $y_0, y_1, y_2, \ldots$ of a filter and trying to fit these observations to an IIR filter equation of the form

$$\hat{y}(n) = \sum_{j=0}^{N} a_j x(n-j) - \sum_{j=1}^{N} b_j y(n-j),$$

where the error $f(a,b) = \|y - \hat{y}\|$ is to be minimized over a given range of index values. This is a convex quadratic optimization problem with a unique minimizer, that can be directly computed by solving the first order necessary optimality condition ($\nabla f(a,b) = 0$). Denoting by

$$w(k) = (x_k, \ldots, x_{k-N}, y_{k-1}, \ldots, y_{k-N})'$$

the windowed joint input-output process, one obtains

$$\begin{bmatrix} a \\ -b \end{bmatrix} = R^{-1}c,$$

where $R = \sum_k w(k)w(k)'$ and $c = \sum_k y(k)w(k)$.

By letting x represent a Dirac's Delta function and y represent the HRIR of interest, the above method allows the computation of an IIR model that is close to the original HRTF in the sense that the errors between the original HRIR and the predicted values given by the above filter equation are minimized.

To obtain best results all HRTFs are time-aligned [10, 1, 12] so that all HRIRs start at time 0 (this is undone later in the interpolation phase). This time-alignment guarantees a minimum-phase property of all systems, thus avoiding a large row of zero coefficients to compensate for the initial time delay (for the arrival of the direct sound).

It should be noted that there is a catch in the above objective function: the measured errors relate the correct output values and the predicted output values, obtained from past input and *from correct past output values*. When this filter equation is applied to a new signal these latter values are not available, and the feedback terms in the equation are fed with *predicted past output values*. The consequence is that the minimal error obtained by this method does not correspond to the error between the original HRIR and the impulse response of the IIR filter, or to the error between the original HRTF and the transfer function of the IIR filter (these two errors are interrelated via Parseval's theorem).

2.2. Structured IIR Filters

The IIR model proposed here addresses both the error issue raised in the previous section as well as the difficulty

of making associations between poles and zeros of different filters raised in [12]. As mentioned in the introduction, this model requires that each frequency band in the z plane contains exactly one pole and one zero, in such a way that the error between the original HRTF and the corresponding IIR transfer function is minimized.

Consider the division of the upper half of the complex plane into N radial frequency bands with endpoints $\omega_0 = 0, \omega_1, \ldots, \omega_N = 2\pi$. The optimization problem has $2N$ complex variables $\pi_1, \ldots, \pi_N$ and $\zeta_1, \ldots, \zeta_N$ corresponding to the poles and zeros in the upper half of the plane, so that π_j and ζ_j belong to the band defined by the interval $[\omega_{j-1}, \omega_j]$. Each pole and zero has a corresponding conjugate dependent counterpart, so that the IIR filter actually has $2N$ poles and $2N$ zeros. In order to ensure compactness of the feasible region and to guarantee the existence of solution, each pole and zero is constrained to lie within the unit circle (it should be noted that poles lying on the unit circle would lead to unstable filters, but these are ruled out by the minimization procedure since the original FIR filters are always stable).

Let $\arg(\cdot)$ denote the argument or angle of a complex number, $*$ denote complex conjugation, and let H be the original HRTF. The complete optimization problem is thus formulated as

$$
\begin{cases}
\min & \|H - \hat{K}(H, \pi, \zeta)\hat{H}(\pi, \zeta)\|^2 \\
\text{s.t.} & \omega_{j-1} \leq \arg(\pi_j) \leq \omega_j \\
& \omega_{j-1} \leq \arg(\zeta_j) \leq \omega_j, \qquad i = 1, \ldots, N \\
& |\pi_j| \leq 1, \ |\zeta_j| \leq 1
\end{cases}
$$

where $\hat{H}(\pi, \zeta)$ is the transfer function given by

$$
\hat{H}(\pi, \zeta) = \frac{\prod_{j=1}^{N}(1 - \zeta_j z^{-1})(1 - \zeta_j^* z^{-1})}{\prod_{j=1}^{N}(1 - \pi_j z^{-1})(1 - \pi_j^* z^{-1})}.
$$

and the overall amplitude factor $\hat{K}(H, \pi, \zeta)$ is set to

$$
\hat{K}(H, \pi, \zeta) = \frac{\langle H, \hat{H}(\pi, \zeta)\rangle}{\|\hat{H}(\pi, \zeta)\|^2}
$$

in order to minimize $\|H - \hat{K}\hat{H}(\pi, \zeta)\|^2$, which is a convex quadratic expression on $\hat{K}$ for fixed H, π and ζ.

