Two approaches for HRTF interpolation

Gustavo H. M. de Sousa[1]*, **Marcelo Queiroz**[1]†

[1]Computer Science Department
University of São Paulo
R. do Matão 1010
05508-900 São Paulo, SP

gustavo@ime.usp.br, mqz@ime.usp.br

Abstract. *This paper deals with the problem of interpolating HRTF filters for the binaural simulation of continuously moving sound sources. Two novel approaches are presented, one based on interpolating impulse responses in time-domain, and another based on interpolating poles and zeros of low-order IIR approximations of measured HRTF filters. Computational tests show that the proposed interpolation schemes are feasible for processing anechoic signals with varying directional information without audible discontinuities (clicks). We present an objective comparison of relative errors of interpolated HRTFs with respect to measured ones, and discuss some of the difficulties of evaluating the quality of the results from both objective and subjective points-of-view.*

Resumo. *Este artigo lida com o problema de interpolar filtros HRTF para a simulação binaural de fontes sonoras em movimento. Duas abordagens inovadoras são apresentadas, uma baseada na interpolação de respostas impulsivas no domínio do tempo, e outra baseada na interpolação de pólos e zeros de aproximações IIR de baixa ordem dos filtros HRTF. Testes computacionais mostram que os esquemas de interpolação propostos são viáveis para o processamento de sinais de áudio anecóicos com informação direcional variável sem introduzir descontinuidades audíveis (clicks). É apresentada uma comparação objetiva dos erros relativos das HRTFs interpoladas com respeito àquelas medidas, e são discutidas algumas das dificuldades em avaliar a qualidade dos resultados dos pontos de vista objetivo e subjetivo.*

1. Introduction

Research in sound spatialization is becoming increasingly important since the second half of the 20th century, from both technological and artistic points-of-view [Roads, 1996]. Its applications include immersive systems and virtual reality, computer games, and musical composition. In this latter field of application sound spatialization becomes an added dimension of artistic experimentation [Stockhausen, 1961, Xenakis, 1992].

A very simple and widespread multichannel sound spatialization technique is called *amplitude panning*, which creates a soundscape by changing the relative amplitude of the signals of each channel [Moore, 1990]. A straightforward extension of this technique is Vector-Based Amplitude Panning or VBAP, which allows for 3D loudspeaker configurations [Pulkki, 2001]. These techniques are intended for loudspeaker reproduction of a sound field; they pressuppose that the listener is seating at the center of the loudspeaker configuration space, and they have very limited results outside this privileged spot (or *hot spot*).

*Supported by CAPES and FAPESP.
†Supported by CNPq.

One alternative to sound field recreation through loudspeakers is the use of binaural simulation through headphones. Much more control can thus be gained over the subjective impression of spatial location of sound sources, and more flexibility on the number of simultaneous users in various different locations, each receiving its own individual soundscape, at the expenses of having to wear a pair of headphones and carry some wireless receiver. Nevertheless, it is a promising path for sound spatialization that imposes no budget or setup burdens.

Although simple schemes like amplitude panning and *Interaural Time Differences* (ITD) have been applied to binaural spatialization, much more realistic results may be obtained through the so-called HRTF filters (*Head-Related Transfer Functions*), which are usually obtained by directly measuring the effect of an incoming signal inside one person's ears (using tiny contact microphones), or by using a dummy head with internal microphones [Algazi et al., 2001]. These filters capture all effects of sound propagation from the original sound source to the listener's ears that modify an incoming signal, including interaural level and phase differences, absorption, reflections and refractions on the environment and around the body of the listener. Among these effects, those resulting from the reflections and refraction of sound on the torso, head and pinnae have a profound impact on the subjective impression of direction [Blauert, 1997, Hofman et al., 1998].

Each HRTF filter (or HRTF for short) is by definition tied to a particular direction of incoming sound, as specified during its measurement. It would be therefore necessary to have one measured HRTF for each direction of incoming sound one would like to represent in a simulation. Several databases of recorded HRTFs exist, such as the CIPIC database [Algazi et al., 2001], that represent large, although finite sets of incoming directions that may be readily used in auralization of static sound sources.

As opposed to that, we are interested in allowing virtual sound sources to describe arbitrary spatial trajectories, that should be subjectively perceived through headphones. In order to change smoothly from one HRTF to another, it is necessary to have interpolation schemes for these filters that satisfy two conditions: on one hand, no audible discontinuities (clicks) should be heared during transitions, and on the other hand, interpolated HRTFs should match as closely as possible the corresponding interpolated directions.

It should be noted that although this problem might seem an easy one at first glance, it remains unsolved in PureData and was only recently addressed in Csound, two very widespread and massively used sound processing computational environments; both Csound's opcode *hrtfer* and Puredata external *earplug~* produce clicks when a pure sinusoidal signal is made to wander around the listener's head; a newer Csound opcode *hrtfmove* uses crossfades to eliminate artifacts when switching between HRTF filters.

