Simulation of a sound field reproduction approach for an innovative musical venue (II): Pre-processing of Head-Related Impulse Responses

Antonio CARDENAS

Supervision:
M.A. Alexander Lindau
Prof. Dr. Stefan Weinzierl

Audio Communication Group
Institut für Sprache und Kommunikation
Fakultät I Technische Universität, Berlin, Germany.
November 2010

Foreword

This work is the second of a three parts project being developed at the Audio Communication group of the Technical University of Berlin, in co-operation with ADAM Professional Audio GmbH. The project aims at a realistic acoustic simulation of an innovative musical venue that will have interactively controlled room acoustics. For that, a way to assess subjective impression of different rendering approaches over the future listening area in the projected venue should be provided.

The first part of the project is being done by Max Röhrbein, and aims a perceptually motivated reduction the sound field components for the purpose of plausible real time synthesis, which would be better suited for a virtual acoustic simulation.

This report treats a fraction of the second part of the project, whose aim is to generate some essential requirements for the auralization of sets of virtual loudspeaker arrays in a virtual room. Therefore it is, among other things, mandatory to be able to retrieve head-related impulse responses (HRTFs) for any direction of incidence. Results of this work will thus be used for the simulation and at the auditive evaluation.

The third part of this project, done by Johnny Nahas, aims to synthesize suitable driving functions for given loudspeaker arrays and to auralizations of the different sound field simulations.

Acknowledgments

Back in 2009 I was looking for an opportunity to continue my studies on acoustics in Berlin. After some job applications, I ended up attending to several very interesting seminars with the Audio Communication Group of the TU Berlin, which I couldn't completely understand (not only because of my very bad German proficiencies!). People discussed about topics such as 'binaural technology' and 'auralization', some things about which I had heard something in the past, but which I was still incapable to grasp totally.

As my interest grew, discussions with Alexander Lindau and Prof. Stefan Weinzierl led to an invitation to take part in this project. This great opportunity was endorsed by Prof. Polack in Paris. I feel extremely grateful for all the help I had from the three of them for the success on this task.

Special thanks to Alex, for your the help, counseling, teaching and corrections. Also for your patience and the time you took for all of this. I will always admire your commitment and great work. Really thank you!

Thanks also to Max, for transmitting always the good spirit needed at EN111. Thanks for being such a good friend and helping me out every time I needed!

Thanks also to Frank, Fabian and Mike, who helped me out when my dsp failed...

And a very special thanks to my family (the now officially extended family included!), which has always been there in the good and hard times. You're just the best!

Contents

Introduction

The acoustical simulation of a technically innovative musical venues is important to determine what will be the use of it. For an electro acoustical musical venue, a modeling of the electroacoustic sound field needs to be implemented for the acoustical simulation.

An auralization is usually created using binaural technology, via dynamic binaural synthesis. One way to achieve that is to do a real time convolution of recorded anechoic audio with head related impulse responses (HRIRs) or binaural room impulse responses (BRIRs). These concepts are further developed in Chapter 1.

For the modeling of the electroacoustic sound field, clean and easily accessible HRIRs should be used. This implies a pre-processing of the measured HRIRs. These pre-processing steps are discussed in Chapter 2.

Chapter 1

Auralization

The motivation behind Virtual Acoustics is to allow the delivery of an acoustical message in a virtual reality system from the source to the receiver as it would happen in a real-world situation. That is, to be able to produce the illusion in a listener of being in a ‘virtual’ acoustic environment which is entirely different from that of where the listener is actually located.

Three main concepts in Virtual Acoustics are ‘simulation’, ‘auralization’ and ‘spatial sound reproduction’. ’Auralization’ for the sound attempts to describe the same concept as ’visualization’ for an image. That is, the description of the process of making an image visible (in visual illustration, movie animation or computer graphics) is the visualization. Auralization in acoustic occurs when an audible result is processed from acoustic effects, primary sound signals or means of sound reinforcement. Michael Vorländer defines it as ‘the technique of creating audible sound files from numerical (simulated, measured, or synthesized) data’ [14].

To auralize the sound field in a room an a priori knowledge of the sound sources, the room geometry and the listeners is needed. Modeling of the sound sources is typically related to approaches of how to model their directive frequency response, position, spatial distribution and orientation, among other parameters, to obtain a so-called ‘dry’ source signal. This ‘dry’ source signal is generally defined as a source signal ‘free of reverberation and of any other cues introduced by sound transmission’ [14] and is the input for the auralization system. The room acoustics modeling can be done using an acoustic modeling software (EASE, CATT-ACOUSTICS, ODEON...), where the room geometry, the sound sources and the receivers are defined as mathematical objects. with the modeled sound source and three dimensional space, the reproduction can be done making use of dynamic binaural synthesis and taking into account the modeling of the human hearing (listener).

1.1 Some fundamentals of binaural technology

This is a small introduction to the basic vocabulary and methods used in binaural technology relevant to this work. For a more exhaustive approach, an excellent text on the fundamentals of binaural technology is given by Hendrik Møller [10]. A great guide to understand 3-D Audio Technologies is given by Durand Begault [1] and very good review of the actual state of the arts was given this year by Rozenn Nicol [12].

The binaural recording technique is a method in which two signals are recorded in the ears of a listener or using a copy of an average human head (dummy head). These signals correspond to the sound pressures at each of the eardrums (left and right) and are a complete physical representation of the sound, including timbre and spacial aspects. If they are then reproduced using frequency compensated headphones (to ensure the sound recorded of one ear just gets to that same one ear), it should be able to reproduce the complete auditive experience [10].

To record the binaural signals, small microphones are usually placed at the eardrum, at the entrance of the open ear canal or at the entrance of the blocked ear canal [10]. This should be made considering the fact that sound transmission, from the beginning from the entrance in the ear canal to the eardrum, is independent of the distance and direction of the sound source [2].
The spacial information contained in the two recorded signals consists of a number of cues that a listener uses to determine the direction and distance to a sound source. These localization cues are namely the ITD (Interaural Time Difference), the ILD (Interaural Level Difference) and the SC (Spectral Cues) [12].

Binaural signals can also be synthesized (binaural synthesis) and this can be done convolving a monophonic 'dry' source signal with a pair of binaural filters measured in free field or anechoic conditions. The filters must reproduce the transfer function of the acoustic path between the sound source location and the listener's ears, defined as the sound pressure located somewhere in the ear divided by the sound pressure in the middle of the head with the listener absent [10] [2]. They are referred to as Head Related Transfer

Functions (HRTFs) or as Head Related Impulse Responses (HRIRs) in time domain. An HRTF describes the listener's morphology response to a sound source located at a given point in a free field. It depends on three main parameters: frequency (or time), angle of incidence (azimuth and elevation) and distance to the sound source, and the listener's morphology (each individual has his own HRTF). These have to be distinguished from Binaural Related Impulse Responses (BRIRs) and Binaural Related Transfer Functions (BRTFs) which describe the transmission of sound from a source to a receiver's ear in an enclosure, thus including multiple reflection paths [14].

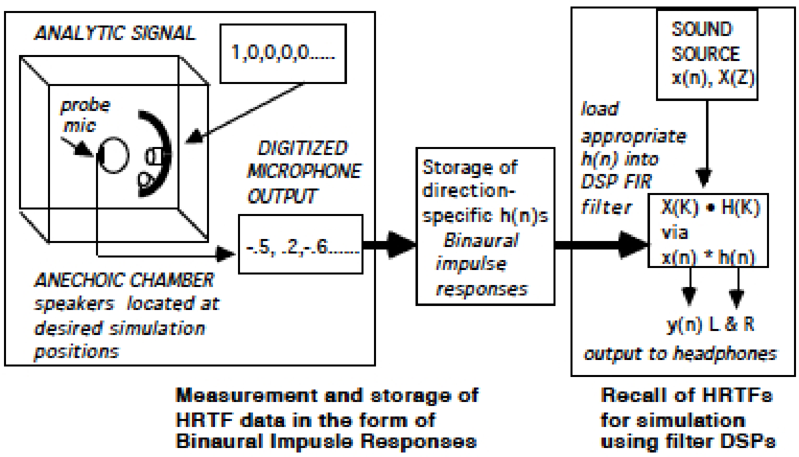

Figure 1.1: Overall plan of the HRTF measurement-storage-simulation technique (from Begault[1]) .

