ACOUSTICS 2012

Auralization of a virtual orchestra using directivities of measured symphonic instruments

S. Pelzer, M. Pollow and M. Vorländer

RWTH Aachen University, Institute of Technical Acoustics, Templergraben 55, 52056 Aachen, Germany
spe@akustik.rwth-aachen.de

Room acoustic simulation algorithms have evolved to powerful tools in recent years. Additionally to traditional tasks, such as prediction of room acoustic parameters (e.g. reverberation time, strength etc.), they are commonly used for auralizations nowadays. For a lively and natural virtual representation of real sound sources, it is important to include their individual characteristics in terms of directivity patterns and spatial dynamics.
For a high quality auralization of a full orchestra, the radiation patterns of symphonic instruments were captured using a surrounding spherical array and stored in the free OpenDAFF directional file format. Using a hybrid image sources and ray tracing method, room impulse responses (IRs) were calculated for each orchestra instrument. The player's movements were simulated by applying an artificial "humanization" to the position and orientation vectors.
The simulation generates binaural signals as well as spatial impulse responses in spherical harmonics (SH) format, which can then be convolved with anechoic recordings of each part. Using the flexible SH representation and head-tracking, a dynamical and immersive auralization can be achieved that reacts on the listeners head movements.

1 Introduction

Modern algorithms for predicting sound propagation in rooms have evolved to powerful tools. Their application is now going far beyond typical tasks of room acoustic quantities estimation (SPL, reverberation time, clarity etc.). The integration of auralization algorithms opened a new field of research, mainly for psychological and psychoacoustical studies, and of course in room acoustics analysis. One of the highest benchmarks for an auralization program is probably the realistic simulation of a complex reverberant space, with multiple natural sound sources and many listeners - a concert hall situation. In this contribution, an exemplary auralization of an orchestra is described step by step, showing how to create a multi-purpose data basis that can be used e.g. for room acoustical and psychoacoustical research.

1.1 Auralization applied in research

Many researchers use the combination of simulation and auralization since simple models can be implemented in MATLAB and also mature tools are available in the form of commercial room acoustics simulations (e.g. CATT, EASE, Odeon). Recent investigations addressed various topics in the context of multiple sources with room acoustics interaction, only a few shall be mentioned here: the number of point sources necessary for representing large ensembles (Lokki [1]), the importance and effect of directivities (Wang [2]), the orchestra formation and signal decorrelation (Vigeant [3]). Most of these studies conclude with still many open questions. For further experiments it would be advantageous to use a flexible transfer path model which is presented in the following section.

2 Acoustic transfer path structure

The auralization process can be divided into three elementary parts. At first, a representation has to be found for describing the the radiation of sound waves from a source. In a second step, the propagation of these sound waves is predicted using a CAD model of the surrounding space and physical models for the interaction of sound with fluid and solid media (usually air and room surfaces). Finally, the sound pressure level at arbitrary points should be available, with respect to characteristics of natural receivers, such as microphones, human listeners or an artificial head. These three connected parts, consisting of (a) source representation, (b) sound propagation / spatial impulse response, and (c) receiver implementation / reproduction of spatial 3D sound, can be handled clearly separated with the big advantage, that each component will be flexible for changes and variations, while the other parts can be kept constant. This principle allows focused investigation on a single item without having additive external influences introducing systematic errors.

3 Sound sources

In common geometrical acoustics models, sources are modeled as surface or point sources, with optional directivity. For most natural sources this is a valid simplification, as long as the real source's extent is small compared to the distance between source and listener. This holds true for most symphonic instruments and the audience in a concert hall, up to the director's position. As musical instruments usually radiate energy in predominant directions, e.g. when it comes to members of the brass section, an auralization using omnidirectional point sources can lead to poor results. In reverberant spaces, Otondo and Rindel found a disturbing change in timbre for this inaccurate source representation [4].

3.1 Instruments directivity measurements

Therefore it is important to apply the natural directivity pattern to each single source. Many attempts have been made to capture the directional energy distribution of symphonic instruments [5, 6, 7, 8]. Recently, we conducted directivity measurements together with colleagues from TU Berlin, featuring the whole range of symphonic instruments, and including comparison of modern and ancient instruments. The data was then interpolated and converted into the OpenDAFF[1] [9] file format and into Spherical Harmonics (SH) domain. OpenDAFF features very fast access functions for use in virtual reality systems, while the SH domain offers an unambiguous, flexible and generalized spatial representation. Details of the measurements as well as its post-processing are published by Behler and Pollow [19, 20].

