Audio Engineering Society
Convention
Paper
Presented at the 110th Convention
2001 May 12–15 Amsterdam, The Netherlands

A Framework for Evaluating Virtual Acoustic Environments

Tapio Lokki[1], Jarmo Hiipakka[2], and Lauri Savioja[1]

[1]Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory
P.O.Box 5400, FIN-02015 HUT, FINLAND
[2]Nokia Research Center, Speech and Audio Systems Laboratory
P.O.Box 407, FIN-00045 NOKIA GROUP, FINLAND

ABSTRACT

A new evaluation framework for virtual acoustic environments (VAE) is introduced. The framework is based on the comparison of real-head recordings with physics-based room acoustic modeling and auralization. The real-head recording procedure and VAE creation method are discussed and new signal processing structures for auralization are introduced. As a case study, recordings were made in a classroom which was also modeled and auralized.

0 INTRODUCTION

In this paper we report a new framework for evaluating the quality of virtual acoustic environments (VAE). The framework is suitable for evaluating both static and dynamic VAEs. In a static environment the positions of the sound sources and the listener, as well as the room geometry are fixed, whereas in a dynamic VAE any or all of these may change.

Room acoustic modeling and auralization can be divided into two different approaches, namely perceptual and physics-based modeling. In perceptual modeling the aim is to find a set of perceptually relevant parameters by which room acoustics can be rendered. For musical and multimedia purposes these parameters include source presence, envelopment, room presence, late reverberance, etc. as described by Jot [1]. The perceptual approach has also been used by, e.g., Pellegrini [2, 3] whose goal has been plausible simulation of a listening room. By contrast, in physics-based modeling and auralization the behavior of sound waves in a modeled space is simulated. Room acoustic modeling is often done using the ray-tracing method [4, 5]. However, for auralization purposes the image-source method [6, 7] is more directly suitable, because it gives the direct sound and early reflections in an efficient manner. With the image-source method late reverberation is usually implemented separately, e.g., by an efficient recursive algorithms.

The basic principles and ideas of physics-based rendering were given already at 1983 by Moore [8]. However, one of the first complete sound

rendering systems that creates natural sounding rendering, was implemented in Helsinki University of Technology (HUT) [9] and it is called Digital Interactive Virtual Acoustics (DIVA). The sound rendering part of the DIVA system has been reported in more detail by Savioja *et al.* [10] and in this paper we present the recent improvements to the system.

The motivation for the evaluation framework presented in this paper came from the need to estimate the quality of DIVA auralization system. Nevertheless, the framework is modular and can be applied in evaluation of any sound rendering system. The modularity allows changing of any component applied in creation of both recorded and auralized soundtrack.

In VAE creation our ultimate goal has been to develop a sound rendering application that can be used in acoustic design. In its ideal form an authentic reproduction of a real environment would be indistinguishable from the real environment without any exception [11]. Of course, this is not possible because of simplifications that have to be made in room acoustic modeling. We try to do plausible, perceptually authentic, auralization that can be used as a reliable tool in room acoustics design. We know very well that the rendering with same level of naturalness can be implemented with more efficient algorithms with perceptual modeling. However, our starting point has been to take the room geometry and all the physics-based data that we can get and with this information to realize natural sounding auralization. To achieve this ambitious goal, we utilize the DIVA auralization system which is a physics-based room acoustic modeling and auralization system that does rendering in time domain. The applied auralization method enables dynamic rendering with normal PC (without any additional DSP-cards), running Linux operating system.

This paper is organized as follows. First, the proposed evaluation framework is introduced. We discuss general aspects of evaluation, room acoustic modeling, auralization, and sound rendering. Then a more detailed description of DIVA auralization is given with a case study in which the proposed framework is applied. Only the recent improvements of DIVA auralization [10] are presented. Also, the description of recording equipment as well as parameters used in sound rendering are presented and possible error sources are discussed. Finally, conclusions are drawn.

1 EVALUATION FRAMEWORK FOR VIRTUAL ACOUSTIC ENVIRONMENTS

In this framework the evaluation of quality of auralization is done by comparing real-head recordings and auralized room acoustic simulation (see Fig. 1). The real-head recordings are used as reference signals. The reference soundtracks are prepared by playing anechoic sound samples in the studied room and recording the sound using the real-head recording technique [12, 13] (see, e.g., [14] for more about the fundamentals of binaural recording techniques). The soundtracks to be evaluated are created by auralizing modeling results as presented on the right column of Fig. 1.

The novelty of this framework is that it is also suitable to dynamic sound rendering which can not be realized with traditional convolution of binaural impulse responses and anechoic stimulus [15]. Actually, traditional convolution based sound rendering has been evaluated respectively by Pompetzki [16, 17]. In his work, he has studied the influence of model complexity and number of image sources required for reasonably authentic perception. He used static rendering with an omnidirectional source.

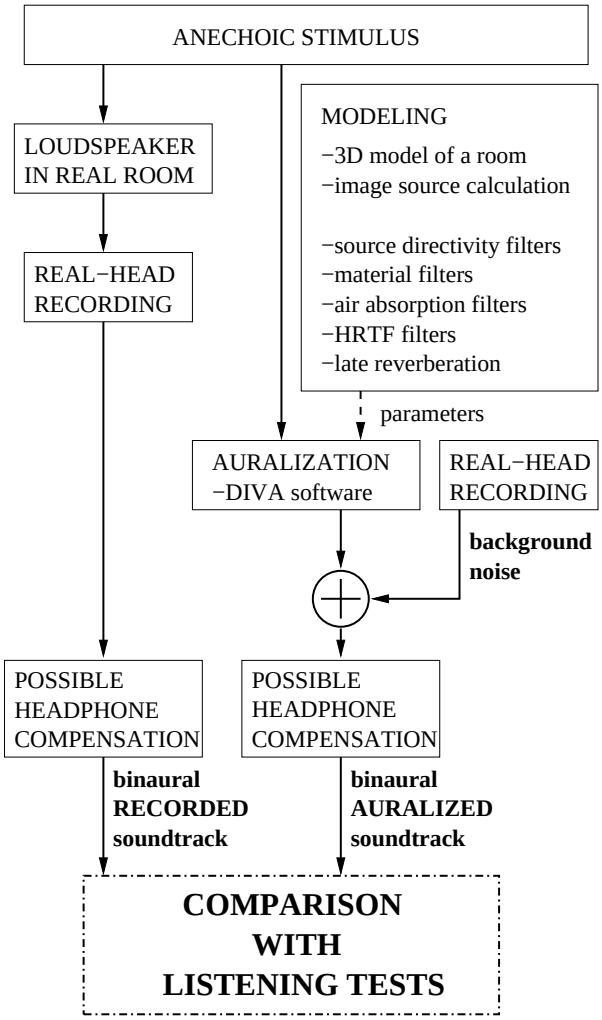

Fig. 1: The framework for evaluating virtual acoustic environments. On the left column the processing blocks for real-head recordings are presented and on the right column the auralization parts are illustrated.