The radial division may be chosen according to the application context, for instance equal division ($\omega_j = j \cdot \frac{\pi}{N}$) or octave division ($\omega_0 = 0$, $\omega_j = \frac{\pi}{2^{N-j}}$, $j = 1, \ldots, N$). Other error measurements, such as the distance between magnitude transfer functions, which avoids the phase uncertainty problem referred to by [8], can also be used at discretion.

This optimization problem has a nonlinear and nonconvex objective function, and requires global optimization strategies such as multistart or branch-and-bound, combined with local descent strategies (Cauchy or Newton-like methods). In the experiments of section 4, Kalman solutions were modified to obtain feasible starting points, from which a multi-resolution first-order descent method was applied to obtain stationary points.

Although the computational cost of solving these problems for a whole HRTF database is high, this procedure has to be done only once at a setup stage, and so this cost has no impact whatsoever on the computational cost of auralization, which is always linear on N.

3. IIR HRTF INTERPOLATION

HRTF databases are built by measuring the incoming HRIR signal at each ear (of a head doll or a human) for a large number of possible incoming directions. The minimum angle between adjacent measured directions can be found by spatial sampling analysis [1] to be about $4.9°$ for a 44.1kHz sampling rate and $18cm$ distance between ears. The CIPIC database [2], for instance, uses 25 azimuth and 50 elevation angles, for a total of 1250 HRTF filters for each ear. Although this number of measurements is reasonably large, if a soundscape rendering of a moving sound source is needed, the need for an interpolation procedure becomes evident.

A moving sound source is defined by an output audio signal $x(n)$ and a spatial trajectory $\tau(n)$ that can be represented as polar coordinates (absolute value, azimuth and elevation). The absolute value in HRTF auralization is used for controlling the overall gain (using a distance-based rule such as inverse-distance gains or other psychoacoustical estimates), while the azimuth and elevation angles define the closest HRTF filters in the database that will be combined in the interpolation.

One of the simplest interpolation approaches is to always choose the filter corresponding to the closest measured direction, and to crossfade transitions to avoid discontinuities [12]. Unfortunately this is not good enough to ensure auralization smoothness if sound sources move slightly fast: with the CIPIC database a 1 second circular round-trip of a sound source would involve 50 crossfades/second, a transition rate which would produce audible artifacts due to phase shifts from one filter to the next.

The way to avoid this is to use adaptive filters whose coefficients are recomputed every sample or every block of B samples (comprising a few milliseconds). For instance, linear and bilinear interpolations [4, 12] are exactly adaptive crossfades between 3 (or 4) measured HRTFs that are closest to the direction of $\tau(n)$ at any given time, where the crossfade weights are linearly (or bilinearly) related to the relative position of the sound source within the triangle (or rectangle) defined by the original HRTF directions. By recomputing the filter coefficients at sample rate, the adaptive filtering will not introduce audible discontinuities in the signal, even with IIR models, provided that the spatial trajectory function $\tau(n)$ is smooth and slowly-varying.

Obtaining the weights in the triangular or rectangular case is a very simple geometrical problem. In triangular

interpolation, if the closest HRTFs have directions d^1, d^2 and d^3 relative to the user, and the virtual sound source has a normalized direction $\hat{d}$ (so that $\hat{d}$ lies in the triangle defined by d^1, d^2 and d^3), the weights w that satisfy $\hat{d} = w_1 d^1 + w_2 d^2 + w_3 d^3$ are given by

$$w = \left[\begin{array}{c|c|c} d^1 & d^2 & d^3 \end{array} \right]^{-1} \hat{d}.$$

In rectangular (or bilinear) interpolation [4], if the azimuth and elevation angles of the closest HRTFs are (α_j, β_k), for $j = 1, 2$ and $k = 1, 2$, and the azimuth and elevation angles of the virtual sound source is $(\hat{\alpha}, \hat{\beta})$, the corresponding weights will be

$$w_{jk} = \frac{|\alpha_{2-j} - \hat{\alpha}|}{(\alpha_2 - \alpha_1)} \frac{|\beta_{2-k} - \hat{\beta}|}{(\beta_2 - \beta_1)}$$

corresponding to the HRTF (α_j, β_k), for $j = 1, 2$ and $k = 1, 2$. These weights are used in a variety of interpolation contexts in the remaining of this section, and will be referred to as the *appropriate weights* with respect to the interpolation method (triangular or rectangular).

