Using spatial information for the synthesis of the diffuse part of a binaural room impulse response

Lukas Aspöck[1], Sönke Pelzer[1], Michael Vorländer[1]

[1] *Institute of Technical Acoustics, RWTH Aachen, D-52074 Aachen, Germany, Email: las@akustik.rwth-aachen.de*

Introduction

For the synthesis of binaural room impulse responses (BRIR), the directional information about the sound source is usually included by convolving the direct sound and the early reflections with the corresponding HRTFs. For the late part of the BRIR, however, there is no standardized method for combining spatial information and the energy decay curve, e.g., obtained by a ray tracing algorithm. In some approaches, diffuse field HRTFs are used for the diffuse part of the impulse response, while others completely disregard spatial information in the reverberation tail. To answer the question, how detailed the spatial information (e.g., the HRTF filter length and resolution) has to be, methods for the spatial synthesis of the diffuse part were implemented and compared. Listening tests were conducted to identify a suitable method, that reduces the computational effort and is perceived similarly compared to a high quality synthesis.

Binaural Room Impulse Responses

It is often assumed that room acoustic simulation mostly depend on the input data and the simulation algorithms. However, if the simulation is used for auralization, the synthesis of the binaural room impulse response represents another important step. The simulations typically use a combination of an image source model and a variant of ray tracing, which is applied once the mixing time has passed [1]. For the image source model, used for direct sound and early reflections of the impulse response, the synthesis of the filter is straight forward: The calculated energy level of the reflection is weighted with a source and a receiver characteristic (e.g., a source directivity and a head-related transfer function). The ray tracing result however does not contain exact temporal information about reflections.

Reverberation Filter Synthesis

The ray tracing algorithms generate energy histograms for multiple frequency bands (octave bands). In the provided implementation the direction of all detected rays is logged (in so called *Directivity Groups*) and a probability density function is used to determine the direction of the incoming rays during a certain time slots. For all time slots (typical length: 1-10 ms), the reflection density is calculated and reflections are accordingly added to the impulse response. The concept of this algorithm is illstruated in Fig. 1, for details see [2].

By using the data of the directivity groups, directional information can be applied whenever a reflection is added

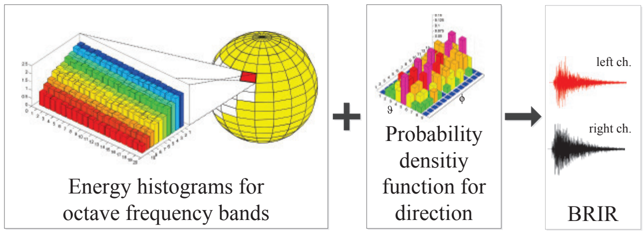

Figure 1: Filter synthesis of the reverberation based on frequency dependent energy histograms and directivity groups.

to the impulse reponse (instead of just adding weighted dirac impulses). The integration of HRTFs is possible, but as numerous reflections have to be accounted for in the reverberation tail, it leads to a high computational workload in relation to the total filter synthesis, which makes it less suitable for real-time applications. This filter synthesis algorithm, based on HRTFs, has been validated and compared to measurements in [3].

Directional Information

Directional information can be included in different levels of detail. In this research, all directional data is based on a HRTF measurement of an artificial head, in a 3° x 3° resolution and a length of 256 samples. To investigate how much effort has to be spent on the synthesis of the late decay, various datasets of gradually reduced complexity were created. This included shorter HRTF datasets with a length of 128 and 64 samples, both with a total energy equal to the original 256 sample HRTF database. Additionally, to reduce the measured data to only the most relevant spatial information, the interaural level difference (ILD) as well as the interaural time difference (ITD) were extracted from the HRTF dataset. The usage of the ITD, ILD or the combination of both datasets represent directional information based on significantly less data. A diffuse field HRTF does not include directional information, however it contains the average spectral information of the human receiver characteristic. The following list shows the investgated input data used for the integration of spatial information into the late part of a BRIR.

- HRTFs of different lengths: 256, 128, 64 samples
- Dirac impulses with ITDs + diffuse field HRTF
- Dirac impulses with ITDs + ILDs
- Dirac impulses with random ITDs + diffuse field HRTF
- Monaural reverberation (no spatial information)

Evaluation

Based on the different input data, BRIRs were created. In all cases, for the synthesis of the direct sound and the early reflections (image source order 1) the original HRTF dataset with 256 samples was used. BRIRs for three rooms, all with a relatively high amount of reverberation (T30 of 2.2s, 5.5s and 6.0s), were analyzed and convolved with pink noise bursts to create stimuli for a listening tests. A first analysis revealed that in time domain, the BRIRs do not show significant differences, in the frequency domain however, clear distinctions can be made, especially in higher frequency bands. In most cases these differences were audible while listening to the convolved files.