1.1 Considerations About the Evaluation Process

All possible error sources that do not depend on room acoustic modeling and auralization should be minimized. This means, e.g., that the same stimulus has to be used in each case. In addition, inaccuracy in sound source and listener modeling should be kept as minimal as possible by using a sound source whose radiation characteristics are known and by recording with a real head with known HRTF characteristics. For example, we are using a high quality loudspeaker whose radiation properties were measured in an anechoic room.

Real-head recordings without any stimulus should also be made to capture the background noise in a studied room. The recorded background noise can then be added to auralized soundtracks so that the background noise is equal in comparison. In subjective comparison binaural soundtracks should be reproduced with headphones so that the acoustics of the listening room do not affect to the results. Alternatively, loudspeak-

ers can be utilized if good listening conditions and a proper way to reproduce binaural sountracks are available.

One problem with headphone reproduction is that headphones seldom have very flat frequency responses and the magnitude (as well as phase) errors should be compensated. A straightforward method to carry out compensation is to measure headphone transfer functions when the microphones are in the entrances of the ear canals and deconvolve these responses in reproduction. This measurement is often carried out at the same time when measuring HRTF functions and headphone compensation can be embedded in the designed HRTF filters. However, in normal case the real-head recordings and HRTF measurements are not made at the same day and it can not be guaranteed that microphone positions in ears are identicals. If embedded compensation is applied to HRTFs this might cause some unwanted differences, especially at high frequencies ($>$ 6 kHz, when the wavelength of sound is the same or smaller than the dimensions of pinnae). As a matter of fact, if utilized headphones are very high quality, no headphone compensation is needed, because the evaluation is done by comparing soundtracks. Anyway, the compensation for the frequency response of the headphones can be done if needed.

1.2 Room Acoustic Modeling and Auralization

In room acoustic simulation our goal is to create a totally artificial, but still plausible, virtual auditory environment. In other words, no measured room impulse responses are used in sound rendering. This means that the sound source characteristics, sound propagation in a room as well as the listener have to be modeled.

In DIVA auralization the modeling of room acoustics are divided into three parts: the modeling of direct sound, early reflections, and late reverberation. The direct sound and early reflections are modeled with the image-source method and late reverberation with an efficient recursive algorithm. With the image-source method the following parameters for each reflection, at each time instant, are calculated:

- orientation (azimuth and elevation angles) of sound source
- distance from listener
- material filter parameters
- azimuth and elevation angle with respect to the listener

These parameters are used in the auralization process that is implemented as the signal processing structure presented in Fig. 2. The signal processing blocks contain the following filters:

- $S_d(z)$ is a diffuse field filter of a sound source.
- $F_d(z)$ is a diffuse field filter of HRTFs.
- $T_{0\ldots N}(z)$ contain the sound source directivity filter, distance dependent gain, air absorption filter and material filter (not for direct sound).
- $F_{0\ldots N}(z)$ contain directional filtering realized with separated ITD and minimum-phase filters for HRTFs.
- R is a late reverberation unit.

Each of these blocks is discussed in mode detail in the case study section.

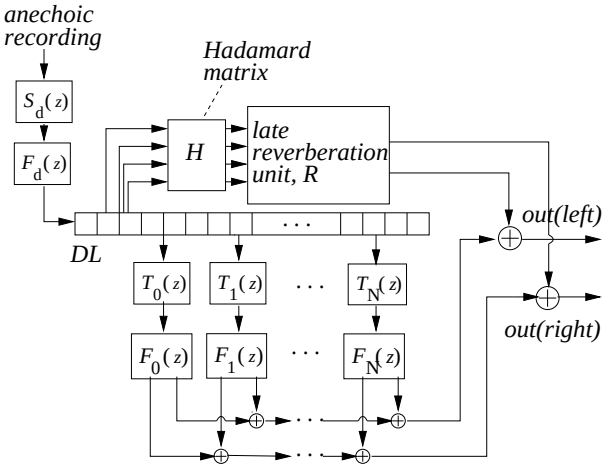

Fig. 2: The DIVA auralization signal processing structure. From the long delay line DL the sound signal is picked to filter blocks $T_{0\ldots N}(z)$ according to the distance of the image source from the listener.

1.3 Sound Rendering

Takala and Hahn [18] have defined the term *sound rendering* as a process of forming a composite soundtrack from component sound objects. Here, the sound objects are such as the direct sound, early reflections and late reverberation. Thus, sound rendering in our case means processing of anechoic recordings (or synthetic sound) so that spatial characteristics are included in the resulting soundtrack.

Static sound rendering, where sound source and the listening point are not moving, is usually done by convolving an anechoic stimulus with binaural room impulse responses [15]. The convolution process can be realized in time or frequency domain and is more or less a trivial operation.

By contrast, dynamic sound rendering, in which anything can move, is not trivial, because binaural room impulse responses can not be used. In fact, if the environment is dynamic, the impulse response is not anymore defined. However, the actual rendering process can be done in several parts. The first part consists of direct sound and early reflections, both of which are time- and place-variant. The latter part of rendering represents the diffuse reverberant field, which can be treated as a time-invariant filter. In practice this means that the direct sound and early reflections are rendered according to parameters obtained by room acoustic modeling, namely image source modeling. These auralization parameters (presented in the previous section) define the coefficients for filters $T_{0\ldots N}(z)$ and $F_{0\ldots N}(z)$ (see Fig. 2) as well as sound pick-up points from the delay line DL.