One interpolation scheme for HRTF filters, known as bilinear interpolation [Savioja et al., 1999, Freeland et al., 2002], aims at expanding the database of measured HRTF by constructing HRTFs for intermediate positions, so-called IPTFs or *Inter-Positional Transfer Functions*, from the available measured HRTFs. These are computed as linear combinations of four adjacent HRTF filters corresponding to a square around the desired direction (we use the term adjacent filters in the sense of filters corresponding to adjacent directions in a finite HRTF database).

We propose two new approaches for HRTF interpolation. The *triangular interpolation* is a small improvement over bilinear interpolation; it combines linearly three HRTFs corresponding to a triangle around the desired direction, with a 25% computational gain for obtaining each interpolated HRTF. The *spectral interpolation*, on the other hand, differs fundamentally from previous approaches in the sense that the HRTF database is substituted by a database of low-order IIR filters that approximate the original HRTFs, computed prior to simulation. Each filter is represented by a set of 6 poles and 6 zeros in the complex plane, that directly affect regions of resonance and antiresonance in the frequency response. By interpolating the positions of corresponding poles and zeros in adjacent filters, we construct intermediate filters that are not linear combinations of adjacent ones, but rather nonlinear combinations that are inherently tied to the spectral shapers (i.e. poles and zeros) of an IIR filter.

Section 2 presents the details of VBAP and the derived triangular interpolation. Section 3 presents the Kalman filter method for obtaining low-order IIR aproximations for the original HRTFs, a technique for matching poles and zeros of adjacent IIR representations, and the spectral interpolation technique for moving sound sources. The implementation of these methods, as well as computational experiments and their discussion are presented in section 4, and some conclusions and further work are presented in section 5.

2. Triangular Interpolation

The original motivation for the triangular interpolation technique was a transposition of the VBAP technique from its original context (amplitude panning over loudspeakers) to a binaural application, where each HRTF-auralized signal would be treated as one of the loudspeakers in order to compute the linear combination coefficients. We will first present the original VBAP technique in order to introduce formally the triangular interpolation.

2.1. Vector-Based Amplitude Panning

The VBAP technique, as mentioned earlier, is a special case of amplitude panning. It aims at recreating the subjective impression of (i.e. positioning) a virtual sound source by sending the same signal over several fixed loudspeakers, each with a different amplitude gain. If $x(t)$ is the original signal, loudspeaker n will play $x_n(t) = g_n \cdot x(t)$, for $n = 1, \ldots, N$, where g_n is the corresponding gain [Pulkki, 2001, Moore, 1990]. These gains depend on the position of each loudspeaker and of the virtual source.

In a tridimensional setting it is customary to consider sets of 3 adjacent loudspeakers forming a triangular cone centered on the listener, that are used for simulating virtual sound sources within the triangle. Let $\boldsymbol{l^n}, \boldsymbol{l^m}$ and $\boldsymbol{l^k}$ be the vectors corresponding to the direction of each loudspeaker relative to the listener, and $\boldsymbol{p}$ the direction of the virtual sound source. Then the gains g_n, g_m and g_k of the loudspeaker must satisfy $\boldsymbol{p} = g_n\boldsymbol{l^n} + g_m\boldsymbol{l^m} + g_k\boldsymbol{l^k}$ or, in matrix notation:

$$\boldsymbol{p}^t = \boldsymbol{g}\boldsymbol{L}_{nmk},$$

where $\boldsymbol{g} = [g_n g_m g_k]$ and $\boldsymbol{L}_{nmk} = [\boldsymbol{l^n l^m l^k}]^t$. The solution is therefore

$$\boldsymbol{g} = \boldsymbol{p}^t\boldsymbol{L}_{nmk}^{-1} = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} l_1^n & l_2^n & l_3^n \\ l_1^m & l_2^m & l_3^m \\ l_1^k & l_2^k & l_3^k \end{bmatrix}^{-1}.$$

The VBAP technique states that the subjective impression of a virtual sound source coming from diretion $\boldsymbol{p}$ is recreated by applying gains g_n, g_m and g_k to the loudspeakers, according to the above equation.

2.2. Using VBAP gains to interpolate HRTFs

Given an input signal $x(t)$, and a measured HRTF corresponding to a direction $\boldsymbol{l}$, given by a finite impulse response $h(n)$, the spatialized signal is obtained by the convolution

$(x*h)(t)$ [Moore, 1990]. The triangular interpolation scheme correponds to treating each spatialized signal $(x*h)(t)$ as coming from a fixed loudspeaker placed at $\boldsymbol{l}$, and applying the VBAP gains to those signals before mixing them in the binaural simulation.

If $\boldsymbol{l^n}, \boldsymbol{l^m}$ and $\boldsymbol{l^k}$ are the vectors corresponding to the direction of each measured HRTF, and $\boldsymbol{p}$ is the intended direction of the virtual sound source, then the gains given by $\boldsymbol{g} = \boldsymbol{p}^t \boldsymbol{L}_{nmk}^{-1}$ will be applied to the signals $(x*h_n)(t)$, $(x*h_m)(t)$ and $(x*h_k)(t)$, leading to an interpolated signal

$$y(t) = g_n \cdot (x*h_n)(t) + g_m \cdot (x*h_m)(t) + g_k \cdot (x*h_k)(t).$$

This is supposed to be done independently for each ear in a binaural simulation (each HRTF actually corresponds to a given direction and one of the ears of the listener).