The HRTF acoustic measurement can be a complex task. They need to be measured in the ears of the participant or dummy head using small inserted (or built-in) microphones. An anechoic chamber is ideally required, and a way of moving the sound source or the participant's or dummy's head accurately (in steps of 1 or 2 degrees at least [14]) needs some very precise equipment. The relative location of the subject in relation to the loudspeaker must be carefully tracked, and a full 3D database requires for a participant to stay still for more than one hour or to use a dummy head. Fortunately, free HRTF databases exist and are an attractive solution (for a list of them refer to [12]). An alternative to measure HRTFs can be to model them. Several methods exists (i.e. Finite Element Method, Boundary Element Method), but they all need a very good representation of the listener's morphology and are highly computing consuming.

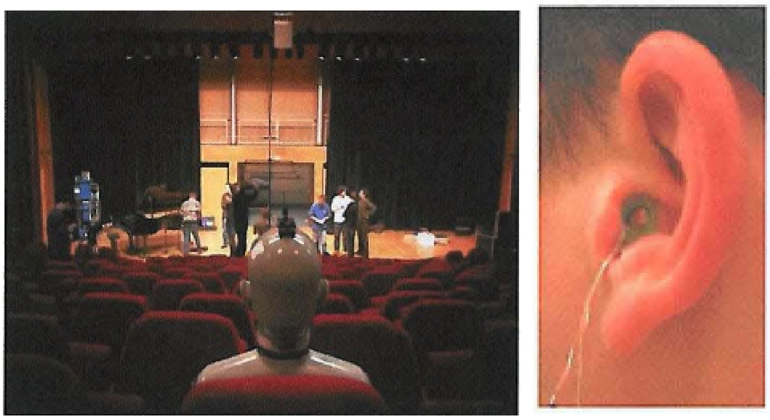

Figure 1.2: HRTF recording: with a dummy head (left) or at the blocked ear canal (right) (from Nicol[12]) .

With a full 3D database of left and right HRTFs (for the left and right ears respectively, and for all directions of incidence in a sufficiently fine angular distribution) a synthetic full 3D spatialization or a Virtual Acoustic Environment can be created via dynamic binaural synthesis. The word 'dynamic' is added to emphasize the fact that the movements of the head are tracked in relation to a reference position. This information is used to adapt in real time the binaural filters and modify the virtual direction of the source in order to keep the source location stable. Likewise, it is possible to select the filters to compensate for changed head orientation and for a different source position thus, allowing to simulate moving sources in dynamic (head-moved) scenario. When no HRTF information for a specific direction has been measured, interpolated HRTFs can be used.

So far, some issues found in binaural synthesis have prevented it to be used in large scale applications. One main issue is that spacial encoding of a sound made by a listener cannot be used for another listener without severe spacial distortion (specially problems with frontal localization [2] [14]). For this, individualization tools of binaural rendering are being developed [12].
In dynamic binaural synthesis several issues can be found, related to the head tracking system and the update of the binaural filters. The head tracking system must be able to provide a fast response and accuracy. Several tracking systems exist (acoustic, optic, magnetic...), and each has its specific advantages and issues (see [12]). A main issue in dynamic binaural synthesis is also due to the low latency needed to provide a natural spatialization, which means that the lapse of time between when the listener moves and

when the binaural filters are updated must be minimal ($\sim$ 75 ms for short stimuli). Also the dynamic update of the binaural filters can lead to audible artifacts due to calculations between sets of coefficients. Cross fading can be use to solve this problem[12].

A proper implementation of dynamic binaural synthesis can lead to a very convincing auralization, taking us a step further into virtual reality.

1.2 Fast and Automatic Binaural response AcquisitioN (FABIAN)

The representation of the acoustical path from sound source to the two ears of the listener in a specific room is called Binaural Room Impulse Response (BRIR), and can be used to simulate natural or artificial soundfields. The measurement of BRIRs is usually made with head and torso simulators (HATS) that are commercially available. These HATS are usually static systems, with limited possibilities to emulate movements of the head, shoulders or torso.

FABIAN (Fast and Automatic Binaural response AcquisitioN) is the first HATS system that allows an automated measurement of BRIRs for a horizontal rotation range of $\pm 75^\circ$ and vertical $\pm 45^\circ$. It was developed by Alexander Lindau at the Audio Communication Group of TU Berlin[7] [5]. Movement is possible using a servo-motorized neck joint and is controlled by a portable PC. This PC can also conduct custom multi-channel impulse response measurements with an implemented application. The BRIRs measurement process is accelerated considerably by successively stimulating the source before FABIAN's reorientation.

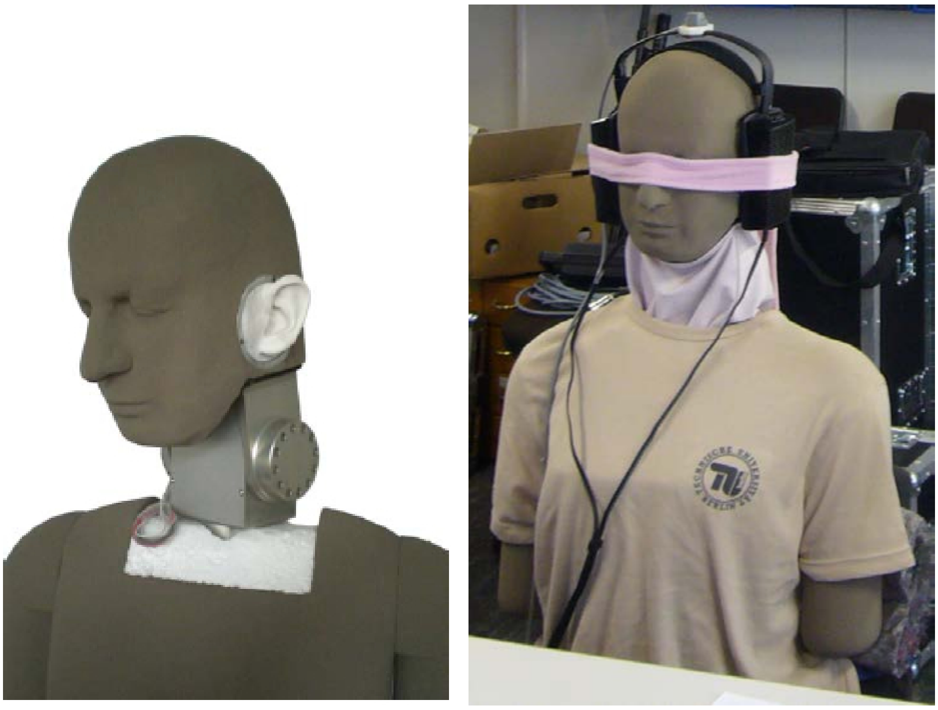

Figure 1.3: Left: Close up of FABIAN, without cover(from Lindau[7]) Right: Head Phone Transfer Function (HPTF) measurement of FABIAN at the TU.

Chapter 2

HRIR Preprocessing

Depending on the method for measuring HRIRs, usually some errors are introduced that have to be corrected by an appropriate preprocessing of the recorded raw data. The raw impulse responses must be modified in time and frequency domain to correct the measurement errors. These errors can originate from the loudspeaker frequency response, unwanted reflections and also from the probe or dummy head microphones response. A post equalization of HRIRs eliminates potentially degradative influences in the measurement and playback chain. Post-processing though should not deteriorate relevant auditory cues as ITD, ILD and SC.

2.1 Moldrzyk HRIRs

There are several HRIR (and/or HRTF) databases publicly available (IRCAM, CIPIC...) [12]. For this project a special dataset acquired with FABIAN was used [9]. This is needed, as FABIAN was used also for measurements of a test room which will be used throughout evaluation in the complete project. Using FABIAN's HRIRs no additional spectral and spatial variation will be introduced when comparing these measurement to future HRIR-based simulations. Christoph Moldrzyk measured a set of HRIRs, using a predecessor of the current FABIAN dummy head that used the same anthropometric base (i.e. Mr. Moldrzyk himself). These HRIRs were measured in the anechoic chamber of the Institute of Technical Acoustics RWTH Aachen, using a rotatable chair. [9] [7]

The horizontal range (azimuth) of the data varies from $0°$ to $359.5°$ (anticlockwise) with a $0.5°$ step resolution. Measurements were also conducted for elevations between $-60°$ to $+90°$, with a $5°$ step resolution. These impulse responses have 336 samples covering approximately 7 ms each. The whole database contains 720(azimuth)x31(elevation)≅ 20000 HRIRs.

