HRTFEARLY & HRTFREVERB: FLEXIBLE BINAURAL REVERBERATION PROCESSING

Brian Carty       Victor Lazzarini

Sound and Digital Music Technology Group, National University of Ireland, Maynooth, Co. Kildare, Ireland

ABSTRACT

A binaural reverberation processor is presented, based on location accurate processing of early reflections and a Feedback Delay Network (FDN) approach to the later diffuse field. Recently developed Head Related Transfer Function (HRTF) dynamic processing algorithms, which have been shown to perform favorably when compared to typically employed methods, are used to allow dynamic direct sources and early reflections. A flexible binaural FDN, which considers interaural coherence, provides an efficient and robust later reverberation model. The overall system is designed to work parametrically, and requires no measured room impulses. This paper introduces the area and gives implementation details of each section of the overall reverberation algorithm from the point of view of recently developed Csound opcodes.

1. INTRODUCTION

In most natural listening circumstances, sound arrives at a listener not only directly from the source, but also after reflecting off obstacles/boundaries (reverberation). Several approaches to reverberation processing have been presented in the literature, varying greatly in computational cost and accuracy (for example, computationally efficient recursive filters [12] to more accurate yet significantly more costly digital waveguide meshes [10]). The impulse response of a room can be split into early reflections and a later, more diffuse reverberant tail. Several artificial reverberation models are based on this decomposition [2, 8, 11], which is also employed here.

The binaural nature of the human hearing system provides our main sound localization cues. Two signals, with possible timing and intensity differences, arrive at the ears and are processed by the brain. HRTFs describe how a sound is altered as it travels from a source at a particular location to the eardrum. HRTF pairs thus inherently include the above interaural differences, as well as other localization cues. The HRTF should thus be considered when endeavoring to recreate artificial reverberation binaurally.

2. HRTF PROCESSING: DYNAMIC TRAJECTORIES

HRTFs can be used to artificially spatialize any source sound to a desired location for headphone listening using convolution. HRTFs are typically measured at discrete points around a listener (the dataset used here is the MIT HRTF dataset [5]). Therefore, to move a source sound from point to point, interpolation is required. Interpolation in the frequency domain is desirable, but poses the difficulty of phase interpolation (phase is a periodic quantity). Decomposing the HRTF into a minimum phase plus linear delay system offers a potential solution to this problem. However, this transformation also potentially introduces inaccuracies [3] at certain locations and involves complex data processing. Novel approaches to the challenge of HRTF interpolation were recently suggested by the authors, which aim to minimize any complex data transformation, preparation or compression, thus allowing empirical data to be used more directly [3, 4].

Briefly, the two approaches suggested both involve interpolating HRTF magnitudes directly and novel approaches to phase interpolation. The first truncates phase, using brief, user definable cross fades to allow for phase changes. The second augments a spherical head model for phase with low frequency scaling, allowing for accurate low frequency interaural phase difference: a psychoacoustically motivated approach. Significantly, the algorithms perform very well in subjective and objective tests. When tested subjectively, the phase truncation and augmented spherical head model both provide a more convincing and artifact free dynamic source trajectory than a minimum phase model (although all perform well) [4]. An anchor condition, with HRTF switching as opposed to smooth interpolation performs less successfully. These results are illustrated in Figure 1. Phase truncation is used in this reverberation application as it is both efficient and performs excellently for dynamic source trajectories.

HRTF processing can offer very convincing results, but has limitations, which should be considered. HRTFs are individual specific, due to the individual nature of outer ear physiology. The binaural nature of HRTFs implies optimal...
reproduction on headphones. Also, artificial spatialisation can be difficult in a vision centric perceptual system.

Figure 1. Preference test results for HRTF dynamic source algorithms.

3. EARLY REFLECTIONS

The image model [1] is a geometric model which can be used for early reflection processing. It uses virtual sources in virtual rooms adjacent to the actual room to be modeled. The geometrical paradigm is illustrated in Figure 2. Each virtual source can be spatialized, filtered, delayed and attenuated accordingly, aiding with the perception of the listening environment.

Figure 2. The image model for early reflections (2 dimensional, 2nd order). S represents the actual source, L the listener and V virtual sources.