By using the linearity property of convolution, it is easy to see that

$$\begin{aligned} y(t) &= g_n \cdot (x*h_n)(t) + g_m \cdot (x*h_m)(t) + g_k \cdot (x*h_k)(t) \\ &= (x * [g_n \cdot h_n + g_m \cdot h_m + g_k \cdot h_k])\,(t), \end{aligned}$$

which shows that the above interpretation is equivalent to combining the HRTFs directly with the VBAP gains, obtaining the interpolated HRTF

$$\hat{h} = g_n \cdot h_n + g_m \cdot h_m + g_k \cdot h_k,$$

and then applying the convolution. The latter strategy corresponds closely to bilinear interpolation [Savioja et al., 1999], but using 3 adjacent HRTFs instead of 4. The triangular approach cuts down 25% of the computational cost of evaluating $y(t)$, regardless of which order of computation is chosen. For direct implementation, the first equation saves about M multiplications, where M is the size of the impulse responses; by using fast convolution on blocks of M samples, the second equation performs better, by requiring a single fast convolution (cost $M \log M$) instead of three.

This method may be applied on a sample-by-sample basis, meaning that the gains applied on each convoluted signal are updated at the audio rate. Since smoothly moving sound sources are bound to update their directions at a much slower rate, all gains will also behave smoothly, and that guarantees that no audible discontinuities will be introduced by the interpolation method. If one decides to update the gains at a slower control rate, this will not be a problem as long as the speed of the virtual sound source does not exceed a certain threshold, related to the rate of filter switching entering the audible range. One solution to the slow update implementation is to lowpass-filter the control signal that represent the directions of the moving sound source, thus forcing the increments of azimuth and elevation to be small compared to the control rate.

The triangular interpolation technique is essentially a time-domain technique, in the sense that the waveforms of impulse responses are combined and time-domain convolution is computed. Due to the linearity of the Fourier Transform, it is possible to apply the same interpolation procedure to the complex transfer functions representing the HRTF filters, using the same gains computed by the VBAP method, and substituting convolution by multiplication of spectra. This is an alternative implementation that does not alter the method at a conceptual level. In section 4 we will discuss important implementation details, such as dealing with phase differences of the HRTFs, and the effects of these in the comparison of interpolated HRTFs with measured ones.

3. Spectral Interpolation

In the previous section we mentioned that the triangular interpolation may be implemented either in time-domain or frequency-domain, by applying linear gains to the impulse responses or to the transfer functions representing the HRTFs. In both cases we

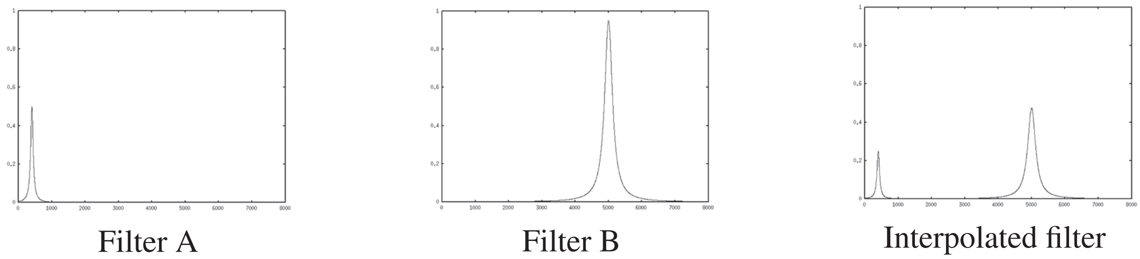

Figure 1: two bandpass filters and a linear interpolation.