3.2 Spherical Harmonics representation

Following an idea, initially proposed by Weinreich in 1980 [10], it is possible to decompose the source radiation into orthonormal base functions on the sphere (spherical harmonics, SH), deriving a spatially continuous solution for the measured radiation pattern. Main advantages of this approach are the flexibility of the components, so that general sources with this orthonormal base can be used in the simulation, and

---

[1] Open Directional Audio File Format: http://www.opendaff.org

the directivity can be easily exchanged after the simulation. This can be used to quickly render auralizations with different source directivities or different properties of these directivities, such as variable spatial resolutions or different source rotations with cheap matrix operations. This also allows a practicable implementation of time-variant directivites in auralizations, which are important for certain sound sources. For example in percussions, the Tam-Tam changes its modal behaviour over time and the Cymbals are usually held up facing the audience after each crash, which drastically changes the radiation character. Also in other areas we encounter time-variant sound radiation, e.g. when auralizing engines or any rotating equipment in different operating conditions [11, 12].

3.3 Anechoic recording of symphonic instruments

Only few real orchestral recordings exist that can be used in auralizations. The requirements to the material are challenging: a full orchestra is needed, sources need to be recorded separately one after each other,n ear-field effects should be avoided, a high signal-to-noise ratio is desirable, and a fully anechoic chamber has to be used for the recordings. Thus only few attempts have been made to record orchestral music in anechoic conditions.[2] But while some did not capture the whole orchestra [13], others did not have enough source separation. Crosstalk is usually too high, when the orchestra is recorded simultaneously with close-up microphones, as shown in Figure 1 during an in-house recording session at ITA Aachen [14]. The crosstalk problem is also present in recordings by Denon in the late 1980's, which also suffer from high background noise [15]. All mentioned recordings share the exposure to floor reflections, too.

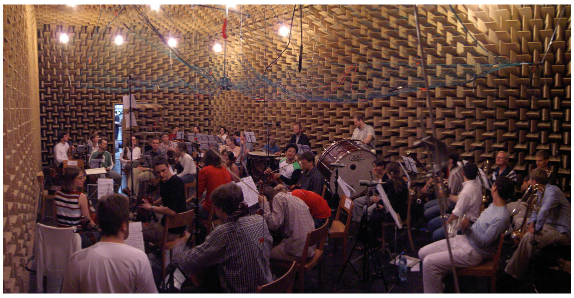

Figure 1: Anechoic recordings of a symphony orchestra at ITA Aachen.

Only recently, Pätynen and Lokki succeeded with recordings that meet most of the requirements stated above, and he also published the results for academic purposes [16]. The recordings feature single tracks of four movements by Beethoven, Bruckner, Mahler and Mozart, and these were used in the present study. Due to the high time consumption of these recordings, each part was recorded only once by one single musician. This means that for a full string section, only one track is available. But the multitude of instruments that play the same part usually shapes the rich sound of a string section in symphonic music. Variations in pitch and timing, together with the varying sound of different instruments and players contribute to a merging chorus effect [5].

Pätynen and Vigeant proposed to apply phase- and pitch-shifting on the single tracks to create copies that can be used for multiple point sources [16, 3]. Surprisingly, their listening experiments revealed no significant differences to non-processed copies of one instrument. This was suspected due to the coloration and mixing by the multiple reflections in the auralized room, especially in combination with the full orchestra playing. These observations suggest that single tracks can be copied without decorrelation or other processing for full orchestra auralizations. More important than the processing of a single recording for the full section seems to be a representation with multiple point sources, one for each musician [1].

4 Receivers

At the end of an acoustic transfer path a receiver implementation is needed. To capture the energy at a certain point (listener position), the energy spectrum of incident sound waves is stored including the time of arrival. This energy decay curve can be evaluated to calculate the sound pressure signal for omnidirectional receivers, such as measurment microphones.

4.1 Binaural signals

For auralizations, the simulation has to provide binaural signals. The receiver is thus modeled as an artificial head, using its head-related transfer functions (HRTFs). Due to the HRTFs, this type of receiver has a certain directivity, so that it is important to know the incident directions of sound waves. To allow different receiver types, e.g. different artificial heads, individual HRTFs, and directional microphones, the energy decay curves can be stored in a general frequency-, time- and direction-dependent data structure, as depicted in Figure 2.