The signal processing parameters in DIVA system is divided into rendering and auralization parameters. The auralization parameters, presented in Section 1.2, are obtained from image source calculation process and need not to be updated for every sample. However, the rendering parameters have to be defined sample by sample basis, and they are created interpolating auralization parameter. The reasonable update rate of auralization parameters depends on the available computational power and speed of movement. If very fast movements are required the update rate has to be higher than with slow movements. Otherwise some audible artifacts occur. These artifacts can be clicks, for example, coming from too big changes in filter coefficients or inaccuracy in localization as reported, e.g., by Wenzel [19].

2 A CASE STUDY OF A LECTURE ROOM

As a case study the proposed evaluation framework was applied to a lecture room (dimensions 12 m x 7.3 m x 3 m) that has quite a simple geometry, as presented in Fig. 3. In this section, we will describe how real-head recordings and sound rendering were made. We also present in more detail the DIVA auralization system and details of all applied filters.

For this case study we recorded and simulated six cases with different listener position and orientation characteristics. Two of these were static cases (listening points s1 and s2), and four were dynamic cases, namely two turnings and two walk-paths. In turning cases listening points s1 and s2 were used with the head movements t1 and t2 illustrated in the corner of Fig. 3. The dynamic walk-paths w1 and w2 were both made so that the listener's head was pointing to the forward direction when walking.

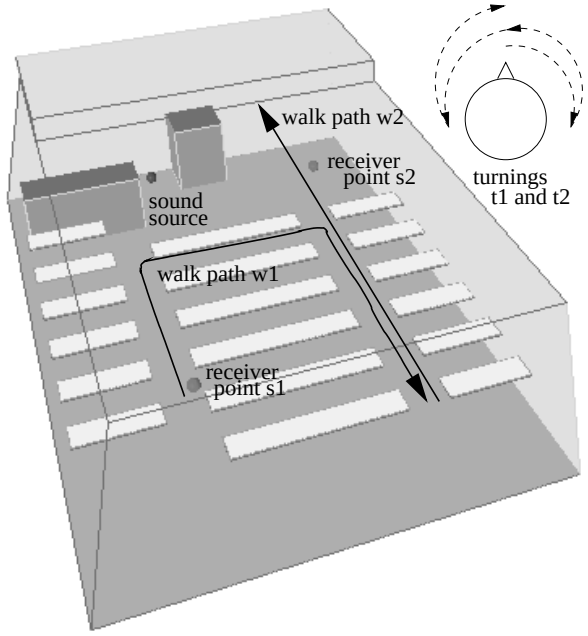

Fig. 3: The 3D model of the studied lecture room. In addition, different rendering cases, two static ones, two turnings, and two walk-paths, are depicted. Head turnings t1 and t2 are applied in the receiver points s1 and s2.

2.1 Real-Head Recordings

The reference soundtracks were prepared by playing anechoic sound samples in the studied room and recording the sound using the real-head recording technique. The sound samples were music, such as singing, guitar, and clarinet. In addition, we used snare drum hits which are wideband transient-like signals. The anechoic stimuli were played with a CD player (Sony Walkman) and reproduced with a small active loudspeaker (Genelec 1029A). Small electret microphones (Sennheiser KE 4-211-2), used with a custom made preamplifier, were placed at the entrances of the ear canals and connected to a DAT recorder. The contents of the DAT tape were then transmitted to the computer, edited and equalized for headphone listening (see leftmost column in Fig. 2). All recordings as well as auralizations were done at 48 kHz sampling frequency.

2.2 Auralization Parameters and Filters

A polygonal 3-D model of the studied classroom is used to calculate auralization parameters for the direct sound and early reflections. With these parameters the direct sound and early reflections are auralized taking into account the directivity of the sound source, air and material absorption, and distance delay and attenuation as well as directional filtering with HRTFs. The applied HRTFs are measured from the same head with which the real-head recordings were done. To obtain a complete simulated soundtrack also late reverberation is added to the auralized early part of the room response. The late reverberation algorithm that creates diffuse late reverberation is adjusted according to an objective analysis of the room impulse response, e.g., reverberation time at different frequency bands.

Filter design has been realized with Matlab [20] software by applying different design methods. All the design problems have been filter fitting problems, where the filter has been matched to a measured magnitude response (applied with sound source directivities and HRTFs) or to some given discrete magnitude values at certain frequencies (applied with air and material absorption).

Before filter design we preprocessed the magnitude responses so that the filter fitting has minimal errors by auditory perception point of view. Applied methods were frequency dependent smoothing and weighting according to ERB scale [21]. The most straightforward way to accomplish smoothing is to use moving averaging of variable window size, in this case window size was an ERB band. In filter design we used algorithms which allow frequency dependent weighting, which has been found to be very useful. In addition, we have used an equivalent method where a given magnitude response is resampled according to ERB resolution and then filter fitting is realized. The applied preprocessing methods yield filters that respect the nonlinear frequency resolution of the human hearing.

Sound source directivity is quite a difficult phenomenon to model. A good overview of the directivity of sound sources has been written by Giron [22]. As mentioned earlier, we used a small Genelec loudspeaker as the sound source in our recording and modeling. For modeling purposes we measured loudspeaker impulse responses in anechoic room from 24 different azimuth angles (every $15°$) and 4 elevation angles ($0°$, $30°$, $60°$, and $90°$), in total 96 responses. From these measurements a filter grid of every $5°$ in azimuth and elevation was designed. Finally, the total amount of designed filters was 1368 and they were applied symmetrically for negative elevations, although the loudspeaker radiation characteristics are not strictly symmetrical in upper and lower hemisphere.

