Dynamic real-time auralization for experiments on the perception of hearing impaired subjects

Lukas Aspöck[1], Florian Pausch[1], Michael Vorländer[1], Janina Fels[1,2]

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

[2] *Forschungszentrum Jülich GmbH (INM-1), 52425 Jülich, Deutschland, Email: j.fels@fz-juelich.de*

Introduction

In recent years, the hearing aid industry developed rapidly, made use of digital signal processing and eventually contributed to improve the social situation of hearing impaired persons by providing adequate technical devices. However, even the latest devices still have a lot of potential for improvement. An important step towards such improvements is the generation of realistic sound environments, which can be carefully controlled and reproduced in laboratory environments. Examples of existing reproduction systems for investigation of the auditory perception (of hearing impaired subjects) can be found in [1] [2].

Various hearing aid-related fields can benefit from such realistically generated complex acoustical scenes. The audiologist can improve the fitting procedure as the test signals are closer to reality, generating difficult listening situations with multiple distracting sound sources. Developers of hearing aid algorithms can create a set of arbitrary acoustical scenarios to test and evaluate their implemented algorithms, e.g., for sound field classification or different beamforming approaches. A third application area is the field of research concerning auditory perception and cognitive processing of hearing impaired persons. Controlled complex acoustic situations help to investigate the perception and to establish a method to describe perceptual performance not only based on audiograms but on results of listening experiments in realistic situations.

Such controlled virtual acoustic scenes can be created by using simulation models based on geometrical acoustics, calculating the propagation paths from each virtual source to the receiver. More powerful computers as well as accelerated simulation models make it possible to use these simulation techniques in real-time applications on a single desktop computer [3]. In comparison to a measured and reproduced acoustical scene, a virtual scene allows interactive modification of the scene, enabling the audiologist, the researcher or the developer to continuously adjust the scene to his or her requirements.

This article presents an auralization system based on a framework for real-time room acoustics simulation which has been extended and adjusted for the use with hearing aids. Instead of using a conventional binaural synthesis based on Head-related Transfer Functions (HRTFs), transfer function measurements of Behind-The-Ear (BTE) hearing aids [4][5] are also considered to realize a multi-channel auralization with direct insertion of simulated signals into the hearing aid. This makes it possible to not only emulate the human perception (binaural hearing) of complex but controlled sound fields, but also to simulate how the sensors of the hearing aid receive the the signal of the same sound field.

Hearing aid auralization system

The difference between creating a normal virtual acoustic environment and an auralization system for hearing impaired (HI) is the usage of the hearing aid (HA). As the listening ability of HI varies strongly, also depending on different HA models and algorithms, it is necessary to use an emulated, but controlled, hearing aid system instead of the individual device for conducting experiments on the cognition of HI. In this way, all system elements up to the perception of the subject, the item of investigation, are known and can be controlled by the researcher. For this reason, the HA of the subjects is replaced by the real-time environment *Master hearing aid* (MHA, [6]) and a dummy BTE reproduction unit (dummy hearing aid, DHA) in this project.

The usage of the DHA makes it also possible to directly insert the simulated signal into the device instead of using the microphones. As the external sound field cannot be accurately reproduced for multiple receiving points (ear canal and BTE microphone positions), the usage of the microphones would lead to an incorrect simulation of the scene and would make it impossible to include different receiving characteristics of HAs.

Concept

The concept of the hearing aid auralization system is depicted in Fig. 1. A virtual indoor or outdoor scene, containing multiple sound sources with assigned signals, directivities and trajectories, is defined and loaded via a user interface. Interactivity evolves as the listener moves or the sound sources are adjusted during the auralization. An optical tracking system is used to constantly adjust the receiver-related auralization to the position and orientation in the real world. The simulation engine generates multi-channel room impulse responses, containing the reflection paths from the virtual sound sources to each receiving points. In case of a common BTE device, two microphones on the HA and the ear canal, resulting in six positions for both sides in total. For the filter synthesis, databases of these transfer functions, describing the sound propagation from a sampled sphere to the receiving point, have to be provided. The first two channels (ear canal) correspond to the conventional HRTF while the other four channels are the Hearing Aid-Related Transfer Functions (HARTF).