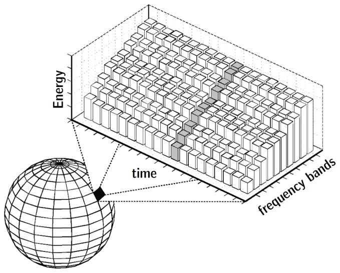

Figure 2: Spatial receiver implementation using energy histograms in a general frequency-, time- and direction-dependent data structure.

When listening to binaural signals, which are produced using the own individual HRTFs, the degree of immersion can be significantly higher due to precise source localization, less unnatural coloration and reduced in-head localization. It is therefore advisable to present auralizations with individual HRTF data. To be able to individualize an already finished simulation, e.g. for better immersion in auralization or as part of listening tests, we can resort again to the spherical harmonics representation, which offers the generalized spatial format for a maximum of flexibility. To capture individual

---

[2] Anechoic recordings of singers are already available. Furthermore, when a free choice of musical content is more important than a reflection-free recording, a full orchestra can be synthesized using sample-based libraries (e.g. Vienna Symphonic Library: http://vsl.co.at)

HRTF data, we presented a measurement system, that allows measuring nearly full-sphere HRTFs in less than 5 minutes, which makes conventional use of individual HRTFs feasible. Details about this system are described by Pollow and Masiero [18, 17].

4.2 Spherical harmonics representation

In the simulation, receivers can be easily implemented in SH domain by successively applying all base functions as receiver directivities. The number of output channels is increased from one/two channels (omnidirectional/binaural receiver) to $N = (M+1)^2$ channels for each base function of SH order M. If desired, high orders are possible as these channels can be generated all at once in one simulation run.

The advantages of an SH receiver implementation are mainly a reproduction independent format. With a real-valued base the SH coefficients can be used directly, comparable to the output of a higher-order spatial microphone (e.g. Soundfield or Eigenmike). In a simulation the order can be much higher than with real microphones, so that the spatial resolution can exceed the localization abilities of the human hearing system.

However, the necessary SH order for binaural reproduction is still not known yet. The goal of the method presented in this contribtion is to faciliate furher work in the topic of necessary spatial resolution and SH orders. Due to the continuity of the SH representation, there is no need for interpolation and it is possible that already low orders provide accurate localization. However, the maximum order used in simulations so far was $M = 69$, which leads to 4900 base functions. The data was derived from discrete measurements on a Gaussian grid with resolution of approximately 2.3 x 2.3 degrees.

4.3 3-D spatial reproduction

The higher-order SH coefficients can be used for loudspeaker reproduction, e.g. using an Ambisonics decoder or spatial processing as proposed by Pulkki (DirAC [21]). At the same time it is possible to render binaural downmixes by just multiplying the SH coefficients with HRTF data in SH domain, enabling direct reproduction through headphones or a crosstalk cancellation network. This includes the application of individualized HRTFs on already finished simulations.

When using a head-tracked headphone system, the immersion of the listener can be considerably higher in the auralized scene. In addition, the human localization abilities are sharpened by small unconscious head movements. This interactive reproduction can be easily achieved as the rotation of the HRTF data in SH domain can be done by a cheap multiplication with the Wigner-D matrix [22]:

$$h_{nm,\mathrm{rot}}(f) = D^{m}_{k,n}(\alpha,\beta,\gamma)\cdot h_{nm}(f) \qquad (1)$$

The SH order of the simulation can be kept variable as well as the SH order of the receiver directivity (e.g. HRTFs). This enables a maximum in flexibility. Quick simulations can be done with low orders, while a final rendering can be performed with very high orders, to ensure a best possible localization, if high-order HRTF data is available. This concept allows for straightforward evaluations on the necessary spatial resolutions in listening tests.

A further idea is to use different SH orders in one single impulse response. With high orders for the direct sound and early reflections a good source localization is ensured, while low orders are sufficient to model the diffuse late reverberation. This idea of hybrid reproduction was proposed by Favrot in 2010 [23]. To carry the idea a bit further, we propose to apply different reproduction techniques for early and late parts of the impulse response. This hybrid simulation and playback with crosstalk-cancelling and Ambisonics was presented in 2011 [24].