The sound source directivity filtering has been realized according to ideas presented previously [23, 24, 25]. The filtering is distributed to two filters (for each sound source and reflection), namely diffuse field filter ($S_d(z)$ in Fig. 2) and directivity filter ($D_{0\ldots N}(z)$ in Fig. 4). The measured magnitude response and fitted IIR filter (order 30) of the diffuse field filter is depicted in Fig. 5. The target diffuse field power spectrum has been derived from all measured responses by power-averaging

$$S_d(\omega,\theta,\phi) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|H_i(\omega,\theta,\phi)|^2\cos\phi} \quad (1)$$

where θ is the azimuth angle, ϕ is the elevation angle, and ω is the angular frequency.

Filtering the input signal with the diffuse field filter of the source ($S_d(z)$ in Fig. 2) has many advantages. First of all, it makes the design of primary directivity filter ($D_{0\ldots N}(z)$ in Fig. 4) much easier, by

flattening the response at low and high frequencies. This is also called diffuse field equalization and it allows the use of lower order filters for the primary directivity filter. In our case we used 7^{th} order IIR filters. The other advantage is that the signal fed to the late reverberation algorithm is filtered by the average directivity of the sound source. This way the late reverberation spectrum is modified with sound source directivity properties as explained in [24].

Air and material absorption filters were designed as presented by Huopaniemi *et al.* [26, 10]. Air absorption transfer functions were calculated based on the standardized equations [27] and first-oder IIR filters were fitted to the resulting magnitude responses.

The material absorption filters were fitted to the absorption coefficients data available at octave bands. When sound is reflected from two or more surfaces the absorption data can be cascaded so that only one filter is needed. The algorithm for realizing cascaded absorption coefficient data with a low-order IIR filter was as follows. First, all possible boundary absorption combinations were calculated and transformed into reflectance data. In the second phase, the resulting amplitudes were transformed into a minimum-phase complex frequency response. A frequency-domain weighted least-squares fitting algorithm was then applied to the complex reflectance data. As a result, a vector containing reflection filter coefficients for all surface combinations was stored for use in the system.

Binaural filtering was realized with a diffuse field filter ($F_d(z)$ in Fig. 2) and with diffuse field equalized HRTF filters. The diffuse field equalization has been discussed in more detail in [14]. This efficient implementation, presented as division into an angle-independent part (the diffuse field part) and a directional transfer function (DTF) by Kistler and Wightman [28], has many advantages. The angle-independent features can be modeled as a general filter for all directions, and as a consequence the DTFs can be designed using lower-order models than those used for full HRTFs. In our implementation the diffuse field filter is also used for spectral modifications of the late reverberation.

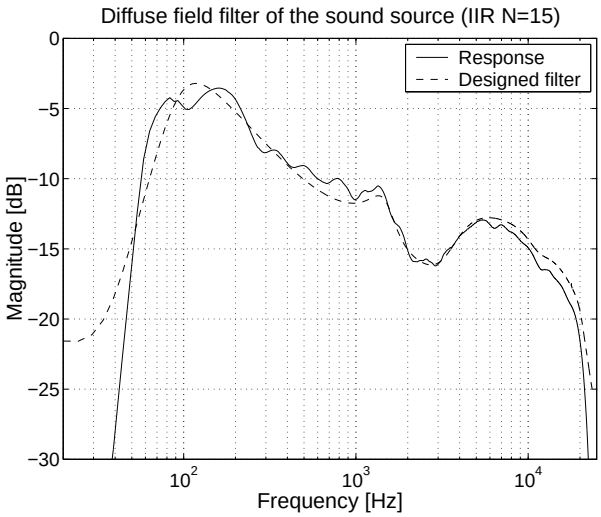

Fig. 5: Diffuse field filter of the sound source ($S_d(z)$ in Fig. 2).

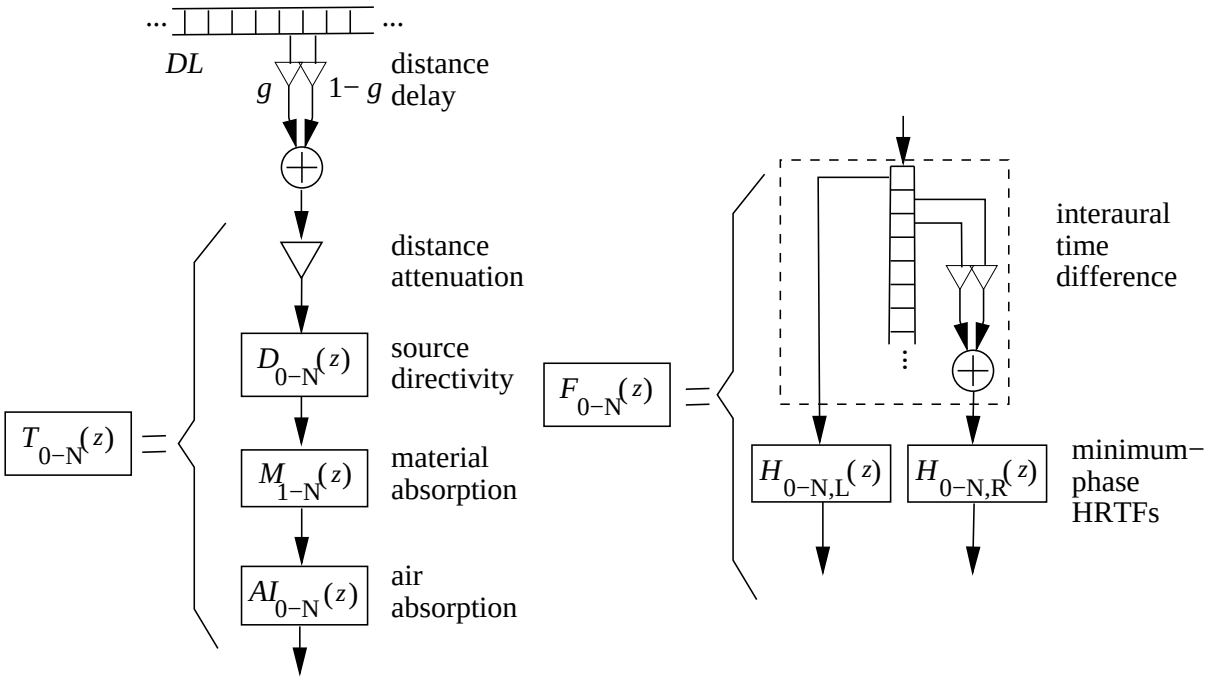

Fig. 4: Detailed filters in $T_{0\ldots N}(z)$ and $F_{0\ldots N}(z)$.