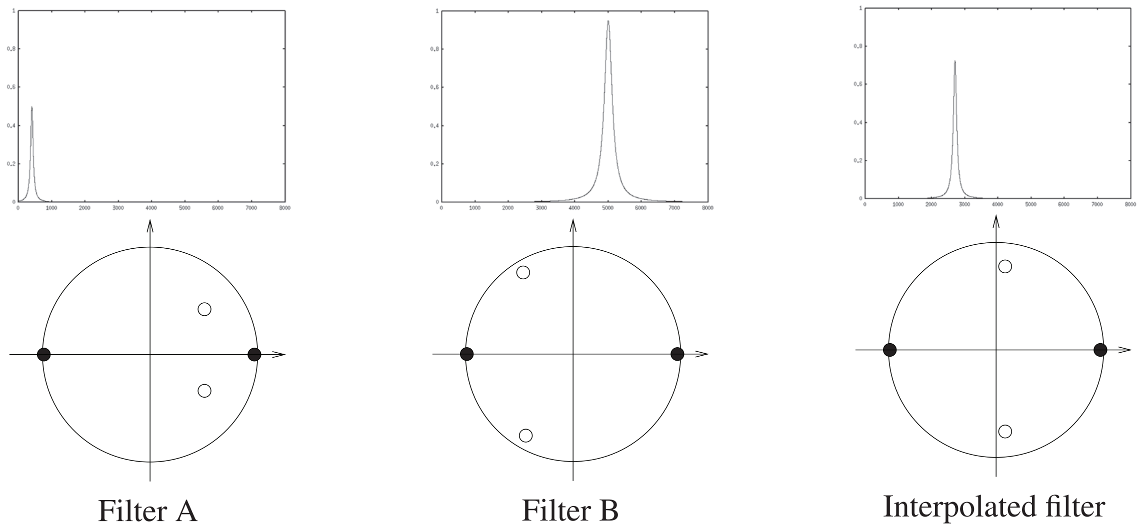

Figure 2: pole-zero diagrams and spectral interpolated filter.

were dealing with *linear combinations* of waveforms, impulse responses, or spectra. The spectral interpolation technique presented here aims at something fundamentally different, namely shaping the spectra of interpolated HRTFs in such a way that the regions of resonance and antiresonance are also "interpolated" (in a weaker sense, to be made precise in the sequel).

To illustrate the idea of spectral interpolation, consider a family of bandpass filters, defined by center frequency and bandwidth. We might consider a smooth transition from filter A to filter B in figure 1. With linear interpolation (of impulse responses, or equivalently of spectra) the intermediate filters wouldn't look like a bandpass filter at all (see the third plot in figure 1). By representing these filters as 2-poles-2-zeros IIR filters, it is natural to define an intermediate filter by placing the poles at an intermediate frequency (angle in the polar representation), and an intermediate distance from the origin (magnitude in the polar representation). This corresponds exactly to interpolating the polar representations of the poles in the pole-zero diagram (see figure 2).

IIR filters are well-known for their compactness: with few feedback coefficients they are able to encode very complex filter responses, thus making them much more efficient than FIR filters in terms of processing time. For spatialization purposes, this means being able to use more spatialization units in real-time and in parallel.

In order to extend this idea to HRTFs we first need to discuss how to obtain pole-zero representations for measured HRTFs. Then we need to address the problem of identifying corresponding pairs of poles or zeros in adjacent IIR filters. Finally we discuss the interpolation of poles and zeros and the use of intermediate IIR filters in the simulation of moving sound sources.

3.1. Obtaining low-order IIR filters

HRTF filters are usually found in the form of HRIR or Head-Related Impulse Responses, which are time-domain signals with a duration of a fraction of a second; the CIPIC database [Algazi et al., 2001], for instance, features recordings consisting of 200 samples on a 44.1kHz sampling rate. Although really small, these samples may be viewed as coefficients of a 200-zeros FIR filter, which is a lot of information in the pole-zero complex plane to deal with. In order to be able to manipulate the positions of poles and zeros we are bound to sacrifice precision in the representation of the filter and use low-order approximations (i.e. with fewer zeros and/or poles).

The Kalman method [Kalman, 1958, Kulkarni and Colburn, 2004] is designed to achieve optimal IIR approximations for a given FIR filter, in the sense that it minimizes the squared error of the approximation over all possible IIR filters of the same order. Suppose we want to approximate a given HRIR $y(n)$ using an IIR filter with P poles and Q zeros, corresponding to the filter equation

$$\hat{y}(n) = \sum_{i=0}^{Q} a_i x(n-i) - \sum_{j=1}^{P} b_j \hat{y}(n-j).$$

The Kalman method minimizes the squared norm of the approximation error

$$\sum_{k=0}^{M} e(k)^2,$$

where $e(k) = \hat{y}(k) - y(k) = \sum_{i=0}^{Q} a_i x(k-i) - \sum_{j=1}^{P} b_j \hat{y}(k-j) - y(k)$. This unconstrained convex quadratic optimization problem has a closed-form solution that is given by

$$\begin{bmatrix} a_o \\ \vdots \\ a_Q \\ -b_1 \\ \vdots \\ -b_P \end{bmatrix} = W^{-1} \left(\sum_{k=0}^{M} y(k) w(k) \right),$$

where $w(k) = (x(k), \ldots, x(k-Q), y(k-1), \ldots, y(k-P))^t$, for $k = 0, \ldots, M$, and the matrix W is a sum of the outer products $W = \sum_{k=0}^{M} w(k) w(k)^t$.

With this method and by choosing a small value for P and Q we are able to convert a complete database of FIR HRTF filters into low-order IIR filters that can be used in the spectral interpolation method.

3.2. Matching poles and zeros

One of the difficulties that appear as we try to interpolate pole-zero diagrams is to assign matching pairs of corresponding poles and corresponding zeros for each group of adjacent IIR filters. On one hand, typical databases such as the CIPIC are not dense enough to ensure that the Kalman method will produce sufficiently close diagrams for adjacent filters. This means that it is not always obvious how to transform one pole-zero diagram into another because we don't always know which pole in the first diagram goes to which pole in the second diagram (see figure 3 for an example). On the other hand, by making careless assignments of pole pairs we risk losing stability of the intermediate filters, as well as producing sound artifacts due to rapidly changing frequency responses.