5 Sound propagation simulation

In general, room acoustic simulation by using geometrical acoustics is implemented with binaural receivers and a given sound source directivity. This way, however, the simulated signals are restricted to a specific set of HRTFs and a fixed radiation pattern of the simulated source. Adjustments such as individualization of the HRTFs or modification of the source directivity pattern cannot be performed after the simulation is finished. Therefore the described general interface between transfer path and the directivities of both sender and receiver as basic radiation patterns (SH) is implemented into RAVEN, a simulation framework under development at ITA Aachen [25].

This geometrical acoustics tool combines an image source method for the realistic representation of early specular reflections with a stochastic raytracing approach to model the diffuse, scattered reflections in the late part of the room impulse response. Sound sources are modeled by inclusion of their directivity and free field sensitivity. By further accounting for the angle of incidence of each sound ray hitting a receiver and making use of SH, the software allows to generate spatial room impulse responses with arbitrary resolution.

The number of sources and receivers is only limited by the available system memory. Computation load is nearly independent of the number of receivers, while it scales linearly to the number of sound sources. An orchestra auralization with many sources is therefore challenging in terms of CPU and memory load. To meet these requirements, the real-time simulation tool RAVEN offers an MPI[3] extension, so that all impulse responses can still be delivered on time [26]. But for many typical users, there will be no compute cluster available and probably no real-time computation necessary, as the impulse responses in a listening test are mostly calculated in advance anyway. Therefore we focus on how an immersive auralization can be done on a single personal computer. By pre-computation of impulse responses for multiple receiver positions (less costly compared to multiple sources), it is possible to allow for listener movements by interpolation of these positions [27].

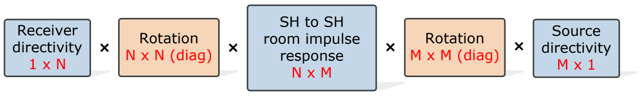

Figure 3: Room acoustics transfer path simulation by multiplication of spherical harmonics coefficients.

---

[3] Message-Passing Interface: www.mcs.anl.gov/research/projects/mpi

5.1 Spherical harmonics representation

A source and receiver SH input and output allows for more lively auralizations by enabling cheap listener rotations (useful for head-tracking) and also source rotations. Latter one can be used to add some ”humanization” to virtual musicians by adding a lively movement to the pointing direction of the virtual instruments. In SH domain, these rotations are done by a simple matrix multiplication, as shown in Figure 3.

A simulation with S sources and R receivers in SH domains results in $S \cdot R$ matrices with dimensions $N_S \cdot N_R \cdot B$, with N_S and N_R being the number of SH base functions of source and receiver, respectively, and B the number of frequency bins. When sources and receivers are consequently modeled both using SH, all operations concerning the directivity, its resolution and rotation, as well as the source signals can be exchanged after the simulation. This can be used to improve the reproduction of auralizations with head-tracking and individual HRTFs, but also has significant advantages for listening tests investigating individual elements of the transfer path and their perceptual properties.

5.2 Implementation

Source directivities of SH base functions were implemented using the existing interface for directional sources, which uses OpenDAFF files. Receivers are for Ambisonics reproduction already modeled using SH base functions and ready for synchronous calculation of all order’s channels. By MATLAB scripting, the source directivities can cycle through all base functions, while one simulation processes all sources and receivers and all receiving base functions paralelly at once. Image sources as well as ray tracing support shared-memory paralellization using OpenMP[4].

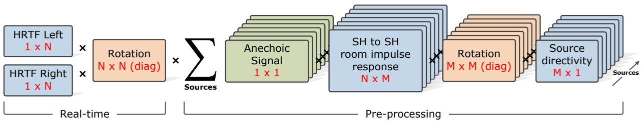

Figure 4: Concept of a full orchestra real-time auralization using spherical harmonics representation of transfer path, source directivities and receiver HRTFs.