There exists several methods to define a diffuse field binaural filter response [29]. We tried two different methods and the magnitude responses as well as the response of the applied filter are depicted in Fig. 7. The first method is equivalent to diffuse field source directivity design and can be calculated with Eq. (1) from a set of free-field measured HRTFs. However, this method produced a magnitude response in which high frequencies are attenuated quite a lot. The other applied method, proposed by Larcher *et al.* [29], allows measuring the diffuse field filters in a normal room. In this method the real-head impulse responses and a monaural impulse response are measured in a normal room, at the same point. The diffuse field binaural response is estimated on the late part (after early reflections) of the measured impulse responses. We used the part between 50 and 75 ms. Then this omnidirectional impulse response part is deconvolved from binaural response parts and the resulting frequency responses are diffuse field binaural responses. Averaging from a few measurements and from both ears, we achieved a response (see Fig. 7) to which we fitted the applied diffuse field filter.

The filter design principles of the primary HRTF filters ($H_{0\ldots N}(z)$ in Fig. 4) has been reported in [10, 30]. As depicted in Fig. 4 the HRTF filter is divided into ITD part (implemented with a pure delay line) and a minimum-phase counterpart of HRTF. The ITD was calculated using a spherical head based ITD model (discussed in, e.g., [31]) with added elevation dependency ($\cos\phi$) [10].

$$ITD = \frac{a(\sin\theta + \theta)}{2c}\cos\phi \qquad (2)$$

where a is the radius of the head, θ is the azimuth angle, ϕ is the elevation angle, and c is the speed of sound. For minimum-phase HRTFs we used FIR filters of order 60.

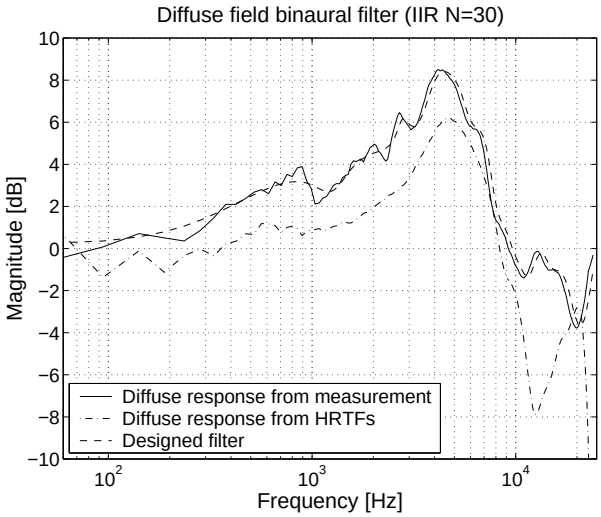

Fig. 7: Diffuse field binaural filter ($F_d(z)$ in Fig. 2).

The late reverberation in a room is often considered nearly diffuse and the corresponding impulse response exponentially decaying random noise [32]. Under these assumptions the late reverberation does not have to be modeled as individual reflections with certain directions. Therefore, to save computation in late reverberation modeling, recursive digital filter structures have been designed, whose responses model the characteristics of real room responses, such as the frequency dependent reverberation time.

The applied artificial reverberation is parameterized based on room acoustic attributes obtained with impulse response measurements in the studied room.

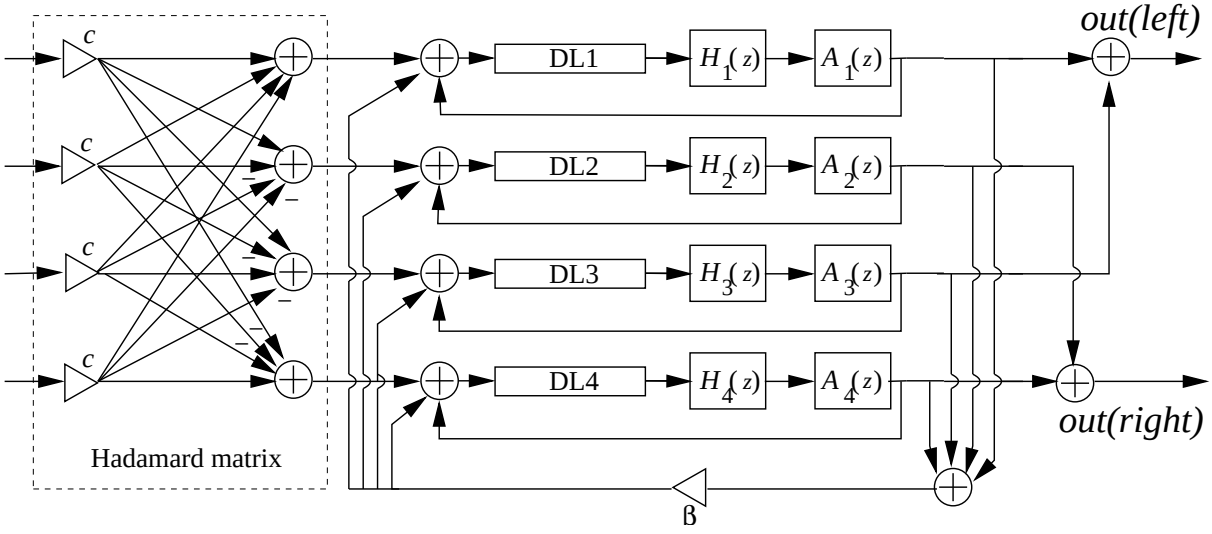

Fig. 6: Late reverberation algorithm. The filters $H_{1\ldots4}(z)$ are lowpass IIR filters of 1^{st} order and $A_{1\ldots4}(z)$ are comb-allpass filters.