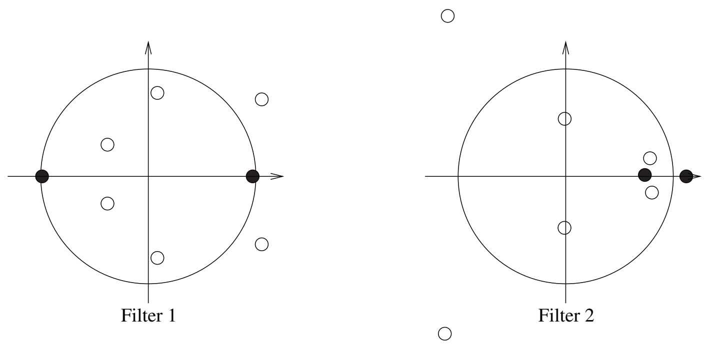

Figure 3: Two adjacent filters represented by pole-zero diagrams.

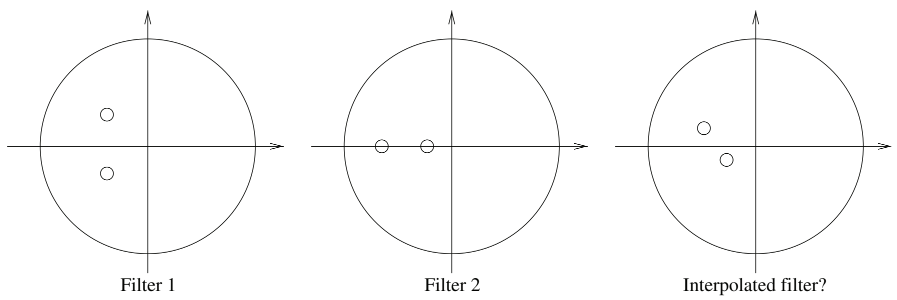

Figure 4: Interpolating from complex towards real poles.

In the absence of a reliable perceptual guideline for assigning such pairs, we adopted a risk-averse strategy, that corresponds to a minimum energy or least effort solution: to select the matching of poles and zeros that minimizes the overall motion between corresponding pairs. This is easy to compute for low-order filters; it suffices to consider all possible permutations of the first set of poles and zeros and evaluating the total distance to the corresponding poles and zeros of the second set. By remembering that all complex poles and zeros come in complex-conjugate pairs, the actual size of the sets where the permutation is applied may be halved.

A second difficulty in this matching is the treatment of poles and zeros that lie on the real line. These are actually the most difficult to deal with, because one cannot safely interpolate between a complex pole and a real one without destroying the complex-conjugate property of complex poles (see figure 4). Forcing the interpolated filter to assume such configurations would produce a filter equation with complex coefficients, which delivers a complex time-domain signal. The only way out is to treat complex poles and real poles as two separate categories (as we were treating poles and zeros as separate categories), and assigning complex poles in one filter to complex poles in the other (and likewise with real poles).

What this solution implies is that we must determine beforehand not only the number of poles and zeros, but also the number of real poles and the number of real zeros in the pole-zero representation. This imposes a constraint on the creation of IIR filters that the Kalman method alone is not able to enforce. One way to cope with this difficulty is to choose the more frequent pole-zero structure in the IIR filters produced by the Kalman method, and then adjust all non-conforming filters to this structure. This involves replacing pairs of real poles for complex poles and vice-versa, depending on the specific case. Substitutions such as those may be implemented by direct search of conforming filter candidates in the neighborhood of the problematic filter.

3.3. Using the interpolated filters

By using Kalman filter approximations to the original HRTF databases, and having a reasonable matching of poles and zeros for adjacent filters, we are now able to discuss the use of the spectral interpolation method for the spatialization of moving sound sources.

Suppose we want to simulate a direction $\boldsymbol{p}$, that lies in a triangular region of the HRTF database defined by the directions l_n, l_m and l_k, and let g_n, g_m and g_k be the gains computed by the VBAP method (section 2). Let p_l^j and q_l^j be the j-th pole and j-th zero of the HRTF $l \in \{n, m, k\}$, and suppose that their polar representations are $p_l^j = \alpha_l^j e^{i\varphi_l^j}$ and $q_l^j = \beta_l^j e^{i\omega_l^j}$. We define the j-th pole and j-th zero of the interpolated IIR filter as

$$\hat{p}^j = (g_n\alpha_n^j + g_m\alpha_m^j + g_k\alpha_k^j)e^{i(g_n\varphi_n^j + g_m\varphi_m^j + g_k\varphi_k^j)}$$

$$\hat{q}^j = (g_n\beta_n^j + g_m\beta_m^j + g_k\beta_k^j)e^{i(g_n\omega_n^j + g_m\omega_m^j + g_k\omega_k^j)}.$$

Each IIR filter also has an overall gain coefficient a_0; we accordingly define the interpolated coefficient as $\hat{a}_0 = g_n a_{0,n} + g_m a_{0,m} + g_k a_{0,k}$.