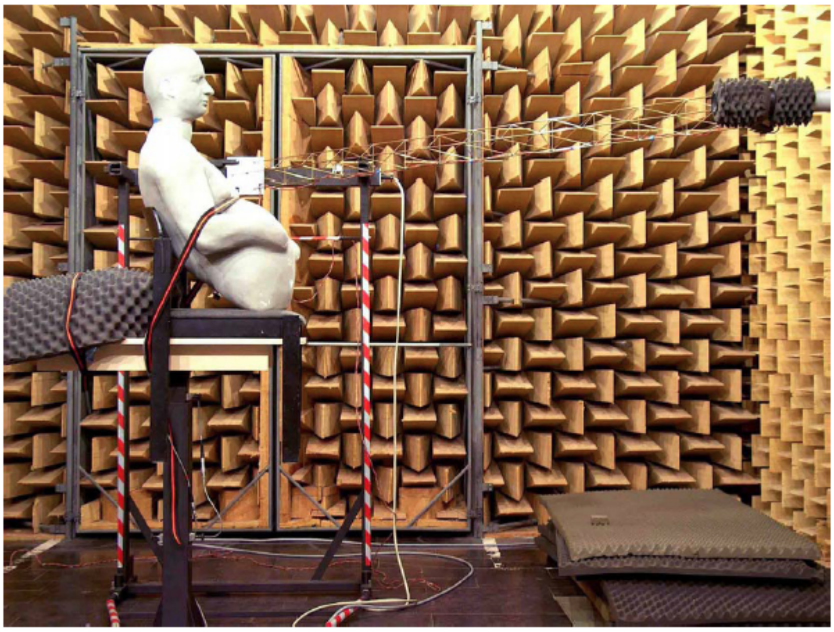

Figure 2.1: Moldrzyk's HRTF measurement setup in ITA Aachen (from Moldrzyk[9]) .

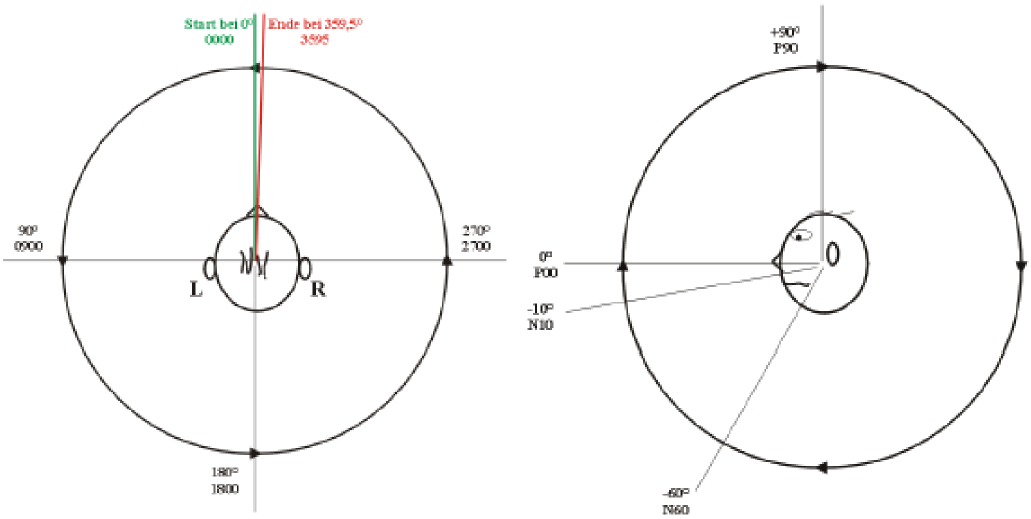

Figure 2.2: HRIRs range and resolution. Left: Horizontal. Right: Vertical. Taken from the technical description of the data made by Moldrzyk.

2.2 Checking HRIRs

Data set was transmitted without description of applied preprocessing. Therefore we first had to make sure whether the given HRIRs, in *.wav files, were equal to the raw data ,in *.spk files (a proprietary data format containing complex spectra of the Monkey Forest measurement software developed at the RWTH Aachen) or if some applied preprocessing had already been done. A parser had to be written in Matlab for reading these files. The parser was written to read the *.wav files following the convention of designation. Additionally some of the original .*spk files had to be read (see 2.2.1). The first step is to determine whether the .wav files were already compensated for the loudspeaker frequency response.

To visualize HRIRs in the frequency domain, a FFT (Fast Fourier Transform) transformation of Matlab was used. To get a first impression of the data we focus on the shape of the frequency response for given direction(s) for both left and right ears.

2.2.1 Loudspeaker compensation

Unwanted spectral distortion is introduced by transducers in the measurement chain (microphones, loudspeakers, interfaces). In our case, microphones are negligible as they were high class measurement microphones (B&K 1/8", microphone pair with high precision alignment). The interface was compensated in the measurement setup within Monkey Forest. Therefore loudspeaker remains as dominant source of spectral deviation. The loudspeaker frequency response might have boots and dips at certain frequencies. With the known loudspeaker frequency response, a compensation can be applied on the measured impulse responses dataset.

Having the raw impulse responses (.spk files), the .wav data and the loudspeaker frequency response, a spectral difference was made for selected HRIRs (chosen arbitrarily) to see if the .wav data is already compensated for the loudspeaker response, and if not, to apply the needed compensation.

Comparing the curves, we found out that the time domain data (*.wav) was already loudspeaker compensated (see figure 2.5).