The applied late reverberation algorithm [33] contains n parallel feedback loops, where a comb-allpass filter is in each loop (see Fig. 6). It is a simplification of a feedback delay network (FDN) structure [34, 35] and produces natural sounding late reverberation. The comb-allpass filters in the feedback loops, denoted by $A_{1...4}(z)$ in Fig. 6, are added to produce an increased reflection density. The filters $H_{1...4}(z)$ implement the frequency dependent reverberation time. Each of them contains a simple all-pole first order lowpass filter whose parameters are calculated automatically when user gives the required reverberation time at low and high frequencies. The lengths of the delay lines in the loops are chosen to be mutually incommensurate in samples to avoid reflections occurring at the same time, and strong coloration caused by coinciding modes in the frequency domain [32].

The inputs to the reverberator are picked directly from the main propagation delay line (DL in Fig. 4). The fixed pick-up points were arbitrarily chosen so that the beginning of the late reverberation slightly overlaps with last early reflections. To increase the reflection density the input signals are fed through a Hadamard matrix. The other advantage produced by the Hadamard matrix is that it makes reverberation outputs highly uncorrelated. This is important from the human perception point of view and helps in externalization in headphone reproduction. However, this uncorrelation is not frequency dependent as in real case when high correlation at low frequencies occurs. The low frequency correlation can be achieved with a correlation filter [24, 1], but we have not implement it.

2.3 Sound Rendering

The auralization process in DIVA system is divided into two processes, namely image-source method and sound rendering. According to the positions and orientations of the surfaces, the sound source(s), and the listener, image source calculation provides auralization parameters presented in Section 1.2. Both static and dynamic rendering have been implemented in the same manner. Naturally, with static rendering the interpolations, presented below, are not needed.

In **dynamic rendering**, in which parameters of the direct sound and early reflections are time variant, every single parameter has to be interpolated for every sample. The image source calculation uses with a chosen update frequency (usually from 20 to 100 Hz) when computing new auralization parameters. With these parameters the sound frame between previous and the new updates are processed. Some filters coefficients, such as material and air absorption and HRTF filters, can be updated without any problems when a new parameter set is received. However, the sound source directivity filters, being 7^{th} order IIR filters, can not be changed without interpolation. This is done by calculating two filters and cross-fading them.

In addition to filter coefficient changes the pick-up points from delay line (DL in Fig. 2) have to be changed when listener's distance from the sound source changes. This operation is a kind of resampling and has to be done using fractional delays [36] if a continuous output is desired. As depicted in Fig. 4, we used the simplest possible fractional delay, a first order FIR filter. The drawback of such a simple but efficient filter is shown in Fig. 8. The high frequencies are attenuated when the desired pick-up point is between two samples. However, this attenuation is not very severe (worst case, when exactly between two samples, 2 dB at 10 kHz and 5 dB at 15 kHz). The worst case occurs rarely because while the pick-up point is moving it seldom happens to be exactly in the middle of two samples. Actually, what happens to the direct sound and each reflection is that in movements the attenuation changes between null and the worst case (see Fig. 8). The updated delay values are always integer values so when the listener does not move no fractional delay is applied. The other case where fractional delays are used is in ITD delays. In this case the fractional delay is always applied to the contralateral ear, the filter of which already has strongly attenuated high frequencies.

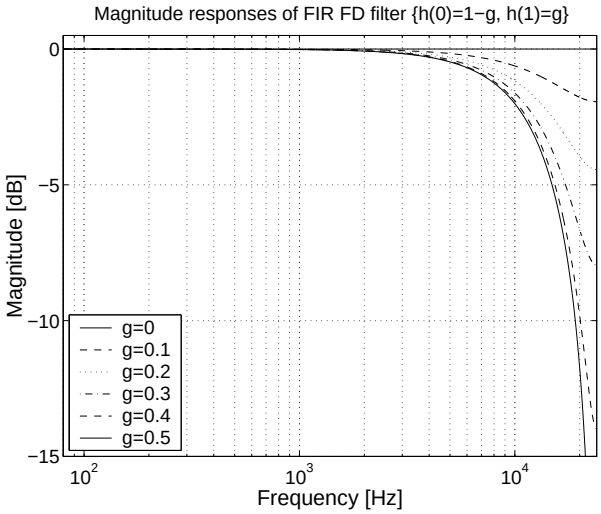

Fig. 8: Magnitude responses of FIR fractional delay filters of 1^{st} order with different gain values.

The advantage of this interpolation is that it gives a continuous output. If this small high frequency attenuation is disturbing for some applications, more sophisticated fractional delays can be applied. Another way, presented by Wenzel *et al.* [37], is to use upsampling before the fractional delay. The interpolation also gives a Doppler effect if the listener (or the sound source) is moving fast enough. Indeed, each reflection has its own interpolation rate depending on the change in distance. According to our knowledge this is also the case in real life, and in reverberant space the Doppler effect is not so easily perceived as in free field conditions.

The Doppler effect indeed affects to the **walk-path creation** needed in modeling. In this case study we had to create by hand the same walk-paths that were recorded. The applied method was to use keypoints and interpolate with spline functions between these keypoints. The keypoints have to be selected with great care so that the interpolation does not create fast accelerations or decelerations. If these phenomena occur they are heard as a slight flanger effect or beating.

Synchronizing recorded and modeled soundtracks is an issue which has to be handled with great care because in subjective evaluation the sound samples are listened with AB comparison test. No automatic synchronization method was applied, instead the soundtracks were synchronized manually. A good and practical way to do it is to listen, e.g., the left channel of both soundtracks with headphones so that the recorded one is played to the right ear and the auralized one to the left ear. This way the unsynchronization is easily detected and can be fixed.

In our simulations the **update rate** for the auralization parameters was 100 Hz. The update rate needed for good perceptual quality depends on the speed of the simulated movement, in our case mainly the angular velocity of the head in the turning points. The necessary update rate may also depend on the method and implementation selected for HRTF filter interpolation; in some cases unwanted disruptions may occur if the update rate is too small. From the perceptual point-of-view the selected update rate should give flawless results (see, e.g., [19]).

For our case study, the simulations to be compared with the recordings were prepared off-line. Thus, the latencies introduced by the auralization system were not a big concern for us. Additionally, because

the recordings were made without any positional tracking, it was impossible to make the real and simulated trajectories exactly the same. However, the simulation system is capable of real-time rendering with system latencies suitable for interactive use.