This set of poles and zeros define a transfer function given by

$$\hat{H}(z) = \hat{a}_0 \frac{\prod_{j=1}^{Q}(1 - \hat{q}^j z^{-1})}{\prod_{j=1}^{P}(1 - \hat{p}^j z^{-1})},$$

which may be rewritten as

$$\hat{H}(z) = \frac{\hat{a}_0 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2} + \cdots + \hat{a}_Q z^{-Q}}{1 + \hat{b}_1 z^{-1} + \hat{b}_2 z^{-2} + \cdots + \hat{b}_P z^{-P}},$$

from which the coefficients for the filter equation

$$\hat{y}(n) = \sum_{i=0}^{Q} \hat{a}_j x(n-j) - \sum_{j=1}^{P} \hat{b}_j \hat{y}(n-j)$$

are readily available.

This method may be likewise implemented in a sample-by-sample update basis, and so the same comment of section 2.2 regarding absence of audible discontinuities apply. If the virtual direction varies smoothly, so do the VBAP gains and therefore all poles and zeros also vary smoothly, and since the filter equation coefficients are continuous functions of the poles and zeros, the output signal will not be affected by audible discontinuities. If one so desires, a slower control rate for updating the filter equation may be employed, provided that the motion of the virtual sound source is slow (or lowpass-filtered) compared to the filter switching rate.

4. Implementation and Discussion

We will discuss in the sequel some implementation details for each of the proposed methods, and also some computational experiments that we made to assess the quality of the interpolations.

4.1. Implementation of the triangular interpolation and discussion

One of the issues that appeared in the implementation of the triangular interpolation method had to do with phase differences between adjacent HRTF filters. This is a sneaky and yet important detail, because it severely affects the interpolated filters.

We used mainly the CIPIC database [Algazi et al., 2001] as input for our experiments. In those recordings, the initial silent gap does not correspond to the distance from the sound source to the ear of the dummy head, but rather it is freely adjusted in order to preserve the main characteristics of each HRIR waveform. That means that adjacent HRIR may have quite different onset times of the direct sound, and thus cannot be carelessly combined. We alleviate that problem by synchronizing onset times of adjacent HRIRs that were going to be used in a triangular interpolation, applying gains and mixing, and only then adjusting the overall delay based on a geometric model of the CIPIC recording settings.

In order to try to objectively evaluate the quality of this interpolation scheme, we considered an automatic experiment where original, measured HRIRs corresponding to given directions would be compared to interpolated versions, for the same directions, obtained from adjacent HRIRs. For each possible HRTF in the database, corresponding to a certain azimuth and elevation angles, we tried to simulate this HRTF by interpolating HRTFs in adjacent azimuth and elevation angles. More specifically, if D is point in the database, we chose triangles ABC containing D, such that A, B, and C are direct neighbors of D. Here all HRTFs for positions A, B, C and D are known, and an interpolated version D' is computed from A, B, C and the corresponding VBAP gains. We then plot a graph of the magnitude of the frequency responses of D and D' (measured in dB), for selected frequencies, as a function of azimuth and elevation angles (see figure 5).

By comparing each pair of graphs in figure 5 one sees that the attenuation patterns of the original HRTFs (left column) are roughly preserved in the interpolated HRTFs (right column). The above frequencies were chosen for illustration purposes only, but as a rule the same preservation of patterns is observed for all frequencies. This observation supports the claim that the interpolation scheme preserves frequency-dependent directional information of the HRTFs.

We have also compared the same pairs of original and interpolated HRTFs by considering the relative errors given by $\frac{\|HRTF(D)-HRTF(D')\|}{\|HRTF(D)\|}$. This measure subsumes attenuation differences for all frequency components at once, but it does not behave very smoothly over the whole database. This is explained by the fact that the original HRTFs also do not behave smoothly on the whole range of possible azimuth and elevation angles, i.e. they vary more rapidly on some regions, and this might explain the difficulty of the interpolation scheme to produce close results in the sense of obtaining similar waveforms by using neighboring HRTFs.

An interesting issue is the relation of the relative errors with the size of the triangles used in the above experiment. We compared the errors of the above experiment with a new setting were points A, B, and C surrounding D were allowed to be neighbors of the neighbors of D (they had distance 2 in terms of database points). It came as no surprise that the errors in the new setting were always larger than the ones in the original experiment. This might indicate that the actual errors of using the triangular interpolation within the small triangles defined by the points in the CIPIC database is smaller than what we have obtained in the numerical experiment. This observation underlines the importance of the density of the database as expressed by the number of directions included in the measurements: the more thinly spaced these measurements are made, the better the resulting spatialization of a moving sound source.