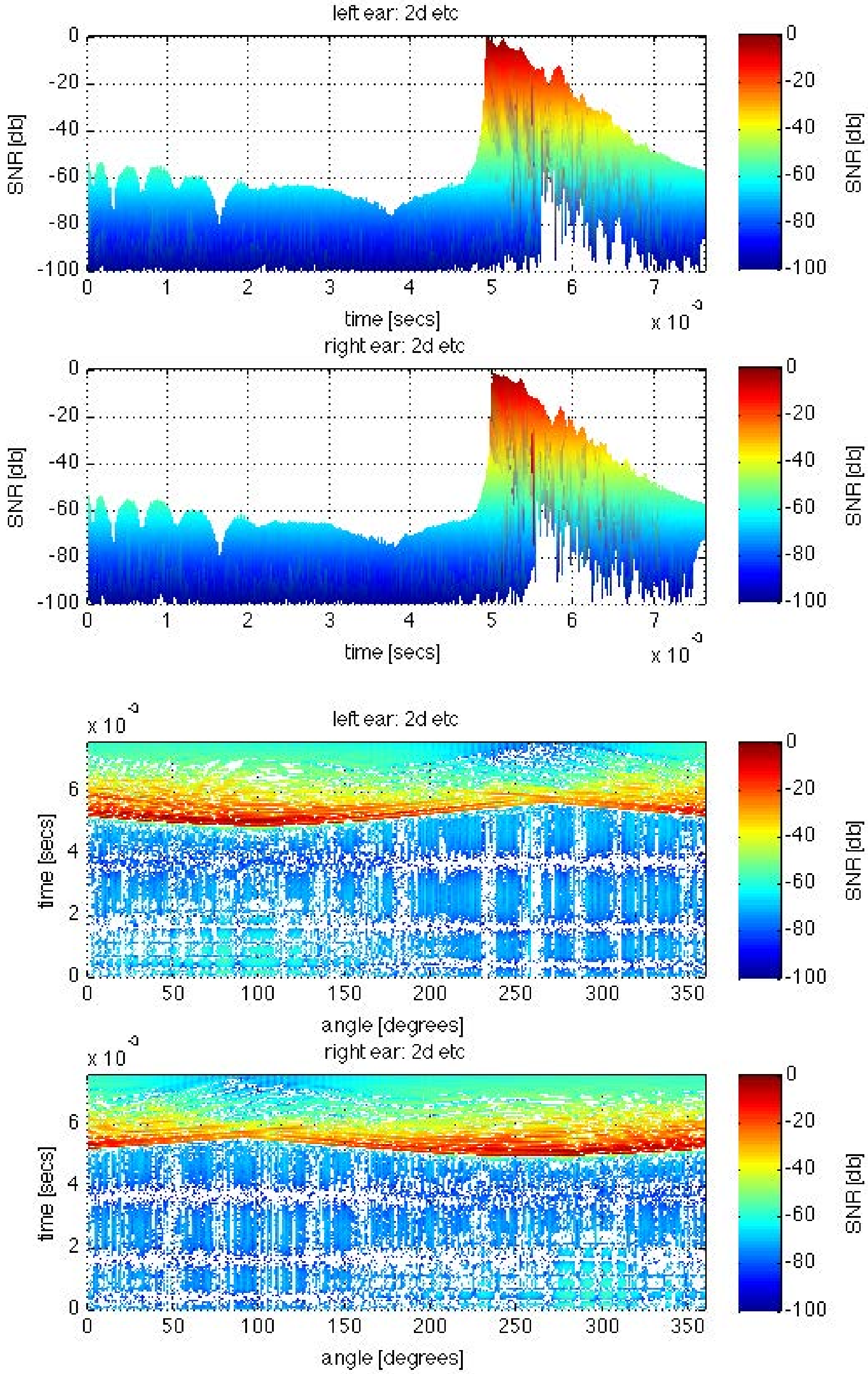

Figure 2.3: Visualizing raw impulses responses in time domain (for elevation 0°), over time in msecs (above) and over angle (below). Notice the large pre-delay and the rather short roll off (only 60 dB SNR).

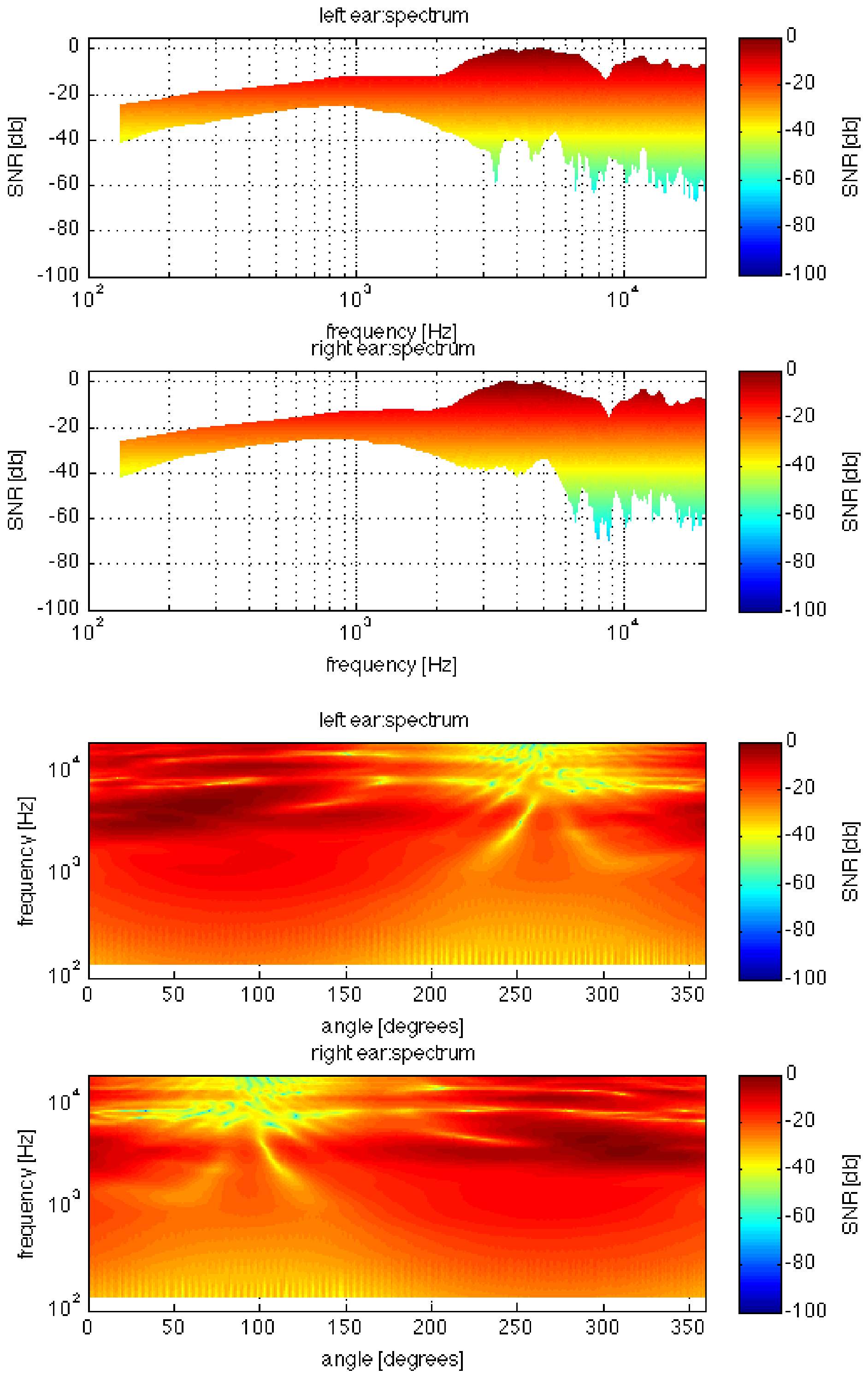

Figure 2.4: Visualizing raw impulses responses in frequency domain (for elevation 0°), over frequency (above) and over angle (below). Notice the roll off below 1kHz. Below 100 Hz no 'plotable' data exists, because it is too small to be interpolated in this configuration. Also notice how SNR has around zero values when the source points direct to the ear (90° for left ear, and 270° for right ear, around 4kHz).