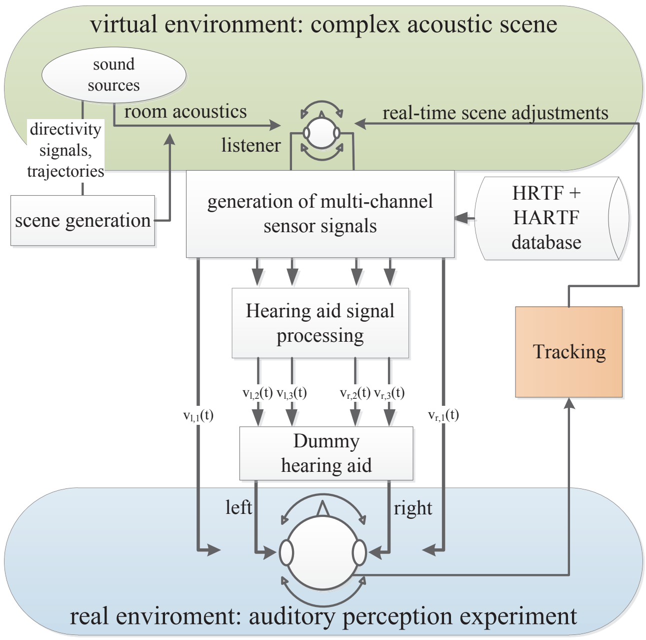

Figure 1: Overview of the hearing aid auralization system

The multi-channel room impulse responses are convolved with the corresponding signals of the sound sources before they are reproduced. While the conventional binaural channels (ear canal) are played back through a transaural loudspeaker system, the additional four channels are processed and then reproduced through the DHA to the HI listener, who receives the virtual scene in the same way as in reality.

System requirements

The system, which is mainly developed for cognitive and auditory experiments with HI subjects, comes with various requirements. The subjects have to be located in a quiet and controlled laboratory environment such as a listening booth. Because of the limited space and the specification to easily set up the system at different locations, the hardware requirements must be kept at a minimum level. Thus, for the standard setup, only one desktop computer with an external sound card, two to four loudspeakers with stands and an optical tracking system with four cameras are chosen.
To realize a natural and interactive situation for the subjects, the auralization and hearing aid simulation software have to operate under real-time conditions and react to user interactions in less than 50 ms. For this reason, the calculations of the direct sound path of moving virtual sound sources are updated with rates between 60-100 Hz. For the calculation of the early reflection paths (10 Hz) and the reverberation (1 Hz) the update rates can be lower.
A programmable real-time module including various hearing aid algorithms is required to control and adjust the hearing aid signal processing. The selection of (individual) processing methods can then be configured and chosen according to the type of perceptual experiment.

Auralization concept

One obvious choice to auralize a complex acoustic scene is to measure a spatial sound field and reproduce it through a loudspeaker array [7]. The recorded scene, however, does not allow for modifications once the measurement is completed. The effort of creating and modifying the scenes could be reduced by using simulations (rendering for higher-order Ambisonics or Vector-Base Amplitude Panning) instead of real measurements. Though, this reproduction techniques do not fulfill the system requirements as it ideally works with a larger number of loudspeakers and as it is not compatible with the hearing aid processing path of the system. Because of the limited amount of reproduction hardware and the desire to provide a natural listening sensation (accounting for translatory and rotatory movements of user), the auralization has to be based on a real-time binaural room acoustics simulation. Based on a provided geometric model of the environment, the sound propagation paths from sound source to receiver are calculated and specified in a binaural room impulse response (BRIR). Although various approaches for conventional binaural synthesis exists [8][9], also for complex acoustic scenes [10], none of these can be directly applied for HA auralization.

BRIR-based auralization

The simulation engine is responsible for calculating a BRIR for each sound source of the scene. These calculations are based on a hybrid simulation approach combining an image source model for the early reflections and a ray tracing algorithm for the reverberation. Both models account for absorption and scattering properties of wall materials and are included in the C++ software library *RAVEN* [11]. The simulation components comprise the direct sound (audibility flag), early reflections (list of audible image sources) and reverberation (echograms for octave bands, including spatial information).

Filter Synthesis

As the existing implementations were designed for standard two-channel binaural synthesis, a new efficient multi-channel filter module was implemented, which separately creates a filter for direct sound, early reflections and reverberation. The number of output channels is equal to the number of channels of the provided receiver characteristic, which is stored as an impulse response database at a defined angular resolution using the *openDAFF* file format [12]. For the HA auralization, the receiver directivity database has six channels in total, two HRTF and four HARTF channels, and a resolution of 3° (azimuth and elevation). The synthesis procedure also includes the source directivity (third octave magnitude spectra) and individual configuration possibilities. As for each incoming reflection an individual receiver-related impulse response is applied, the computational effort of the reverberation filter is much higher in comparison to the direct sound and early reflections filter. Because of this, three different reverberation modes are implemented: (1) Individual reflections for the entire impulse response [13], (2) direction dependent reflections up to a point in time related to the perceptual mixing time [14], followed by a static reverberation tail, or (3) a static reverberation filter for the entire impulse response. The static reverberation filter can be precalculated for one exemplary situation or can be based on a diffuse-field impulse response of the receiver.