4.2. Implementation of the spectral interpolation and discussion

As discussed in section 3, having a database of low-order IIR aproximations of the original HRTF database is a prerequisite for applying the spectral interpolation. We therefore applied the Kalman method on the CIPIC database, obtaining IIR filters of 6 poles and

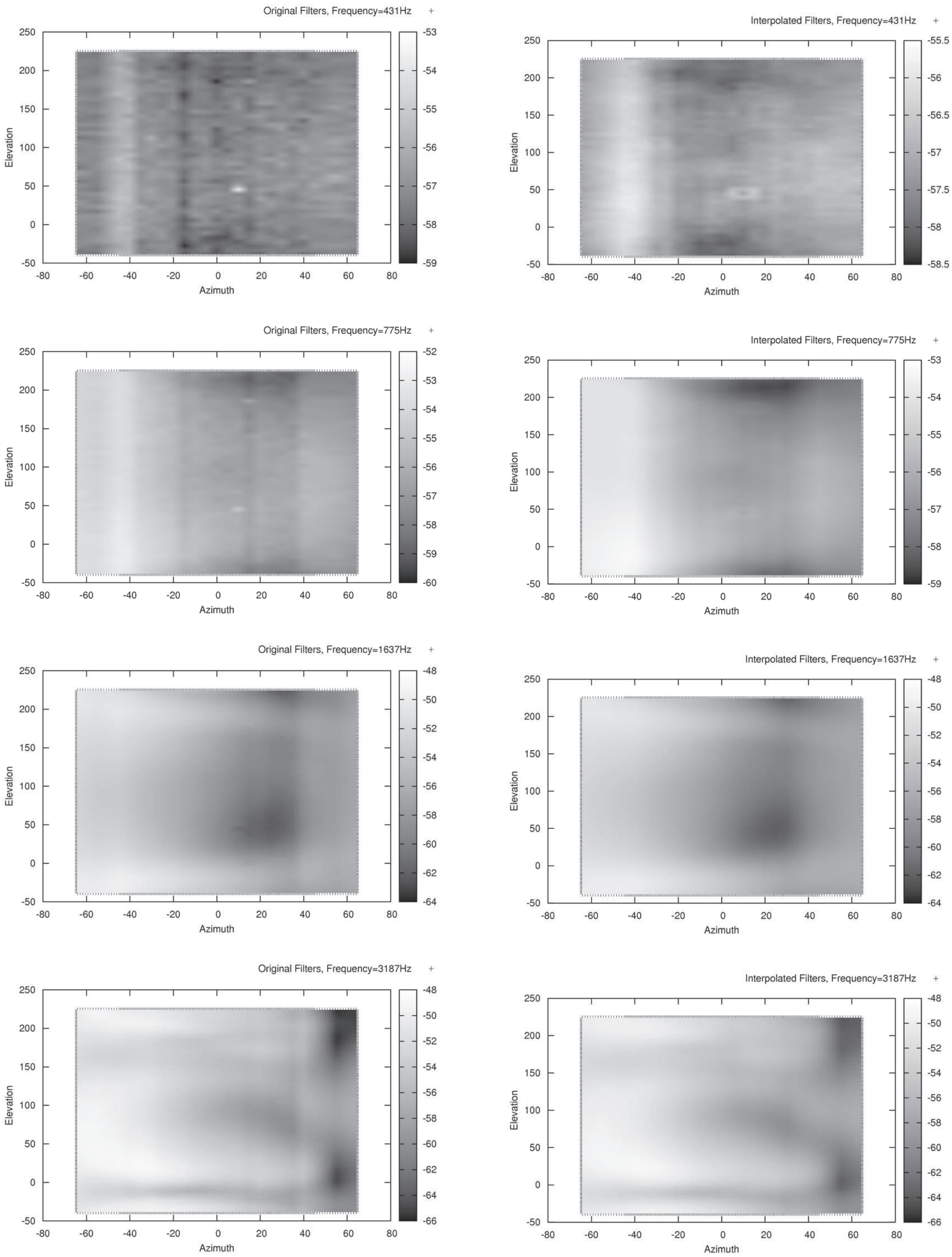

Figure 5: Comparison between original and interpolated HRTFs for selected frequencies. Each graph shows the pattern of attenuation for a given frequency as a function of the direction of incoming sound.

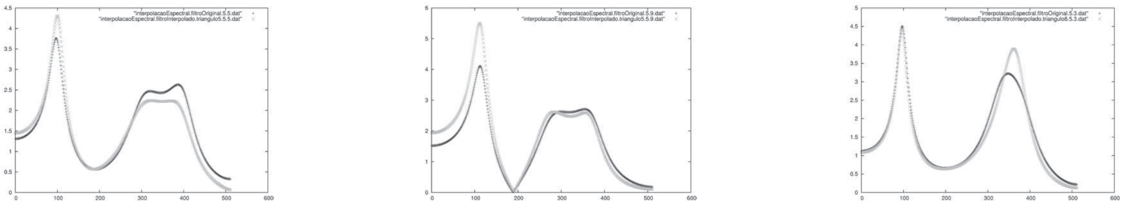

Figure 6: Measured and interpolated HRTFs.