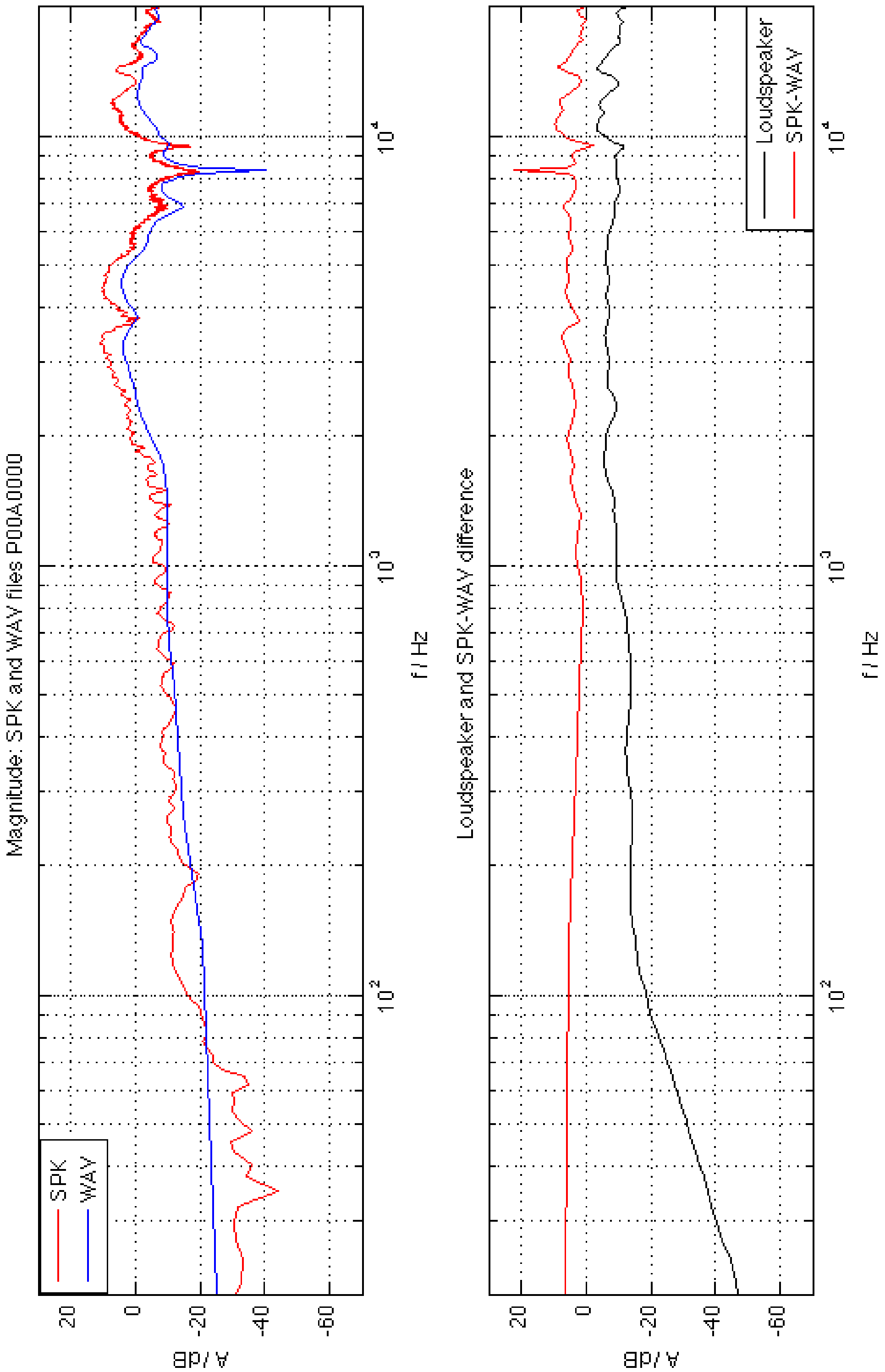

Figure 2.5: Above: raw .wav and .spk spectral data. Below: Magnitude difference of one .spk and .wav data and the loudspeaker frequency response. Both are for the 0°/0° HRIR. Notice the similarity between the magnitude difference and the loudspeaker frequency response.

2.2.2 Reflections

Another important issue is to check for unwanted reflections. When measuring HRIRs in an anechoic room, one must be careful to cover any reflective surface near the source that could provide any kind of reflexion with sound absorbing material. Reflections/Echos are most easily be identified by looking at the HRIR data, where they form distinct peaks following the direct sound, which can then be smoothed. From the delay of the peak one can infer the distance of the obstacle. In the measurement setup in ITA Aachen only a semi-freefield anechoic chamber was in place, thus the setting suffered from floor reflection (see figure 2.1). Even more problematic was that these reflections changed their distance from the direct sound as the loudspeaker boom was moved for measurements. Therefore in the preprocessing done on the *.wav files in Aachen the HRIRs were cut pretty early, at around 100 samples (rather short, compared to [1]). Therefore, Moldrzyk *.wav data had been already appropriately windowed (see in figures 2.3, 2.4).

2.3 HRIR manipulation

Visualizing HRIRs was needed as a prerequisite for actually manipulating them. Is not comfortable, nor efficient, to have a bunch of .wav files that must always be open to handle the impulse responses. The idea would be to have a specific file where all impulse responses are saved and ready to use. The first approach is to use a software like MATLAB to create vectors with all the data and transform them as we want them to be.

2.3.1 Minimum-phase HRIR and pure delay

Separation of ITD and ILD can be a good approach for ITD individualization, thus to compensate for the dummy heads wrong morphology or head size [6]. With this approach we can also get rid of long pre delay in the Moldrzyk *.wavs which reduces dataset size and simplify later interpolation.

HRIRs can be represented as a minimum-phase filter and pure delay [12]. This represents the magnitude spectrum of the HRIR (including ILD and SC) as a minimum phase filter. The ITD portion of the HRIR can be represented as a decimal value computed estimating the HRIR inter-channel delay, and stored as a metadata. This interaural delay corresponds to the difference between the left and right HRIRs start, which is introduced by sound paths from a certain source to both ears having unequal lengths. Time information of the raw HRIR can be reintroduced by adding the delay (depending on sign) at one or the other minimum phase HRIR channel's start.

This approach to model the HRIRs has been validated in several papers, leading to the conclusion that all-pass portion of the ITD is mostly inaudible, and that a pure delay (linear phase part) is sufficient to represent the delay [3] [12].

To calculate the minimum phase HRIR in Matlab, a zero padding is done to double the HRIR lenght. Then the 'rceps' function is used for the minimum phase reconstruction, and it's then cut in half to account for the padding previously done.
For the pure delay component calculation a recently revisited "Inter- Aural Cross-Correlation (IACC)" method, proposed by Nam et al. was used [11]. In this method the cross-correlation function of the HRIRs and their minimum phase components are computed individually per channel. The maximum of this calculation corresponds to the excess-phase delay's component, called 'arrival time'. The ITD is then calculated as the difference between left-ear and right-ear HRIR arrival times. As the sample exact granularity (at 44.1kHz) is larger than just noticeable ITD-difference (10 μsec, [8]), we must be more exact and an upsampling by 10 was in our case applied to obtain more accurate results. The resulting data is the minimum phase HRIRs vectors (left and right ears) and the pure delay also stored as decimals in 64 bit floating point resolution (see figures 2.6, 2.7 and 2.8).

2.3.2 Low frequency compensation

Because of the strong windowing performed in Aachen (see 2.2.2), a low frequency loss was found on all the data. From HRIRs we know, frequency response should be nearly flat below 1kHz (because of pressure receiver, long wavelength, no scattering, no shadowing). Thus to reestablish the theoretical behavior of what we know, an low frequency compensation must be done. Otherwise HRIRs (and thus also auralizations) would sound very thin.

In our HRIRs, a roll-off can be seen for frequencies below 1kHz. Thus a 'bass boost' must be applied to obtain a flat response there. This creates a new, spectrally altered versions of our measured HRIRs. The frequency boost was done using a designed shelve filter applied to each HRIR. The values used for the filter were found calculating an average HRIR and applying the filter to it (see figure 2.9).