The implementation of the early reflections processor as the opcode \texttt{hrtfearly} for the computer music language Csound uses phase truncation HRTF processing to spatialize and move the direct sound source and early reflections in accordance with the image model. HRTF processing can be costly, particularly when compared to the more efficient later diffuse field algorithm. Therefore, the user can choose the order of the early reflections. Order 0 processes just the direct source, order 1 the first reflections, etc. It may be desirable to simplify early reflection spatial accuracy if higher order/more efficient processing is required [8]. Three dimensional processing is also optional, whereby reflections from the floor and the ceiling are considered. In a natural environment, a user may move his/her head, thus reorienting sound sources. This is also possible in the presented opcode. Source and listener location are also dynamic parameters.

The room parameters define the nature of the reverberant sound. A value for the low and high frequency absorption coefficients of each surface of the (presumed rectangular) room are used to calculate the cutoff frequency of a simple low pass filter which models the surface’s response. A three band equalizer is also offered for each surface to allow for implementation of multiband reflective surfaces. Source and reflection location are dealt with using HRTFs. Distance processing is implemented using an interpolated delay line (as well as attenuation), which also provides any appropriate Doppler Effect. As well as the processed input, the opcode outputs the rooms mean free path and low and high frequency reverberation times, based on the Norris-Eyring reverberation equation. These outputs can then be used in the diffuse field binaural reverberation opcode.

4. DIFFUSE FIELD

After the period of discrete early reflections in the evolution of the reverberation, sound begins to arrive from all around a listener, in a diffuse manner. Therefore, spatially accurate single reflections are no longer required. Without the necessity of considering each reflection individually, a much more efficient late reverberation can be implemented. The opcode \texttt{hrtfreverb} offers efficient binaural reverberation processing. It can be used with \texttt{hrtfearly} to provide spatially accurate source location as well as reverberation, or independently as an efficient, more general binaural reverberator.

A reverberant tail can be captured/modeled and convolved with the input sound to impose the reverberant characteristics of the space onto the input. However, convolution, although optimizable, can introduce processing delays and be computationally expensive for long impulses. A feedback system is perhaps a more appropriate solution. FDNs offer a subtle yet effective approach. Jot’s comprehensive treatment of the topic covers the scenario of modeling a measured impulse response [6] and using parametric inputs [7]. Jot also discusses binaural output of the results.

Recently, the binaural element of Jot’s measured impulse model was improved by considering interaural coherence [9]. The current model furthers this work by considering the parametric scenario, as well as independent early reflection processing.
Briefly, the Jot FDN works using a number of mutually prime delay lines, with frequency dependent feedback, in accordance with the desired frequency dependent reverberation time. The feedback loop includes a matrix which increases the density of the diffuse tail. In \textit{hrtfreverb}, the frequency dependent reverberation time can be chosen by the user, or values calculated by an instance of the \textit{hrtfearly} opcode, based on the inputted room geometry and surface characteristics, can be used.

The suggested reverberation frequency density of 0.15 modes per Hz [12] is achieved using a sufficiently long overall delay (the sum of the delay lines used). The mean free path of the environment provides a suitable average delay line length (it is also important that the shortest delay line is shorter than the overall diffuse onset delay for real time processing). A flexible number of delay lines is thus required for various reverberation times. 6, 12 or 24 are used in \textit{hrtfreverb}, based on the input reverberation properties. Frequency dependent reverberation times are achieved using simple first order Infinite Impulse Response (IIR) filters [7]. These filters (\(f(z)\) in Figure 3, which represents the overall process) also reduce the spectral energy for low reverberation times, which is compensated for using a tone correction filter (\(t(z)\)). Two outputs of the FDN are taken, which should be uncorrelated (vectors \(b\) and \(c\) ensure this). Coherence of the HRTF dataset is calculated using equation 1 [9]. The coherence matching filters \(u(\omega)\) and \(v(\omega)\), which provide accurate interaural coherence, are defined by equations 2 and 3 [9].