6 zeros each. The first difficulty here was the structure of real and complex poles, as discussed in section 3.2. Since most of the Kalman filters had a similar structure of 4 complex poles and 2 real poles, we forced all non-conforming filters into this common structure.

With respect to an objective evaluation of the quality of the interpolation, it must be made clear that there are two different kinds of errors involved. First, there are the errors produced by approximating the original 200-point FIR filters for 6-pole-6-zeros IIR filters using the Kalman method. These are a price to be paid on the accuracy of the simulation for bringing the computational cost down to a much lower level, which supposedly have benefits on reducing real-time and parallelization constraints. Second, there are the errors introduced by using interpolated versions of the IIR filters. These are not to be compared with the original FIR filters (otherwise we would count the first kind of error twice), but with the IIR filters obtained by the Kalman method. In terms of the experiment of section 4.1, we consider the IIR filter corresponding to directions A, B, C and D to have been obtained with the Kalman method, and compare the IIR at D with the interpolated version of IIR at D obtained by using the VBAP gains on the IIRs of A, B and C. Figure 6 shows graphs of measured and interpolated HRTFs at selected directions of the database.

The method of spectral interpolation proposed here might be enhanced in several and important ways, for instance, by using HRTF databases more complete than CIPIC, and also by refining the pole-zero model using higher-order filters, thus allowing the simplified IIR model to capture more complex phenomena such as shoulder or pinnae reflections.

5. Conclusions and Further Work

In this article we presented two novel approaches for interpolating HRTF filters in the simulation of moving sound sources. The triangular interpolation is a time-domain linear interpolation technique akin to the bilinear interpolation, and the spectral interpolation presents a whole new way of dealing with moving sound sources.

With respect to allowing smooth transitions between filters for continuously moving sound sources, both methods may be made to update the interpolated filters on a sample-by-sample basis, which guarantees a smoothly varying output signal without audible discontinuities. Other strategies, such as lowpass-filtering the motion signal, may be used to ensure smooth transitions even when using a slower control rate.

One of the (already expected) conclusions of the experiment with the triangular interpolation is that the density of the HRTF database has a profound impact on every interpolation scheme. As the number of available measured HRTFs increases, the smaller the distance of a virtual sound source to an original HRTF will be, and the smaller the corresponding interpolation error.

We presented an attempt at objectively evaluating the quality of both methods, by recreating interpolated filters on top of existing ones and measuring the relative errors. The preservation of the overall attenuation patterns for all frequencies endorses the interpretation that frequency-dependent directional information is also preserved. The relative errors of interpolated filters were also measured, but these values were not well-behaved and did not help in providing objective conclusions about the quality of the interpolation procedure. This method of objectively comparing original and interpolated HRTFs proved to have a limited power in assessing the usefulness of these methods in the context of practical spatialization, where the illusion of direction is actually more important than a rigorous physical evidence of the relation between the direction and the corresponding filter.

Subjective tests are usually difficult to apply and to analyse for a number of reasons, such as our perceptual equipment lack of accuracy in differentiating nearby directions, and the fact that we are highly suggestible to visual imagery and other hints when evaluating direction of incoming sounds. Nonetheless, subjective tests are strongly needed in order to provide the clues to the relation between nearby directions and nearby HRIR waveforms or nearby HRTF frequency responses, and also to reconfirm the psychoacoustical effectiveness of these methods in simulating moving sound sources.

References

Algazi, R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). The CIPIC HRTF database. In *WASSAP '01*. 2001 IEEE ASSP Workshop on Application of Signal Processing to Audio and Acoustics.

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization*. MIT press.

Freeland, F. P., Biscainho, L. W. P., and Diniz, P. S. R. (2002). Efficient HRTF interpolation in 3D moving sound. In *WASSAP '01*. AES 22nd International Conference: Virtual, Synthetic, and Entertainment Audio.

Hofman, P., Van Riswick, J., and Van Opstal, A. (1998). Relearning sound localization with new ears. *Nat Neurosci*, 1(5):417–421.

Kalman, R. E. (1958). Design of a self-optimizing control system. In *Transactions of ASME*, volume 80, pages 468–478.

Kulkarni, A. and Colburn, H. S. (2004). Infinite-impulse-response models of the head-related transfer function. *The Journal of the Acoustical Society of America*, 115(4):1714–1728.

Moore, F. (1990). *Elements of computer music*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Pulkki, V. (2001). *Spatial sound generation and perception by amplitude panning techniques*. PhD thesis, Helsinki University of Technology. Espoo, Finland.

Roads, C. (1996). *The computer music tutorial*. MIT press.

Savioja, L., Huopaniemi, J., Lokki, T., and Väänänen, R. (1999). Creating interactive virtual acoustic environments. *J. Audio Eng. Soc.*, 47(9):675–705.

Stockhausen, K. (1961). *Two lectures*. T. Presser.

Xenakis, I. (1992). *Formalized music*. Pendragon Press Stuyvesant, NY.