The choice between triangular or rectangular interpolation may be based on computational performance (triangular being roughly 25% faster because it combines 3 instead of 4 HRTFs) or on minimizing the number of transitions between different sets of HRTFs (rectangular subdivision of the database creates half the number of regions as does triangular subdivision).

Three adaptive interpolation approaches based on the IIR filter model are presented in the sequel. The first one uses the above weights to combine filter coefficients in the filter equation, and the remaining two use the weights to position poles and zeros in the z plane representation of the IIR filters obtained by the Kalman method or by the structured IIR approach.

3.1. Interpolation of Filter Coefficients

A first approach to interpolate IIR models of HRTF filters is to look at the IIR filter equation

$$y(n) = \sum_{j=0}^{N} a_j x(n-j) - \sum_{j=1}^{N} b_j y(n-j)$$

and to consider the coefficients a_j and b_j as functions of n, defined as linear combinations of the coefficients of the closest IIR models of measured HRTFs, using the appropriate weights as discussed above.

This idea is almost a direct translation of the usual linear interpolation of HRTF FIR filters, with the notable exception that in the FIR case the interpolation of the coefficients is the same as the interpolation of the impulse responses, and

also of the transfer functions (due to linearity of the z transform), whereas in the IIR case the interpolation affects simultaneously the numerator and denominator polynomials of the transfer function

$$H(z) = \frac{a_0 + a_1 z^{-1} + \cdots + a_N z^{-N}}{1 + b_1 z^{-1} + \cdots + b_N z^{-N}}$$

that corresponds to the given IIR filter equation. Since the interpolation of filter coefficients affects the position of the poles and zeros in highly nonlinear ways, there is no a priori guarantee that interpolated filters will always be stable. Unstable configurations were not observed in the numerical experiments (section 4), and they should not be the cause of troubles in the context of moving sound sources, where an unstable configuration would probably be soon replaced by a stable one.

3.2. Pole-Zero Interpolation (PZI)

The pole-zero interpolation takes the positions of the poles and zeros (zeros of the numerator and of the denominator of the transfer functions, respectively) of the IIR models of 3 or 4 HRTFs that are closest to the direction of the virtual source, and combines them linearly using the appropriate weights. Each pole (or zero) of the interpolated filter is obtained by a linear combination of exactly one pole (zero) from each of the 3 or 4 filters, where this set of matching poles (zeros) has been defined beforehand (this problem will be discussed in the sequel).

Consider the set of filters $\{F_j\}$ for $1 \leq j \leq 3$ (triangular case) or $1 \leq j \leq 4$ (rectangular case), defined by their pole-zero structure

$$F_j \leftrightarrow \{\pi_{jk}, \, \zeta_{jk}, \, k = 1, \ldots, N\}.$$

and overall gain factor K_j, that are supposed to be combined with appropriate weights w_j to create a new interpolated filter. By assuming that the poles and zeros are ordered according to their association, in such a way that $\pi_{1k}, \pi_{2k}, \pi_{3k}$ are corresponding poles and $\zeta_{1k}, \zeta_{2k}, \zeta_{3k}$ are corresponding zeros, the structure of the interpolated filter will be given by

$$\hat{F} \leftrightarrow \left\{ \begin{array}{l} \hat{\pi}_k = \sum_j w_j \pi_{jk}, \\ \hat{\zeta}_k = \sum_j w_j \zeta_{jk}, \, k = 1, \ldots, N \end{array} \right\}.$$

The interpolation of the overall gain factor could simply be computed as $\hat{K} = \sum_j w_j K_j$, but in many contexts this leads to poor interpolation results, because this expression assumes that the overall gain of the interpolated filter is similar to those of the closest filters used in the interpolation. This assumption often fails to hold, for instance with Kalman filters, which lack any kind of constraints on the position of poles and zeros. We propose to recompute the overall gain factor from the original HRTF data, as in section 2.2. In the case of interpolation, where the original HRTF is unknown, this means finding the best overall gain with respect

to the closest measured HRTFs. Specifically, if $\{H_j\}$ are the transfer functions used in the interpolation, we define

$$\hat{K}_j = \frac{\left\langle H_j, \hat{H}(\hat{\pi}, \hat{\zeta}) \right\rangle}{\|\hat{H}(\hat{\pi}, \hat{\zeta})\|^2},$$

and the interpolated overall gain factor as $\hat{K} = \sum_j w_j \hat{K}_j$.