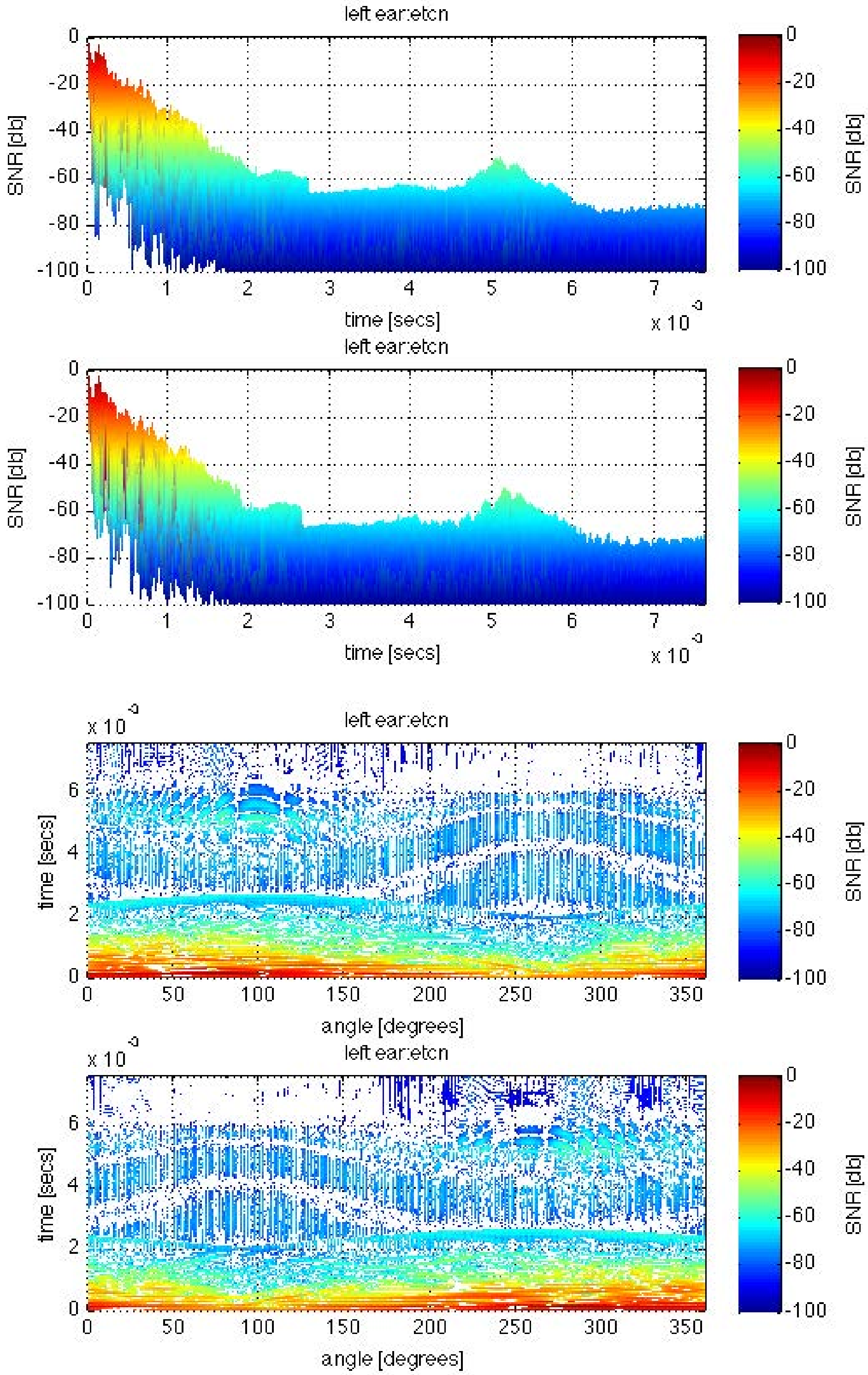

Figure 2.6: Minimum-phase impulses responses in time domain (elevation 0°), over time in msecs (above) and over angle (below). Notice that now the signal starts at 0 and decays, and are all aligned.

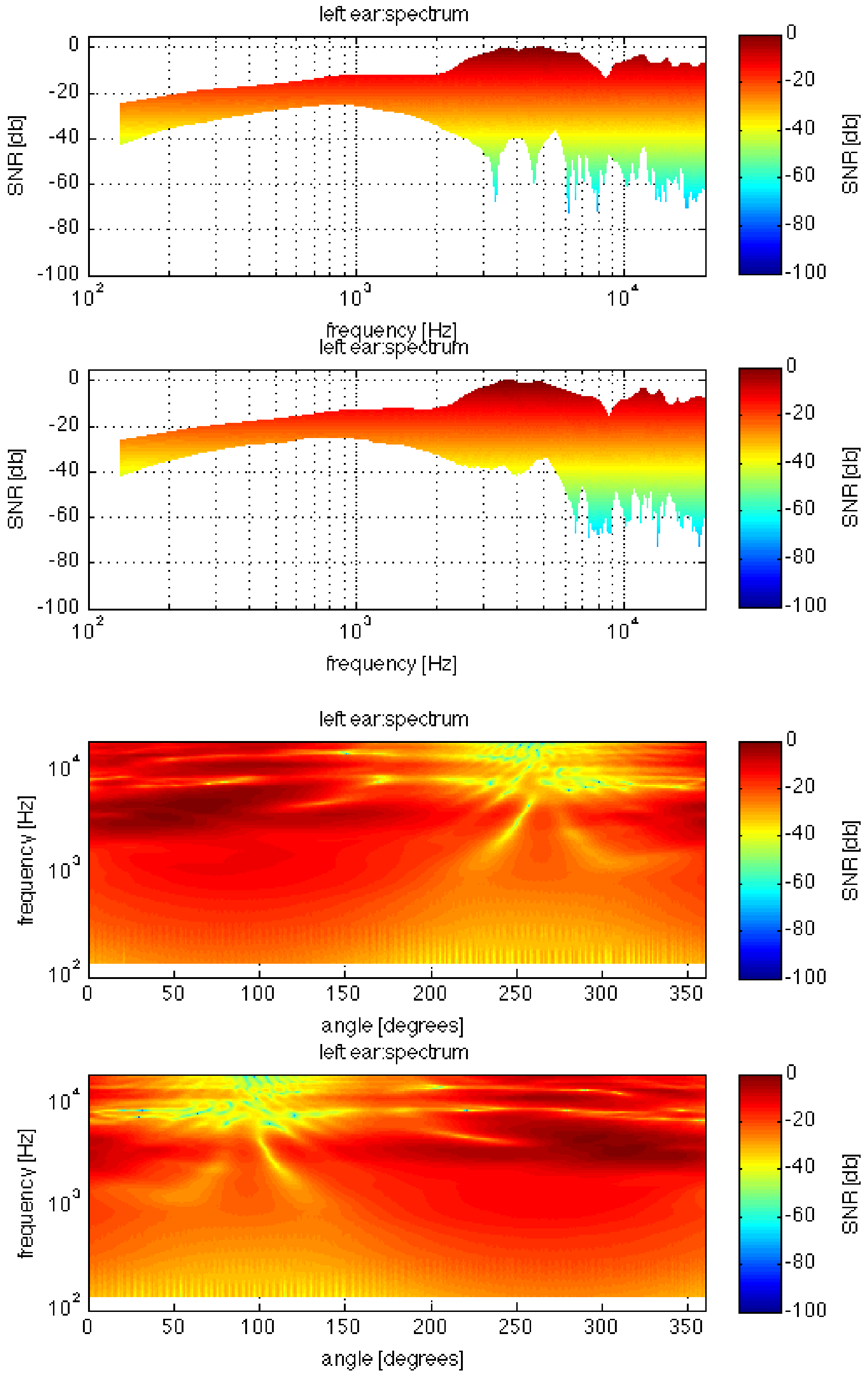

Figure 2.7: Minimum-phase impulses responses in frequency domain (for elevation 0°), over frequency (above) and over angle (below). This shows that the HRIRs spectral information remains unchanged after the minimum phase transformation.

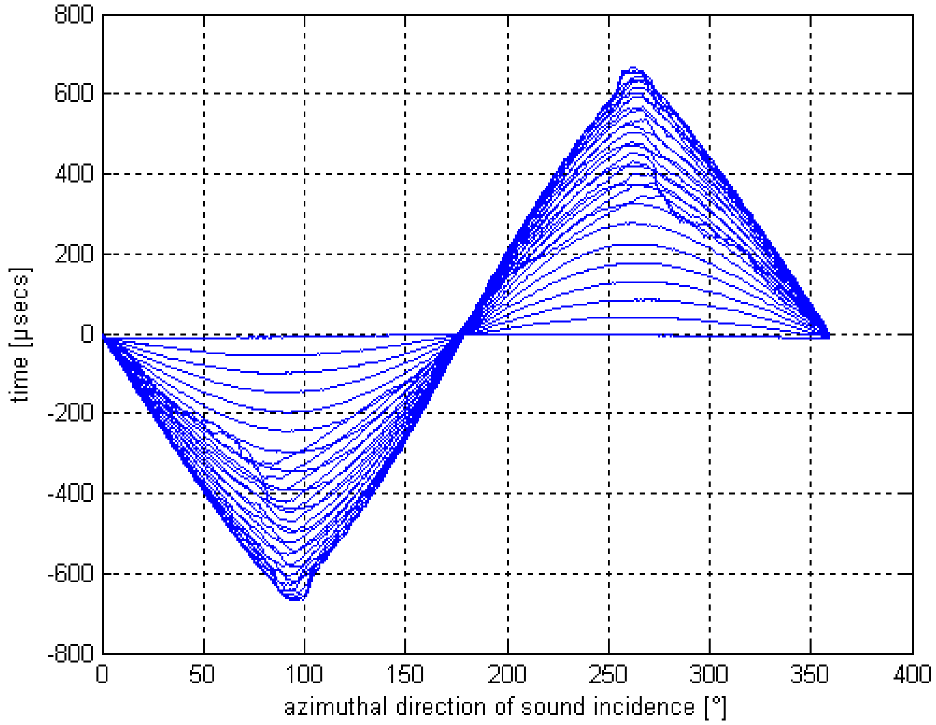

Figure 2.8: ITD in microseconds for all azimuthal directions of sound incidence. Each curve represents a different elevation, being the outer curve for elevation 0° (biggest ITD value). The flat curve is for elevation 90°. Notice the symmetry between the left and right ear ITD values.