3 DISCUSSION

The simulation system contains dozens or even hundreds of digital filters and other processing components. For this evaluation, our goal was to create a virtual acoustic environment that would sound as natural as possible and could be directly compared with real recordings. Thus, we did not aim for ultimate efficiency at this point, but we will make simplifications to the simulation system later checking the system for perceptual differences after each modification. The compromises made in computational efficiency and optimal filter orders resulted in simulation times about twice the time duration of the sound samples on a normal powerful PC computer (a 500 MHz Intel Pentium III Processor).

The optimal solution for filter orders and the chosen filter design methods are hard to define. For example we know that HRTF filtering can be performed with less computational requirements by using, e.g., principal component basis functions [28, 38] or the interaural transfer function model [39]. However, our first goal is to make as perceptually authentic auralization as possible and the second goal is to optimize the computational requirements.

3.1 Preliminary Results

In this case study image sources were calculated up to the third order reflections. In the studied space (Fig. 3) this means 30-50 reflections that arrive to the listener in the time window of 50 ms after the direct sound. The late reverberation algorithm starts to create output "reflections" after 30 ms, thus it is a little bit overlapping with the last modeled early reflections. In Fig. 9 two impulse responses, measured and modeled, are depicted to clarify this overlapping. In addition, the responses show that our modeling is not yet perfect and the responses look different in the beginning. However, these errors in modeling are at low and high frequencies, but the first listening tests [40] showed that at mid frequencies our modeling is quite reliable. The static impulse responses can also be analyzed with objective methods such as traditional room acoustic attributes based on sound energy decay of impulse response on one-third octave or octave bands. This analysis is out of the scope of this article, as well as the applied novel auditorily motivated time-frequency analysis method [41]. However, objective analysis will be presented in near future [42].

Possible modeling errors can be divided into three groups as presented in Table 1. This coarse division is based on the results of informal listening tests with the DIVA auralization system. All "negligible" and "clearly audible" errors can be fixed by using longer and more accurate filters as well as more time in refining the models. However, "modeling errors" can not be overcome using the image-source method, but we do not know the relevance of these errors from perception point of view. For example, the diffraction needs a more elaborate modeling method [43].

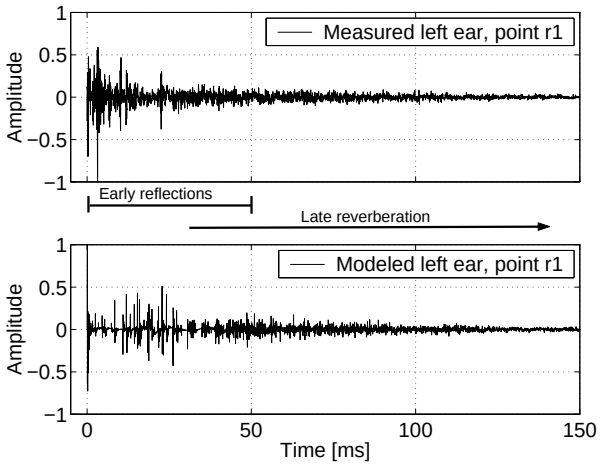

Fig. 9: Measured and modeled impulse response in listening point s1 (measured in the entrance of the left ear).

4 CONCLUSION

The evaluation framework for virtual acoustic environments has been introduced with a case study done in an ordinary lecture room. In our evaluation framework, real-head recordings (both static and dynamic) are used as reference sound signals. The methods and equipment for creating these reference soundtracks are reported, by pointing out the design and implementation issues of good quality real-head recordings. Also the post-processing, i.e., the headphone equalization, is discussed.

In the case study section the description of the recent improvements made to the DIVA auralization system were presented. The filter design methods were overviewed and auralization components were dug in more detail. Also the sound rendering issues, such as dynamic rendering, update rate and computational efficiency were discussed. Finally, the quality of the DIVA auralization system—to be evaluated with proposed framework—was discussed.

Negligible or barely audible	Clearly audible	Modeling errors
small errors in distance delay modeling	errors caused by interpolations	lack of diffraction model
small errors in air absorption modeling	errors in HRTF modeling	lack of diffuse reflection model
small errors in walk-path creation	errors in source directivity modeling	
small errors in material absorption modeling	errors in late reverberation modeling	
small errors in distance attenuation modeling	number of image sources used	

Table 1: The error sources between recorded and simulated soundtracks using the DIVA auralization system.

ACKNOWLEDGMENTS

This work has been partly financed by the Helsinki Graduate School in Computer Science. The authors also wish to thank Nokia Foundation, and Finnish Foundation of Technology Development (Tekniikan edistämissäätiö) for financial support.

REFERENCES

[1] J.-M. Jot. Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces. *Multimedia Systems, Special Issue on Audio and Multimedia*, 7(1):55–69, 1999.

[2] R.S. Pellegrini. Comparison of data- and model-based simulation algorithms for auditory virtual environments. In *the 106th Audio Engineering Society (AES) Convention*, Munich, Germany, May 8-11 1999. preprint no. 4953.

[3] R.S. Pellegrini. Perception-based room rendering for auditory scenes. In *the 109th Audio Engineering Society (AES) Convention*, Los Angeles, Sept. 22-25 2000. preprint no. 5229.

[4] A. Krokstad, S. Strom, and S. Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *J. Sound Vib.*, 8(1):118–125, 1968.

[5] A. Kulowski. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18(6):449–469, 1985.

[6] J. B. Allen and D. A. Berkley. Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.*, 65(4):943–950, 1979.

[7] J. Borish. Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Am.*, 75(6):1827–1836, 1984.

[8] F. R. Moore. A general model for spatial processing of sounds. *Computer Music J.*, 7(3):6–15, 1983 Fall.

[9] T. Takala, R. Hänninen, V. Välimäki, L. Savioja, J. Huopaniemi, T. Huotilainen, and M. Karjalainen. An integrated system for virtual audio reality. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4229.

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

[11] R.S. Pellegrini. Quality assessment of auditory virtual environments. In *Proceedings of Internoise 2000*, volume 6, pages 3477–3483, Nice, France, Aug. 27-30 2000.