In order to create associations of poles and zeros for different IIR filters one has to define exactly what it is that must be achieved with the association. In the case of interpolation one wants to define intermediate IIR filters, and so the structure of the interpolated poles and zeros must satisfy the requirement that for each complex pole (or zero) its complex conjugate is also a pole (zero) of the same filter. This is the same as saying that the filter coefficients (and the output filtered signal) have to be real-valued.

Even this minimal requirement already turns the association problem into an intrinsically difficult problem. Real poles (or zeros) may be readily associated with other real poles (zeros). When it comes to complex poles or zeros one is free to associate any members of different filters without violating this requirement, as long as the exact same associations are made with the corresponding complex conjugates. But there is no way of matching a filter with 2 zeros at $z = -1$ and $z = 1$, for instance, with another filter with 2 zeros at $z = -i$ and $z = i$. Any linear combination will produce a complex filter as a result of interpolation. Only under very special circumstances may a real pole (or zero) be associated with a complex pole (zero), namely when the multiplicity of the real pole or zero is even, and an association of a pair of overlapping real poles (or zeros) with a pair of complex poles (zeros) is made possible.

In the sequel this association problem will be addressed in the specific contexts of IIR filters produced by the Kalman method, and of structured IIR filters obtained by optimization as seen in section 2.2.

3.3. PZI Using Kalman IIR Filters

Since the Kalman method has no way of regulating the presence or multiplicity of real poles or real zeros, some preprocessing of the original filters is necessary in order to use the produced IIR database in an interpolation context. Specifically, since overlapping poles or zeros are usually not produced by this method, the number of real poles in any filter should be made equal, as well as the number of real zeros. Two possible strategies for this preprocessing are: (1) allow only complex poles and zeros in the filters; or (2) choose the most frequent number of real poles (and of real zeros) and force every non-conforming filter to adapt to the chosen number of real poles (and of real zeros). In both cases a search among modified versions of the filter is needed to enforce the chosen structure of poles and zeros.

Once a conforming IIR database is produced, the association of poles and zeros may proceed by considering separately the problem of associating real poles of one filter with real poles of another filter, and so on for real zeros, complex poles and complex zeros.

One association strategy was previously proposed in [12] using a minimum distance criterion. Consider for instance a set of complex poles $\pi_1, \pi_2, \ldots, \pi_M$ that must be associated with another set of complex poles $\rho_1, \rho_2, \ldots, \rho_M$. The proposed scheme is as follows: out of all possible permutations p of the indices $1, 2, \ldots, M$, choose the one that minimizes the overall distance $\sum_j |\pi_j - \rho_{p(j)}|$. This association is reflexive (that is, p^{-1} minimizes $\sum_j |\rho_j - \pi_{p^{-1}(j)}|$), but it lacks transitivity, meaning that the best association between filters A and C may be incompatible with the best associations between filters A and B and between filters B and C. This is a huge problem when trying to create associations between poles and zeros of 3 or 4 HRTF filters at a time (as required for interpolation).

We propose another association, based on frequency ordering. For every IIR model in a preprocessed database, sort all complex poles and zeros according to their angles, and use this ordering as the association for complex poles and zeros of different filters, matching the j-th complex pole (zero) of any filter with the j-th complex pole (zero) of any other filter. Correspondingly, sort all real poles and zeros according to their values, and use this ordering for association as above. This association satisfies reflexivity and transitivity, being a feasible substitute for the minimum distance approach.

3.4. PZI Using Structured IIR Filters

Structured IIR filters were proposed with the PZI strategy in mind, so the association of poles and zeros is already established through the association of each frequency band to exactly one pole and one zero. This is closely related to the frequency-based association just proposed for IIR filters obtained by the Kalman method.

It should be noted that real poles and zeros may occur in structured IIR filters, either on the extreme frequency bands (the first starting on 0 and the last ending on π) or in any other band (if the magnitude of the pole or zero is zero after optimization). When that happens, the corresponding filter has actually two coinciding poles (or zeros) produced by the same complex variable of the optimization model. During interpolation, if the spatial trajectory of the virtual sound source departs from a filter with a real pole in frequency band j and reaches another filter where the matching pole in the same frequency band is complex, the interpolation scheme will produce mirrored, complex-conjugate pole trajectories in frequency bands j and $2N - j + 1$ (on the lower half of the unit circle), ensuring that the intermediate filters will be real-valued, as they should be.