Signal processing and reproduction

Fig. 2 shows the signal flow starting from the simulated signals and ending at the listeners ears. Once the multi-channel filters for the current situation are synthesized and convolved with the corresponding anechoic signal, the binaural channels are passed to a crosstalk cancellation (CTC) network and reproduced through loudspeakers. This represents the external sound field, which can also be used in experiments involving normal-hearing subjects. The HA-related channels are passed to the MHA module for the HA signal processing before they are reproduced by the receivers of the DHA. In case of different processing times of the CTC network and the MHA module, a delay compensation is applied. Here, the output does not necessarily have to be synchronized, but could also be set to a latency of a typical hearing aid (below 10 ms).

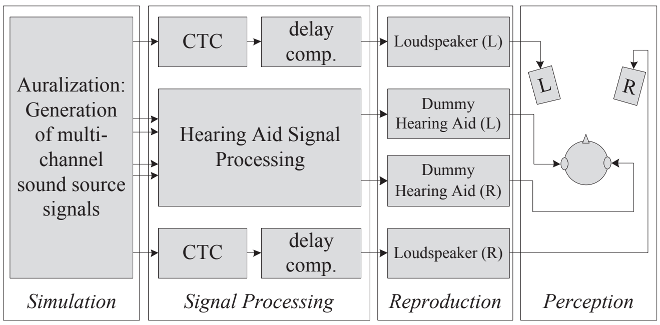

Figure 2: Signal flow from simulation to listener

Performance analysis

With the aim of implementing an interactive auralization system, it is important to investigate the processing performance of the system. In a first step, the room acoustical simulation and the multi-channel filter synthesis is analysed. While the simulation is identical to conventional binaural simulation, the filter synthesis is adjusted to the HA auralization concept.

All calculations were done for an example restaurant scene (387 polygons) with three sound sources on a common desktop computer (*Core i7* @ 3.40 GHz, 8 GB RAM, *Windows 7*). To account for other simultaneously running tasks, only 33% of the CPU time is assigned to the simulation engine. This value might vary depending on the current workload, but has been established as a reasonable estimation. The results for the simulation (see Table 1) show that the direct sound (DS) can be updated at very high rates while the update rates of the early reflections (IS, approx. 55 audible image sources per sound source) are satisfying. If higher update rates are desired the image source order could be decreased to 1. The ray tracing (RT) calculation times are significantly higher, the calculation of a physically correct energy decay of the room (statistical level deviation of 1.37 dB) for each sound sources of the scene would allow update rates of only 0.16 Hz. Even the calculation of a short ray tracing up to a time of 200 ms an update can only be provided every two seconds. This time is determined in dependency of the perceptual mixing time. As it was recently shown [15] that changes in a BRIR even past the perceptual mixing time can be distinguished, a factor of 6 was chosen conservatively. However, even if further perceptual investigations proof that this factor can be reduced, it is a reasonable choice to use precalculated data and to avoid an extensive real-time calculation of the reverberation tail. Though, in rooms with a simple shape the reverberation does not vary strongly for different receiver and source positions, it is sufficient to update the reverberation at low update rates [16].

Table 1: Calculation times and update rates for room acoustics simulation

Task	Source #0	Source #1	Source #2	∑	Update rate (33% CPU time)
DS Audibility	0.73 µs	2.03 µs	0.29 µs	3.05 µs	10.93 kHz
IS - 2nd Order Audibility	10.09 ms	12.35 ms	9.55 ms	31.99 ms	10.42 Hz
IS - 2nd Order (moving sources)	16.83 ms	16.15 ms	16.27 ms	49.25 ms	6.77 Hz
RT - 2000 ms 3000 Particles	678.18 ms	692.57 ms	663.79 ms	2034.54 ms	0.16 Hz
RT - 200 ms 3000 Particles	257.05 ms	258.12 ms	244.09 ms	759.26 ms	0.44 Hz

Table 2 shows the calculation times of the multi-channel filter synthesis of all three simulation components. A filter update is required frequently, i.e. in case of a user rotation. While the DS and IS filter synthesis allow very high update rates, the standard reverberation synthesis (1) does not allow update rates above 1 Hz. Changing the configuration to the mixing-time dependent implementation (2) increases the update rate to a satisfactory level. The calculation time of the static filter (3) is not shown here as it is calculated during the initialization process and does not require updates.

In rooms with simple shape and static sound sources, the echograms as well as the multi-channel filters should be calculated in advance and exchanged in case of movement of the subject. This does not provide a physically correct simulation of the reverberation for all positions but it leads to a plausible and interactive representation of the virtual scene with at a low latency.