5.3 Real-time auralization

For an immersive real-time auralization of the simulation results with a high number of sources (orchestra auralization), all operations with very high computational costs have to be done in a pre-processing step. This concerns mainly the simulation of spatial room impulse responses as well as the convolution of several minutes of recorded solo instruments with long room impulse responses for all combinations of spherical harmonics base functions. The entry point for the real-time operations in the auralization chain is right after these convolutions, as shown in Figure 4. The receiver directivity is still variable to enable various HRTFs (different artificial heads, individual HRTFs, different resolutions etc.) as well as free rotation. The remaining operations (rotation and HRTFs) are simple multiplications in the SH domain and thus real-time capable even for high spatial resolutions. In addition, the HRTFs are stored in a very compact format by separating the delay and interaural time-differences from the impulse response, resulting in filters with only 65 frequency bins. Using SH there is no need to interpolate data due to the spatially continuous definition. This is interesting especially when dealing with low orders. For the full real-time auralization, another advantage is that the number of sources and the length of the impulse response do not affect the computation load. The convolutions which are very expensive for many sources and long room impulse responses and their superposition is done in a pre-processing step.

Compared to a pure binaural simulation, only the full SH approach offers this potential for pre-computation with a flexible result. If no such flexibility (e.g. for reproduction independence of loudspeaker setup, tracked headphones etc.) is intended, the contruction of static binaural filters is very lightweight and results after convolution with source signals and superposition in only two channels. But for immersive simulations with a head-tracker, this binaural filter has to be updated on every head rotation. This requires to exchange the HRTFs for the direct sound and each single reflection of all sources in the binaural room impulse response (BRIR). These BRIRs have to be convolved in real-time for every source signal, which can be performed on a single PC only for a very limited number of sources (approx. less than 5 on current hardware).

5.4 Performance

With the SH approach and pre-processing, any number of sources and any length of room impulse responses can be auralized. With high SH orders, the available disk space can become problematic, as an order of 30 results alreaedy in nearly 1000 channels, so that 1 Minute of orchestral music takes 10GB of disk space.

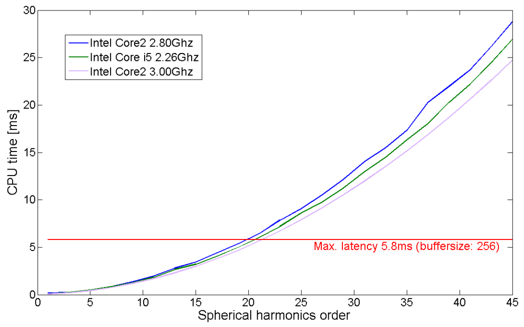

Figure 5: Used CPU time for real-time convolutions of spherical harmonics with different orders.

The real-time convolution of these 1000 channels with HRTFs of 256 samples length took 11.7ms per frame on a single core of an Intel Core2Quad 3.00Ghz CPU. Unfortunately, with an ASIO buffer size matched to the 256 samples of the convolved excerpt, the sound card driver will ask for new data every 5.8ms so that no continuous playback is possible. This problem could be solved with using all 4 cores by parallelization. But on the other hand, the SH concept allows auralization also on slow machines by not using the full maximum order. Convolving only the first 21 of all available orders (484 channels in total) takes only 5.7ms of CPU time, so that this can be done in real-time, as shown in Figure 5. This fully scalable framework enables even weak systems to auralize all sound sound sources without artifacts or dropouts, just with reduced spatial resolution as necessary.

---

[4] OpenMP. Application Program Interface Version 3.0, www.openmp.org

6 Conclusion and Outlook

We propose to apply generalized simulations whenever it can be advantageous to vary the spatial characteristics of sources or receivers after the simulation is finished. Therefore all sources and receivers directivities are modeled using spherical harmonics, so that a single spatial room impulse response is represented by a NxM matrix for an M^{th} order source and an N^{th} order receiver. This fully flexible transfer path can be used for immersive real-time auralization by multiplying HRTFs in SH domain with respect to the current view direction of the listener (head-tracking assumed). Scaling the rendered SH order to the available system resources is possible, independent of the number of sources or room impulse response length. The necessary order for accurate spatial reproduction of sources and receivers is not yet known, but using the presented method, a listening experiment design is straight-forward and planned for the near future. Further research using the technique can aim at perceptual effects of musicians movements, time-variant directivities, signal conditioning (for multi-channel point sources), spatial reproduction techniques and room influence (volume, mean absorption, diffuseness).