[12] P. Maijala. Better binaural recordings using the real human head. In *Proc. Int. Congr. Noise Control Engineering (Inter-Noise 1997)*, volume 2, pages 1135–1138, Budapest, Hungary, Aug. 25-27 1997.

[13] H. Møller, C.B. Jensen, D. Hammershøi, and M.F. Sørensen. Using a typical human subject for binaural recording. In *the 100th Audio Engineering Society (AES) Convention*, Copenhagen, Denmark, May 11-14 1996. preprint no. 4157.

[14] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3-4):171–218, 1992.

[15] M. Kleiner, B.-I. Dalenbäck, and P. Svensson. Auralization – an overview. *J. Audio Eng. Soc.*, 41(11):861–875, Nov. 1993.

[16] W. Pompetzki. *Psychoakustische Verifikation von Computermodellen zur binauralen Raumsimulation.* PhD thesis, Ruhr-Universität Bochum, Verlag Shaker, Aachen, 1993.

[17] W. Pompetzki and J. Blauert. A study on the perceptual authenticity of binaural room simulation. In *Proc. of the Wallace C. Sabine Centennial Symposium*, pages 81–84, June 5-7 1994.

[18] T. Takala and J. Hahn. Sound rendering. *Computer Graphics*, SIGGRAPH'92(26):211–220, 1992.

[19] E. Wenzel. Analysis of the role of update rate and system latency in interactive virtual acoustic environments. In *the 103rd Audio Engineering Society (AES) Convention*, New York, Sept. 26-29 1997. preprint no. 4633.

[20] MATLAB. Signal processing toolbox, version 4.2, 1998. MathWorks Inc.

[21] B.C.J. Moore, R.W. Peters, and B.R. Glasberg. Auditory filter shapes at low center frequencies. *J. Acoust. Soc. Am.*, 88(1):132–140, July 1990.

[22] F. Giron. *Investigations about the Directivity of Sound Sources.* PhD thesis, Ruhr-Universität Bochum, Verlag Shaker, Aachen, 1996.

[23] M. Karjalainen, J. Huopaniemi, and V. Välimäki. Direction-dependent physical modeling of musical instruments. In *Proc. 15th Int. Congr. Acoust. (ICA'95)*, pages 451–454, Trondheim, Norway, June 1995.

[24] J.-M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. In *the 98th Audio Engineering Society (AES) Convention*, Paris, France, 1995. preprint no. 3980.

[25] J. Huopaniemi, K. Kettunen, and J. Rahkonen. Measurement and modeling techniques for directional sound radiation from the mouth. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'99)*, Mohonk Mountain House, New Paltz, New York, Oct. 1999.

[26] J. Huopaniemi, L. Savioja, and M. Karjalainen. Modeling of reflections and air absorption in acoustical spaces — a digital filter design approach. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'97)*, Mohonk, New Paltz, New York, Oct. 19-22 1997.

[27] Standard ISO 9613-1. *Acoustics — Attenuation of Sound During Propagation Outdoors — Part 1: Calculation of the Absorption of Sound by the Atmosphere*. 1993.

[28] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3):1637–1647, 1992.

[29] V. Larcher, J.-M. Jot, and G. Vandernoot. Equalization methods in binaural technology. In *the 105th Audio Engineering Society (AES) Convention*, San Francisco, USA, Sept. 26-29 1998. preprint no. 4858.

[30] J. Huopaniemi. *Virtual acoustics and 3-D sound in multimedia signal processing.* PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, report 53, 1999.

[31] J. Blauert. *Spatial Hearing. The psychophysics of human sound localization.* MIT Press, Cambridge, MA, 2nd edition, 1997.

[32] M. R. Schroeder. Natural-sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3):219–223, 1962.

[33] R. Väänänen, V. Välimäki, and J. Huopaniemi. Efficient and parametric reverberator for room acoustics modeling. In *Proc. Int. Computer Music Conf. (ICMC'97)*, pages 200–203, Thessaloniki, Greece, Sept. 1997.

[34] J.-M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physique et perceptifs.* PhD thesis, l'Ecole Nationale Superieure des Telecommunications, Télécom Paris 92 E 019, Sept. 1992.

[35] D. Rocchesso and J. O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE Trans. Speech and Audio Processing*, 5(1):51–63, Jan. 1997.

[36] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine. Splitting the unit delay – tools for fractional delay filter design. *IEEE Signal Processing Magazine*, 13(1):30–60, Jan. 1996.

[37] E.M. Wenzel, J.D. Miller, and J.S. Abel. Sound lab: A real-time, software-based system for the study of spatial hearing. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5140.

[38] V. Larcher, J.-M. Jot, J. Guyard, and O. Warusfel. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition on hrtf data. In *the 108th Audio Engineering Society (AES) Convention*, Paris, France, Feb. 19-22 2000. preprint no. 5097.

[39] G. Lorho, J. Huopaniemi, N. Zacharov, and D. Isherwood. Efficient HRTF synthesis using an interaural transfer function model. In *the 110th Audio Engineering Society (AES) Convention*, Amsterdam, the Netherlands, May 12-15 2001. Accepted for publication.

[40] T. Lokki and H. Järveläinen. Subjective evaluation of auralization of physics-based room acoustic modeling. In *Proc. Int. Conf. Auditory Display (ICAD'2001)*, Espoo, Finland, July 29 - August 1 2001. Abstract submitted.

[41] T. Lokki and M. Karjalainen. An auditorily motivated analysis method for room impulse responses. In *Proc. COST-G6 Conference on Digital Audio Effects (DAFx-00)*, pages 55–60, Verona, Italy, December 7-9 2000.

[42] T. Lokki. Objective comparison of measured and modeled binaural room responses. In *Proc. 8th International Congress on Sound and Vibration*, Hong Kong, China, July 2-6 2001. Accepted for publication.

[43] R.R. Torres, P.U. Svensson, and M. Kleiner. Edge diffraction in room acoustics computations. In *Proc. EAA Symposium on Architectural Acoustics*, Madrid, Spain, Oct. 16-20 2000. paper AAQ13.