Conclusion

A concept and parts of the implementation of a real-time simulation environment for experiments with hearing impaired subjects were presented. In addition to the conventional binaural auralization, the simulated signal is

Table 2: Calculation times and update rates for six channel filter synthesis

Task	Source #0	Source #1	Source #2	Σ	Update rate (33% CPU time)
DS filter 6ch, 128 samples	0.12 ms	0.12 ms	0.12 ms	0.36 ms	925.93 Hz
IS filter 6ch, 128 samples	0.60 ms	0.58 ms	0.59 ms	1.77 ms	188.33 Hz
Reverb filter #1 6ch, 2000 ms individual refl.	178.10 ms	179.43 ms	179.32 ms	536.85 ms	0.62 Hz
Reverb filter #2 6ch, 200 ms + static Reverb	18.67 ms	20.14 ms	20.23 ms	59.04 ms	5.65 Hz

processed by a hearing aid simulation environment and reproduced by a dummy hearing aid. This makes it possible to investigate the perception of realistic complex sound fields via hearing aids. To allow natural behavior of the listener, the auralization is constantly adapted according to the subject's position and orientation, which is detected by an optical tracking system.
The auralization is based on a real-time simulation and a filter synthesis of multi-channel room impulse responses, which fulfills the requirements of scene adjustments and a high level of interactivity. A new filter module adjusted to the demands of the generation of the simulated hearing aid signals was implemented.
The performance analysis showed that a multi-channel direct sound and early reflections simulation of three dynamic sound sources can be calculated at sufficiently high update rates. Depending on the complexity of the scene simplified concepts or precalculated filters should be considered for the calculation of the reverberation.
The next steps involve the implementation, configuration and evaluation of the hearing aid processing as well as of the reproduction systems before the complete system can be evaluated and finally used for experiments with hearing impaired subjects.

References

[1] Ericson, M.A., Multichannel Sound Reproduction in the Environment for Auditory Research, *Proceedings of 131st Convention of the Audio Engineering, New York*, 2011.

[2] Oreinos, C., Buchholz, J.M., Validation of realistic acoustic environments for listening tests using directional hearing aids, *Proceedings of 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014.

[3] Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., Aloza, D., Progressive Perceptual Audio Rendering of Complex Scenes, *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2007.

[4] Kayser, H., Ewert, S.D., Anemüller, J., Rohdenburg, T., Hohmann, V., Kollmeier, B., Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses, *Eurasip Journal on Advances in Signal Processing*, 2009.

[5] Oreinos, C., Buchholz, J.M., Measurement of a Full 3D Set of HRTFs for In-Ear and Hearing Aid Microphones on a Head and Torso Simulator (HATS), *Acta Acustica united with Acustica*, 2013, vol. 99, no. 5, pp. 836–844.

[6] Grimm, G., Herzke, T., Berg, D., Hohmann, V., The Master Hearing Aid: A PC-Based Platform for Algorithm Development and Evaluation, *Acta Acustica united with Acustica*, 2006, vol. 92, no. 4, pp. 618–628.

[7] Merimaa, J., Pulkki, V., Spatial Impulse Response Rendering I: Analysis and Synthesis, *J. Audio Eng. Soc*, 2005, vol. 53, no. 12, pp. 1115–1127.

[8] Menzer, F. (editor), *Efficient Binaural Audio Rendering Using Independent Early and Diffuse Paths*, 2012.

[9] Aspöck, L., Pelzer, S., Vorländer, M., Using spatial information for the synthesis of the diffuse part of a binaural room impulse response, *Proceedings of Fortschritte der Akustik : 40. Deutsche Jahrestagung für Akustik, Oldenburg*, 2014.

[10] Nicholas Tsingos, Perceptually-based auralization, *Proceedings of 19th Intl. Congress on Acoustics 2007, Madrid*, 2007.

[11] Schröder, D., Physical-based real-time auralization of interactive virtual environments, *PhD Thesis, RWTH Aachen University*, 2011.

[12] Wefers, F., OpenDAFF - A free, open-source software package for directional audio data, *Proceedings of Fortschritte der Akustik : 36. Deutsche Jahrestagung für Akustik, Berlin*, 2010.

[13] Aspöck, L., Pelzer, S., Wefers, F., Vorländer, M., A real-time auralization plugin for architectural design and education, *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin*, 2014.

[14] Lindau, A., Kosanke, L., Weinzierl, S., Perceptual evaluation of model- and signal-based predictors of the mixing time in binaural room impulse responses, *J. Aud. Eng. Soc.*, 2012, vol. 60, pp. 887–898.

[15] Stade, P., Perzeptive Untersuchung zur mixing time und deren Auswirkung auf die Auralisation, *Proceedings of Fortschritte der Akustik : 41. Deutsche Jahrestagung für Akustik, Nürnberg*, 2015.

[16] Lentz, T., Schröder, D., Vorländer, M., Assenmacher, I., Virtual Reality System with Integrated Sound Field Simulation and Reproduction, *EURASIP Journal on Advances in Signal Processing*, 2007.