Even though all poles and zeros used in the interpolation are constrained within the unit circle, we still obtained best results by using the interpolation of the recomputed overall gain factors $\hat{K}_j$, as defined in section 3.2, instead of the interpolation of the original gains K_j.

4. NUMERICAL EXPERIMENTS

The methodology for building IIR databases and for using HRTF interpolation in binaural rendering presented in the previous sections may be applied to any HRTF database, regardless of being a tailored, individualized HRTF database or a generic, non-individualized HRTF database. To provide an objective evaluation of the methodology and of the interpolation procedure, we chose the CIPIC database [2] as an illustration.

The goal of this section is to objectively quantify the errors involved in the whole process. At first two sources of errors must be identified: the errors introduced by going from the original, measured HRTF filters to the simplified IIR models obtained by the Kalman method or the structured IIR approach, and then the errors that the interpolation procedure adds on top of the first errors.

While it is true that all errors are undesired regardless of their source, it should be noted that the first source of errors (of building a replacement IIR database) is part of a trade-off between audio fidelity and computational performance or, in other words, they are a price we are willing to pay for having more simultaneous users. The errors arising from the interpolation procedure are useful for comparing different approaches.

4.1. Impact on Performance

The first question we would like to answer regards the computational costs of using IIR models of various numbers of poles and zeros compared to the cost of using the original FIR filters. In this comparison we assumed that all filter parameters are recomputed at a sample-by-sample basis, so that all filters are computed according to their filter equations (in a context of *static* sound sources, the corresponding FIR filters might be computed using FFT convolution, but this is not the case here).

Figure 1 shows the the percentage of time required to compute each output sample of the IIR models of the HRTF filters for several N values, relative to the time required to compute the corresponding output sample of the FIR filter. This comprises the time spent in interpolating the filter coefficients to obtain the interpolated filter equation, as well as the time required to apply the filter equation to past input/output values. The second curve in the figure is the reciprocal of the first curve, representing the (relative) number of simultaneous users with IIR filters requiring the same computer power as a single user with a FIR filter. It can
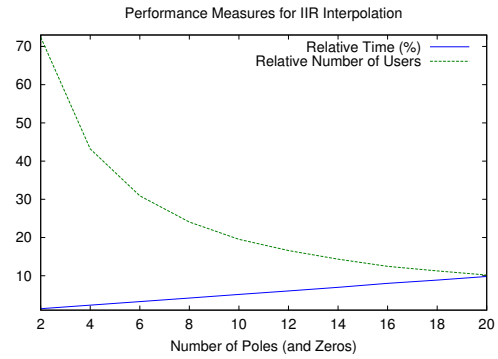

Figure 1. Relative times for computing IIR output for several N values, and corresponding increase factor for the number of users.

be seen that for $N = 20$ there is a ten-fold gain in computer time and number of simultaneous users allowed, and this only improves as smaller values of N are considered (with an implied increase in approximation errors, as will be seen in the sequel).

4.2. Fitness of the IIR Models

The second addressed question is exactly how many poles and zeros in the IIR models are necessary or desired in order to build the alternative database of filters. To help answer this question, figure 2 shows graphs of the relative errors of the filters obtained by the Kalman method and the structured approach, measured using magnitude transfer functions of the IIR and FIR versions, as functions of the number of poles and zeros used. Each graph shows the average relative error over all 1250 directions in the CIPIC database, as well as the minimum and maximum error values for each N.

It can be seen that both models have their quality improved as the filter orders increase, as should be expected, and that the structured approach manages to produce filters that are better than Kalman filters, on the average, up to 12 poles and 12 zeros, and not much worse after that (up to 20 poles and 20 zeros). Reasonably good fits start appearing for $N = 4$ with structured IIR models (minimum curve), and the error of the worst-case fits (maximum curves) also drop with N for both methods.

For any given value of N, the relative errors of the models as functions of azimuth and elevation angles are not uniformly distributed, but instead they are concentrated in certain extreme directions, such as lower-front-left and lower-back-right, as can be seen in the example of figure 3.