Room Acoustic Parameters

Although room acoustic parameters are meant for evaluation of monaural RIRs, the parameters were also analyzed for the generated BRIRs. It is expected that a reasonable filter synthesis should only lead to minor variations of the room acoustic parameters for the different input data sets. While for the reverberation time this holds true (variations below JND), more sensitive parameter such as the C80 vary up to 4 dB. This requires further investigation. In addition, also spatial parameters such as the interaural cross correlation should be evaluated.

Benchmarks

Fig. 2 shows the calculation times required for the synthesis of the late part of the BRIR (of a concert hall; filter length: 2 s, 335220 reflections in total) based on the different input datasets. The measurements were done on a *Intel Core-i7* desktop PC and repeated 10 times.

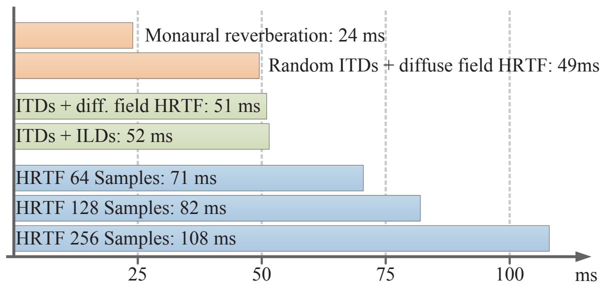

Figure 2: Calculation times for the synthesis of BRIRs using different directional input data for the late part of the BRIR.

Listening Tests

Listening tests have been conducted. In a 3AFC-test 21 participants were asked for audible differences between the generated stimuli (BRIR of different datasets convolved with pulsed pink noise) and a reference stimulus (convolved with BRIR based on the 256 sample HRTF). Each synthesis method was compared for three rooms and repeated once, resulting in a total amount of 48 questions. The stimuli were reproduced via headphones in a quiet laboratory. A boxplot showing the percentage of correctly answered questions is shown in Fig. 3.

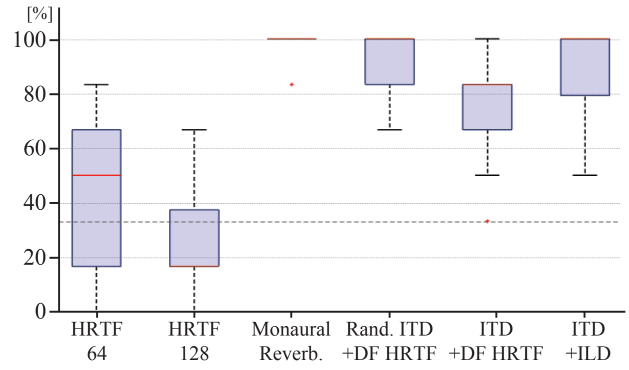

Figure 3: Listening test results: Percentage of correctly answered questions (detected different stimuli).

Conclusion

The results of the parameter analysis as well as of the listening tests show that small variations of the input data (e.g., changing the HRTF length) can lead to audible differences in auralizations of reverberant rooms. However, in more than 50% of the trails, auralizations based on short HRTFs (64 samples) were not detected. Most of the participants also described the auralizations based on the interaural chracteristics as plausible and as similar in comparison to the reference. Considering the test persons all were experienced listeners and had very good conditions to dinstinguish the auralizations (quiet laboratory, critical stimuli, static and reverberant room situations), also the synthesis methods based on the interaural characteristics might be suitable for dynamic real-time applications with multimodal feedback. This solution would be attractive due to significantly less computation cost (50%). Because of the clear result for the non-directional cases it can be stated that due to prominent spectral differences monaural reverberation is not sufficient in any case. The application of diffuse field HRTFs and a decorrelation by random ITD shifts improves the situation, but auralizations are still clearly distinguishable from the ideal reference case. In the future, a more efficient method for the synthesis of late part of the BRIR creating perceptually similar results will be implemented based on the results of this research. This implementation will combine temporally gradual spatial information.

References

[1] Lindau, A., Kosanke, L., Weinzierl, S., Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses, *Audio Engineering Society Convention 128*, 2010 .

[2] Schröder, D., *Physically Based Real-Time Auralization of Interactive Virtual Environments*, Ph.D. thesis, RWTH Aachen University, 2011.

[3] Pelzer, S., Aretz, M., Vorländer, M., Quality assessment of room acoustic simulation tools by comparing binaural measurements and simulations in an optimized test scenario, *Acta acustica united with Acustica*, 2011, vol. 97, no. S1, pp. 102–103.