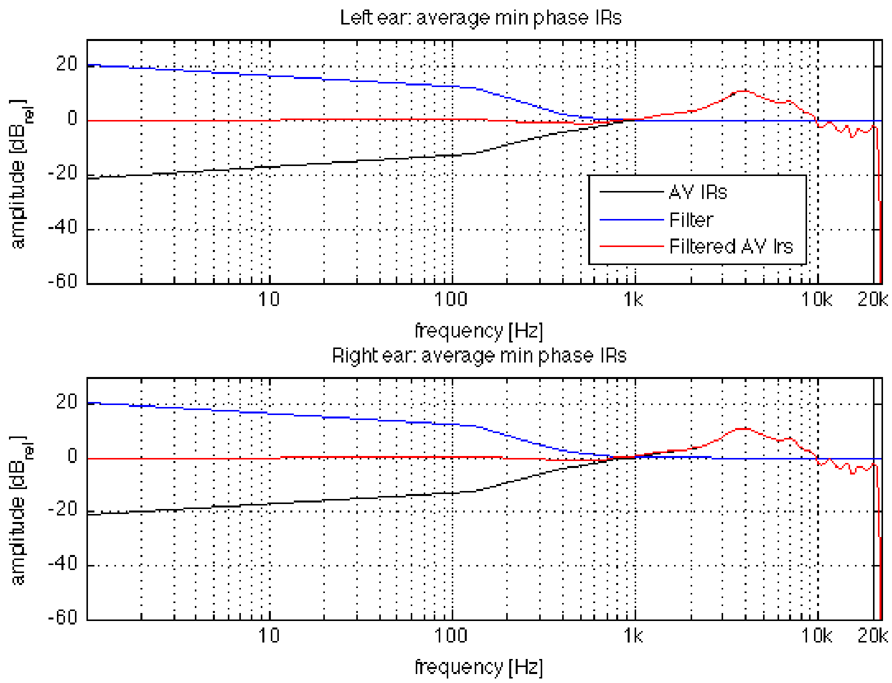

Figure 2.9: Applying filter compensation to the average spectral HRIR value.

2.3.3 Fade out and normalization

A sudden drop-off from the HRIR noise level to that of the playback engines idle noise can be heard after an HRIR has been convolved, for instance, with an impulse. To avoid that, a fade out is done here applying a left-slope squared cosine window on the whole dataset, on the last 150 samples.

The last step is to normalize the whole dataset. This is done dividing all the data by the maximum value found in the HRIR data set. The result from this pre-processed impulse responses can be seen in figures 2.10 and 2.11.

2.4 Re-saving and accessing HRIRs

Our pre-processed HRIRs were so far stored as vectors in a .mat file, with another vector for the pure delay data. This is a format not very handy to work with for somebody who has not worked on them, unless a specific documentation for this data is made. This is one of the main problems for HRIRs databases, each one of them has a specific way of being recorded. It is not only about the filetype used (as .wav files, .dat files, .mat files....), but also (and most problematic) about the orientation of the given data and the applied preprocessing. Some people have them stored in a 360° system, others in ±180° systems... It can be then difficult to use a same programmed routine to compare two different datasets. A new solution for this might be the OpenDAFF project presented early 2010 at the DAGA conference in Berlin [15]. One of the OpenDAFF team's goal is to provide a universal format for saving and accessing directional audio (HRIR/HRTF, etc). The project is still developing and on an alpha version, but it is working for some of the basic functions that we need right now.

Developed at the Institute of Technical Acoustics (ITA) at RWTH University (Aachen, Germany), the Open Directional Audio File Format (OpenDAFF) is an open source format which enables an easy interchange of directional audio data. With this format, different HRIRs set could be shared without needing to know all the measurement details, and used in any personal routine. Its is designed to be optimized for speed and real-time applications, and to be used in any platform for free and open-source.

In this format the data is defined at spherical grid points, and you can access it with in-build routines. They are mostly written in C++, with a Matlab Toolbox that is easily used to create the daff file and manipulate the data. It also comes with two compilable programs, the DAFF Viewer and

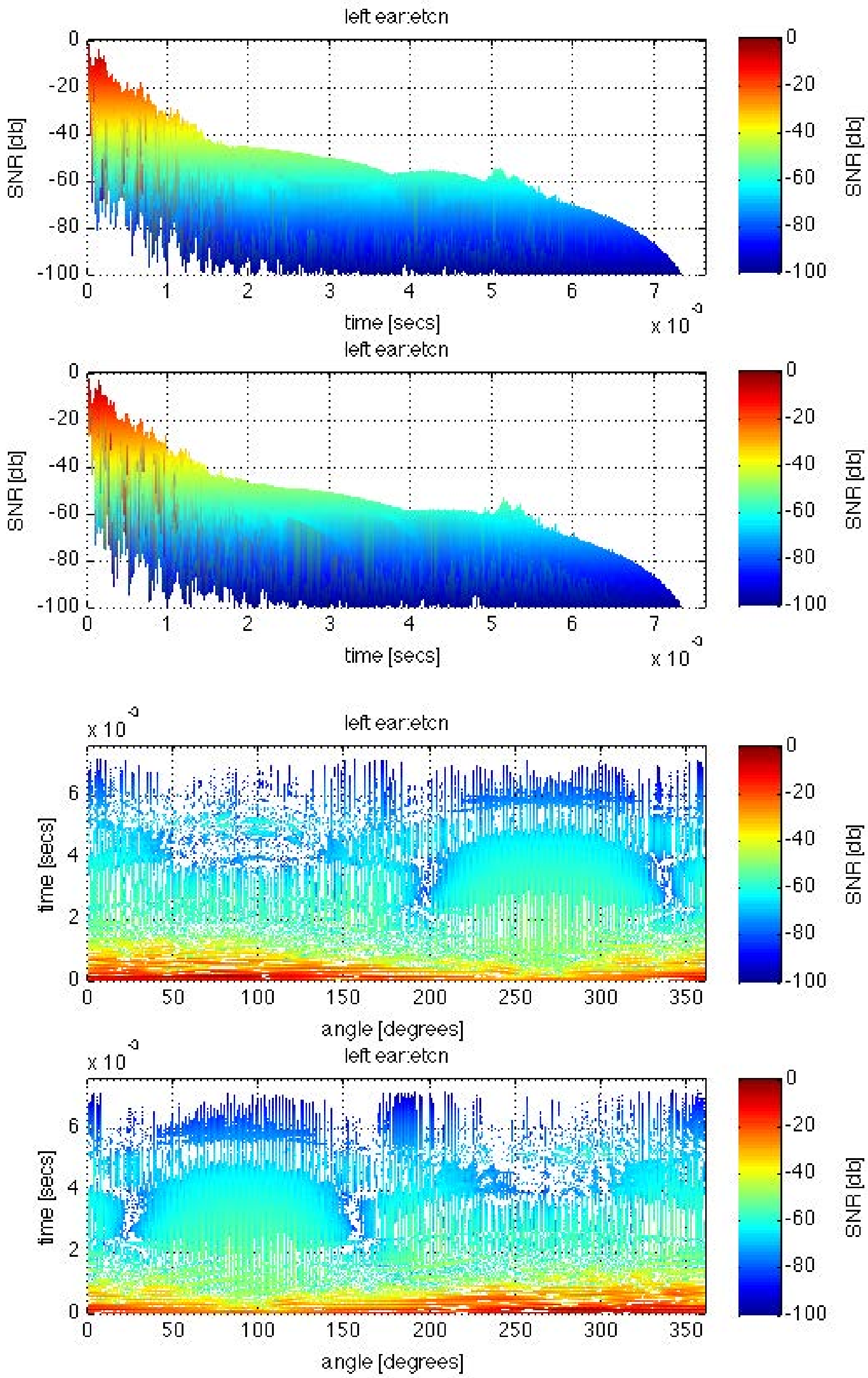

Figure 2.10: Final pre-processed impulses responses in time domain (for elevation 0°), over time in msecs (above) and over angle (below). Notice the smooth roll off after fade out.