4.3. Fitness of the Interpolation Scheme

The next experiment aims at objectively evaluating the quality of the interpolation schemes presented in section 3. Since
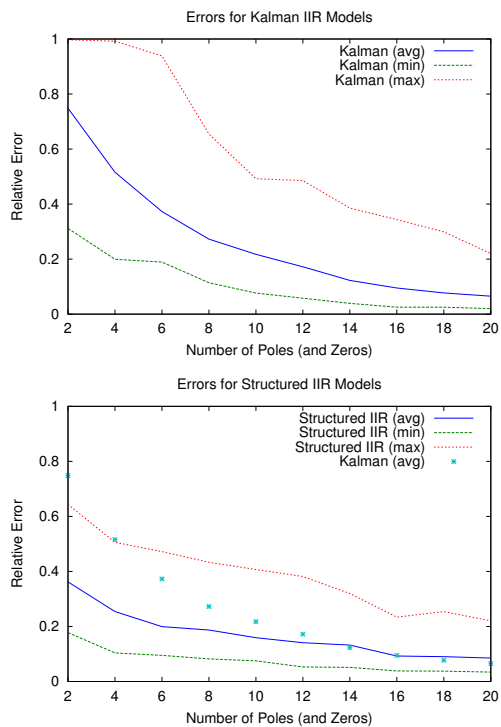
Figure 2. Relative errors of the Kalman filters (top) and Structured IIR filters (bottom) as functions of the number of poles and zeros.

Figure 3. Error of structured IIR model for left ear and N=16 as a function of azimuth and elevation angles.

Figure 4. Relative errors for the filter coefficient interpolation applied to Kalman filters and structured IIR filters, as functions of the number of poles and zeros.

it would be difficult to objectively assert anything about interpolated directions whose original HRTFs are not known, this experiment tries to recreate HRTFs for directions that do exist in the database, but interpolating from other close directions.

Figure 4 presents error graphs similar to those of figure 2, but referring to the relative errors of the filter coefficient interpolation. Unlike the modeling errors of the previous section, interpolation errors exhibit a nearly Gaussian distribution, and for this reason we present average and standard deviation values for each method.

The comparison of PZI applied to Kalman filters and structured IIR filters is shown in figure 5. As explained in section 3.3, PZI is only applicable to Kalman filters if the database is preprocessed, so that all filters have the same configuration of real poles and zeros. Here we forced all poles and zeros to become complex, by substituting every pair (A, B) of real poles (or real zeros) for two overlapping poles (zeros) at $(A + B)/2$, and performing a gradient search over the set of complex conjugate pole (zero) pairs in order to minimize the relative error with respect to the original HRTF. For Kalman PZI the errors represent the error contribution of both preprocessing and interpolation.
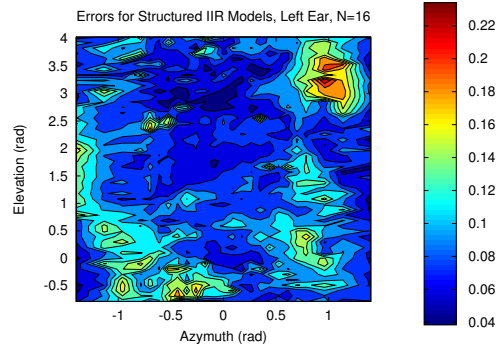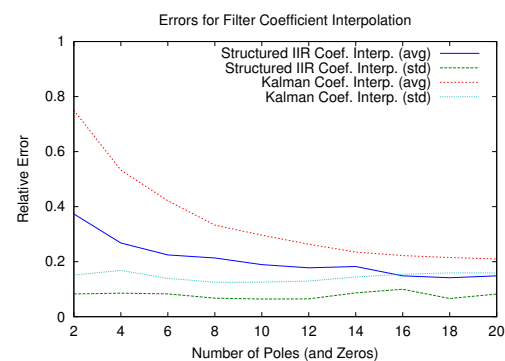
It can be seen that the relative errors of the structured IIR approach with filter coefficient interpolation for $N = 18$ is 14% of the norm of the corresponding HRTF filters, and for PZI it is 19% with $N = 16$, whereas the best values obtained by Kalman filters are 20% (with filter coefficient interpolation for $N = 16, 18, 20$) and 40% (with PZI for $N = 4, \dots, 12$).