References

[1] T. Lokki, "How many point sources is needed to represent strings in auralization?", International Symposium on Room Acoustics (ISRA 2007), Seville, Spain, 2007

[2] L. M. Wang, M. C. Vigeant. "Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities", Applied Acoustics, Vol. 69, Issue 12, 2008

[3] M. C. Vigeant and L. M. Wang. "Investigations of multichannel auralization technique for various orchestra arrangements, with phase-shifted string sections", International Symposium on Room Acoustics, 2007

[4] F. Otondo, J. H. Rindel. "The influence of the directivity of musical instruments in a room", Acta Acustica united with Acustica 90, No.6, 2004

[5] J. Meyer. "The sound of the orchestra", J. Audio Eng. Soc. 41, 203-213, 1993

[6] F. Otondo and J. H. Rindel. "A new method for the radiation representation of musical instruments in auralizations", Acta Acustica united with Acustica 91, 2005

[7] J. Pätynen J., T. Lokki. "Directivities of Symphony Orchestra Instruments", Acta Acustica united with Acustica 96 138-167, 2010

[8] M. Noisternig, F. Zotter, B. Katz. "Reconstructing sound source directivity in virtual acoustic environments", Principles and Applications of Spatial Hearing, World Scientific Publishing, 2011

[9] F. Wefers. "OpenDAFF - Ein freies quell-offenes Software-Paket für richtungsabhängige Audiodaten", Fortschritte der Akustik, DAGA, 2010

[10] G. Weinreich and E. B. Arnold. "Method for measuring acoustic radiation fields", J. Acoust. Soc. Am. 68:2, 404-411, 1980

[11] M. v. d. Giet, J. Blum, P. Dietrich, S. Pelzer, M. Müller-Trapet, M. Pollow, M. Vorländer, K. Hameyer. "Auralization of electrical machines in variable operating conditions", IEEE International Electric Machines & Drives Conference (IEMDC), 2011

[12] F. Wefers and M. Vorländer. "VATSS - Ein Virtual Reality System zur interaktiven Simulation von Fluglärm", Fortschritte der Akustik, DAGA, 2012

[13] CD Bang&Olufsen 101. Music for archimedes, 1992

[14] I. Witew, J. Paprotny, G. K. Behler. "Auralization of Orchestras in Concert Halls Using Numerous Uncorrelated Sources", Proc. of the Institute of Acoustics, Vol. 28 Pt., 2006

[15] Denon. Anechoic orchestral music recording. Audio CD, Denon Records, 1995. ASIN: B0000034M9.

[16] J. Pätynen, V. Pulkki, T. Lokki. "Anechoic recording system for symphony orchestra", Acta Acustica united with Acustica, vol. 94, nr. 6, pp. 856-865, 2008

[17] M. Pollow, B. Masiero, P. Dietrich, J. Fels, M. Vorländer. "Fast measurement system for spatially continuous individual HRTFs", 25th AES UK Conference: Spatial Audio in Today's 3D World, 2012

[18] B. Masiero, M. Pollow, J. Fels. "Design of a Fast Broadband Individual Head-Related Transfer Function Measurement System", Proc. of Forum Acusticum. Aalborg, Denmark, 2011

[19] G. K. Behler, M. Pollow, M. Vorländer. "Measurements of musical instruments with surrounding spherical arrays", Acoustics 2012, Nantes, France, 2012

[20] M. Pollow, G. K. Behler, M. Vorländer. "Post-processing and center adjustment of measured directivity data of musical instruments", Acoustics 2012, Nantes, France, 2012

[21] V. Pulkki. "Spatial Sound Reproduction with Directional Audio Coding", JAES Volume 55:6, 2007

[22] J. D. Pendleton. "Euler angle geometry, helicity basis vectors, and the Wigner D-function addition theorem", American Journal of Physics, vol. 71, p. 1280, 2003

[23] S. Favrot, J. M. Buchholz. LoRA: A Loudspeaker-Based Room Auralization System, Acta Acustica united with Acustica Vol. 96, 364-375, 2010

[24] S. Pelzer, B. Masiero, M. Vorländer. "3D Reproduction of Room Acoustics using a Hybrid System of Combined Crosstalk Cancellation and Ambisonics Playback", International Conference on Spatial Audio, 2011

[25] D. Schröder. "Physically-based real-time auralization of interactive virutal environments", PhD thesis, RWTH Aachen University, Aachen, Germany, 2011

[26] M. Schlütter. "Parallelisierung von Algorithmen für die Echtzeit-Raumakustiksimulation", Master's thesis, RWTH Aachen University, Aachen, Germany, 2009

[27] B.-I. Dalenbäck. "Real Time Walkthrough Auralization - The First Year", M. Strömberg Valeo Graphics, IOA Copenhagen, May, 2006