\[
\phi(\omega) = \frac{\sum_{i=1}^{N} L_i(\omega) R_i^*(\omega)}{\sqrt{\sum_{i=1}^{N} |L_i(\omega)|^2 \sum_{i=1}^{N} |R_i(\omega)|^2}} \tag{1}
\]

where \(L_i(\omega)\) and \(R_i(\omega)\) represent the \(i^{th}\) left and right HRTF in a dataset of \(N\).

\[
u(\omega) = \frac{1 + \phi(\omega)}{2} \tag{2}
\]

\[
v(\omega) = \sqrt{\frac{1 - \phi(\omega)}{2}} \tag{3}
\]

Left and right HRTF dataset average power filters (\(l(z)\) and \(r(z)\)) are then used to make the output binaural. The opcode outputs the appropriate delay time (according to the mean free path, processing order and inherent delay) for the late reverberation, as well as the processed audio. This delay, followed by a scaling factor, completes the process, which is illustrated in Figure 3.

5. IMPLEMENTATION

The opcodes are designed to balance efficiency, accuracy and usability. They must also be robust to a large number of scenarios. As discussed, \textit{hrtfreverb} is designed to be used with \textit{hrtfearly}, or equally, as an independent, efficient binaural reverberator. If a number of sources exist in the same environmental infrastructure, they can be summed and processed with the same instance of \textit{hrtfreverb}. Individual trajectories/locations can be dealt with using multiple instances of \textit{hrtfearly}. For ease of use, a number of default rooms are available, with standard surfaces. Processing at various sampling rates is also offered. Csoud provides suitable opcodes to implement the appropriate \textit{hrtfreverb} delay (calculated in a straightforward manner, appropriate to the parametric nature of the processing), as well as other opcodes which add a further dimension of user control (e.g. further low pass filtering).

As mentioned above, HRTF processing can be computationally costly, particularly when processing
multiple sources with multiple reflections each. Therefore, several code optimizations have been implemented, for example, optimal real FFT processing is used, and interpolation is only performed when relative source location changes. Real time performance in most typical scenarios is therefore achievable.

Overall, it is hoped that the opcodes offer an intuitive, flexible approach to binaural reverberation. The desirable balance between ease of use (default values, the standalone nature of the intuitive hrtfreverb opcode, etc.) and advanced processing (expert optional parameters, giving a fine degree of control) allow for both immediate and detailed use.

6. APPLICATIONS

The most obvious application of the HRTF reverberation opcodes is the binaural spatialisation of audio with accurate localization and environmental processing. More specifically, they allow for binaural multi-channel audition (the opcodes were developed with this application in mind). As any source can be spatialized to any location in any desired room, it is possible to place virtual loudspeakers in virtual listening rooms. The phase truncation interpolation algorithm allows a listener to move around this virtual environment without jeopardizing audio quality. Therefore, multi-channel algorithms can be auditioned on headphones. A composer/sound designer can thus work with a multi-channel setup using only headphones.

Using the infrastructure available in Csound, the application under development will allow a user to setup a room with a desired number of virtual loudspeakers. Each of these virtual loudspeakers is then fed with an appropriate audio stream (for example, the outputs of an Ambisonic/VBAP/Wave Field Synthesis mix). The overall output can then be auditioned. The user can also move through the listening space, for example to investigate any sweet spot issues. This approach is meant as an audition tool, as opposed to offering a general, optimized binaural solution.

Compositionally, the reverberation tools can also be useful. Indeed, their development was part motivated by the need for an accurate reverberation when composing using HRTFs. Interestingly, the parametric approach to the algorithms involved allows non natural scenarios, within the boundaries of stability, which may provide appealing compositional material. For example, massive rooms with very reflective walls, non natural levels of late reverberation and distant sources.

7. CONCLUSIONS

New tools for binaural reverberation are presented. The algorithms developed constitute a consolidation of classic and more recent approaches to reverberation. Several updates are suggested to allow for a flexible implementation, including recently developed HRTF interpolation algorithms and more accurate early reflections. The resulting algorithms are presented as efficient Csound opcodes, allowing both immediate application and a fine degree of control.

8. ACKNOWLEDGEMENTS

This Research is supported by the Irish Research Council for Science, Engineering and Technology: funded by the National Development Plan and NUI Maynooth.

9. REFERENCES