To provide a more musical measure of what these relative errors mean, we also computed the differences between interpolated and measured filters on a decibel scale, as a function of frequency. The average error values (over all directions in the database) for structured IIR models with N=18 and filter coefficient interpolation are 0.93dB for frequencies up to 5kHz, 1.35dB between 5kHz and 10kHz, 1.87dB between 10kHz and 15kHz, and 3dB above 15kHz. This increase in errors as a function of frequency had already been observed by [1] when interpolating FIR HRTF filters, and its growth rate can be related to the angular sampling of the HRTF database.
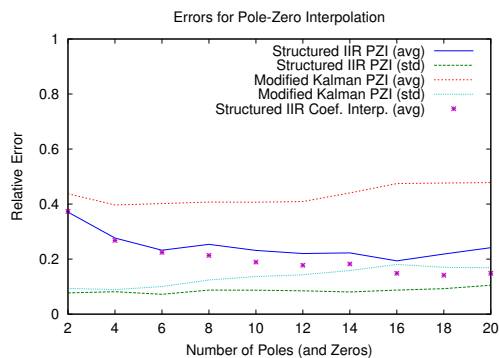
Figure 5. Increment in relative error for PZI applied to Kalman filters and structured IIR filters, as functions of the number of poles and zeros.

5. CONCLUSION

We presented in this paper a new class of IIR filters that are suitable for HRTF interpolation. We showed that these filters can be used in the context of filter coefficient interpolation and pole-zero interpolation for increasing the number of simultaneous users allowed by a factor of 10, with relative errors about 14% (or 19% for PZI) of the original HRTFs.

One of the possible directions of further research is to make the interpolation scheme more flexible by allowing different values of N in the same IIR database. The motivation for this lies in figure 2, where even for very small values of N it is possible to obtain good structured IIR models for *some* directions in the HRTF database. This flexibility might be accomplished by resorting to overlapping poles and zeros: whenever a pole lies on top of a zero, their contributions to the transfer function cancel out and the filter is equivalent to a filter with fewer poles and zeros.

The present experimental evaluation of the interpolation procedures was based on objectively quantifiable comparisons. These should be compared in the future to psychoacoustical experiments, which are of utmost importance in a context such as auralization where the last word comes from the user, namely whether he or she perceives the spatial trajectory of the virtual sound source as intended by the composer or sound designer.

6. REFERENCES

[1] T. Ajdler, C. Faller, L. Sbaiz, and M. Vetterli, "Interpolation of head related transfer functions considering acoustics," in *Proc. 118th AES Convention*, Barcelona, 2005.

[2] R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, 2001.

[3] C. Cheng, "Moving sound source synthesis for binaural electro-acoustic music using interpolated head-related transfer functions (HRTFs)," *Computer Music Journal*, vol. 25, no. 4, pp. 57–80, 2001.

[4] F. P. Freeland, L. W. P. Biscainho, and P. S. R. Diniz, "Efficient HRTF interpolation in 3D moving sound," in *Proceedings of the AES 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*, Espoo, 2002.

[5] K. Hartung, "Comparison of different methods for the interpolation of head-related transfer functions," in *Proceedings of the 16th AES International Conference on Spatial Sound Reproduction*, Rovaniemi, 1999.

[6] R. E. Kalman, "Design of a self-optimizing control system," *Trans. ASME*, vol. 80, pp. 468–478, 1958.

[7] A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 115, no. 4, pp. 1714–1728, 2004.

[8] V. Lazzarini and B. Carty, "New csound opcodes for binaural processing," in *Proc. 6th Linux Audio Conference*, Köln, 2008.

[9] R. Martin and K. McAnally, "Interpolation of head-related transfer functions," Air Operations Division, Department of Defence, Australian Government, Tech. Report DSTO-RR-0323, 2007.

[10] M. Matsumoto, S. Yamanaka, and M. Tohyama, "Effect of arrival time correction on the accuracy of binaural room impulse response interpolation," *J. Audio Eng. Soc.*, vol. 52, no. 1/2, pp. 56–61, 2004.

[11] P. R. Runkle, A. Blommer, and G. Wakefield, "A comparison of head related transfer function interpolation methods," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, 1995.

[12] G. H. M. Sousa and M. Queiroz, "Two approaches for HRTF interpolation," in *Proc. 12th Brazilian Symposium on Computer Music*, Recife, 2009.

[13] L. Wang, F. Yin, and Z. Chen, "An indirect interpolation method for head-related transfer function pole-zero models," *Acoust. Sci. & Tech.*, vol. 29, no. 5, pp. 329–331, 2008.

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