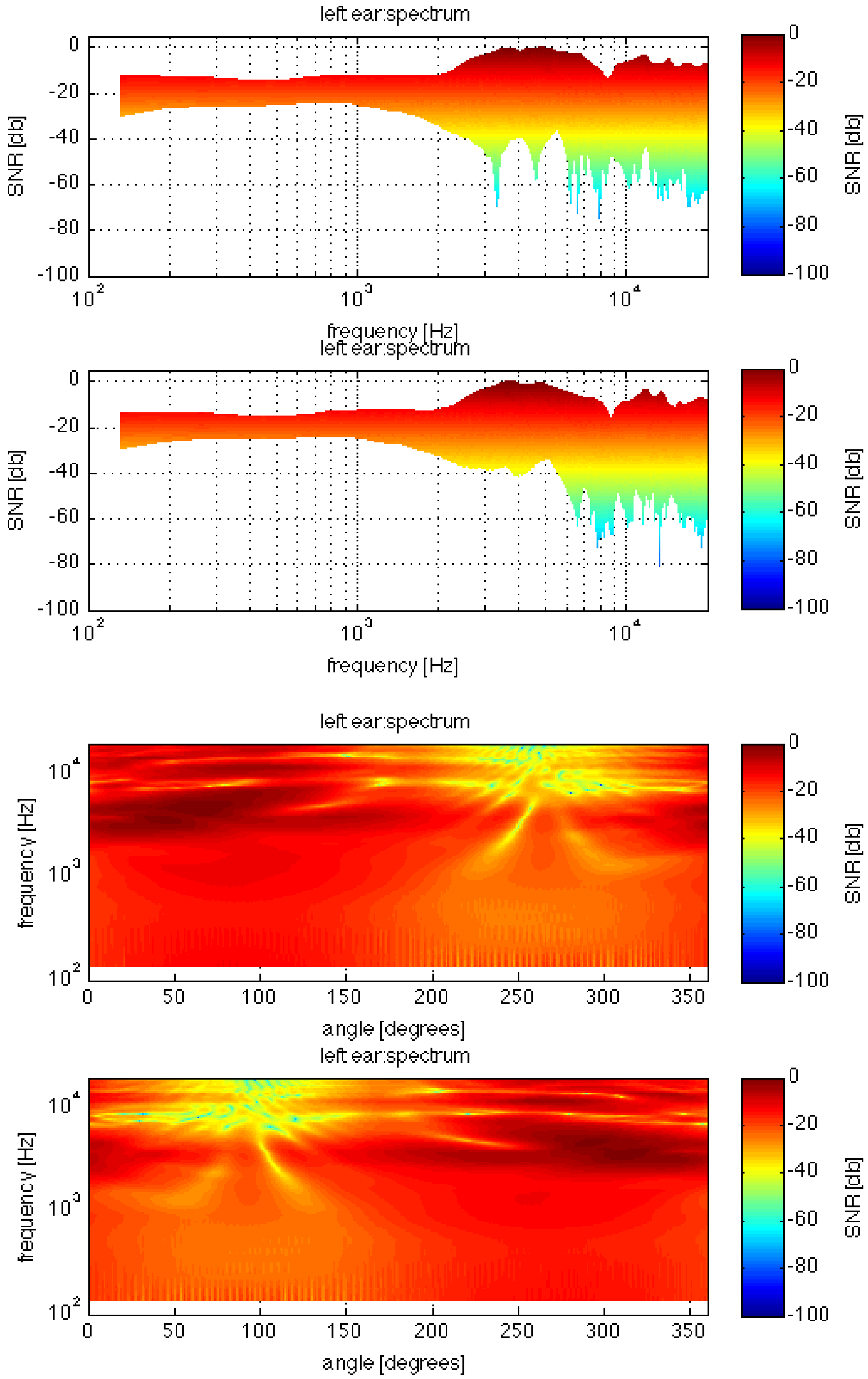

Figure 2.11: Final pre-processed impulses responses in frequency domain (for elevation 0°), over frequency (above) and over angle (below). Notice the flat frequency response below 1kHz, after applying the compensation filtering.

the DAFF Tool to visualize, analyze, inspect and debug the data.

There are two defined coordinate systems: one for writing the daff file, and one for reading it (see figure 2.12). For writing it the data spherical coordinate system (DSC) is used, defining azimuth as counterclockwise 360° alpha angles and elevation as 180° beta angles (with A0° and B0° always in the south pole). For reading the daff file, the object spherical coordinate system (OSC) is used. It facilitates the compatibility with other known coordinate systems that use directional audio (as EASE), defining the azimuthal angle from −180° to 180° and the elevation from −90° to 90° with universal semantic (P0°, T0° is always frontal). Data can be 3-D rotated (yaw, pitch, roll) when the daff file is created.

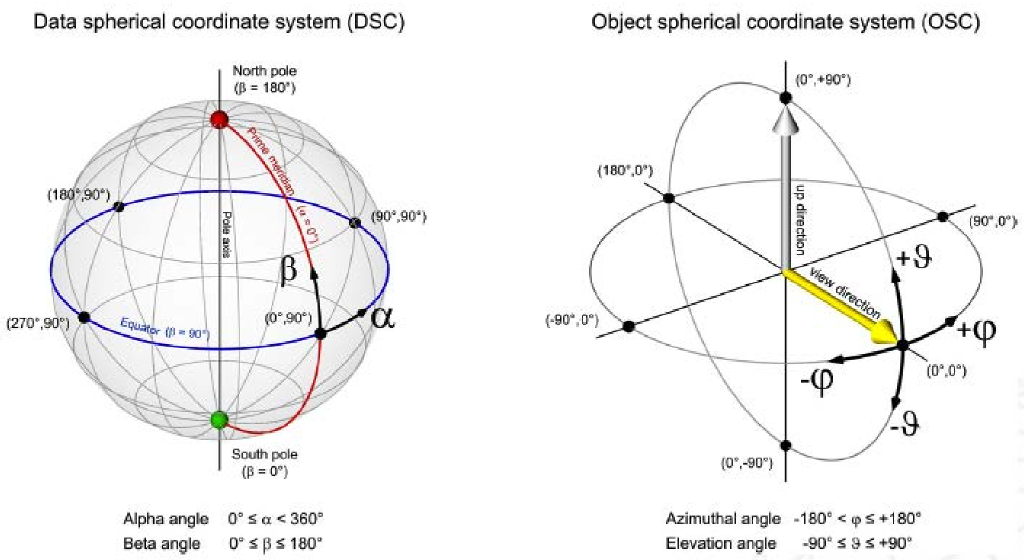

Figure 2.12: OpenDAFF Spherical Coordinate Systems (from [15]).

To write a daff file an intermediary function must be created. This function converts your data to the daff reading format (alpha, beta) and provides the data for each needed channel. It can also provide a metadata per record, if needed. In our case, two channels were used (left and right ear impulse responses vectors) and a metadata for every direction (the ITD value as decimal in floating point resolution). Another metadata information can be given in the file creation to describe your daff file (author, contact, applied processing, etc).

The project is still on an alpha testing phase. This project is one of the first to use it, so many other functionalities are being added and will be ready for its release. For more information about the project refer to the webpage [13].

With our dataset written successfully as a daff file, the access to the data is now made using daff defined functions.

The first requirement for the post-processed impulse responses is was deliver them for any asked direction. As the data has only a 0.5° (azimuth) x 5° (elevation) resolution and is limited to an elevation range of -60 to 90°, a method to deliver impulses responses for any asked direction outside our data range must be implemented.
The OpenDAFF format has a pre-built function that looks for the nearest neighbor data for a given azimuth and elevation angles. The delivered impulse response found is from the nearest point in the equiangular spaced sphere grid that contains any data, calculated using the spherical law of cosines. The function always returns a valid data, even if the angles are outside the given data boundaries, to maintain a consistent behavior.

To compute the impulse response in a direction where the measured HRIR is not available, HRIR interpolation could also be used instead of just giving the nearest available data. It is necessary for dynamic binaural synthesis, where the virtual sound source must be moved so the trajectory can be rendered with enough intermediary directions between start and end points. It also allows to measure and store less HRIRs, because the missing ones can be computed. Several interpolation methods exists, and a good comparison of them is given by Hartung et al [4]. From this paper, an easy method to implement is the inverse-distance weighting method in the time domain. Our minimum phase signals have no initial delay, which is already extracted as a pure delay data, so it makes this method easier to implement as the peaks and zeros are roughly time aligned and do not cancel each other when cross-fading(interpolating). This method makes use of the inverse of great-circle distance to weight the surrounding cell HRIRs and obtain the desired impulse response, and will be developed in the following at the Audio communication Group of the TU Berlin.

Conclusion

Pre-processing of HRIRs is a must to create a realistic auralization. Each set of measured HRIRs is different, and for each of them a special way to pre-process them is needed. The steps taken here are prototypical, and can could be taken also with most other HRIR data sets. The first pre-processing steps should always be the same for any impulse response set (loudspeaker and microphone compensation, normalization), but the way to deal with ITD and spectral magnitude might change. The minimum-phase impulse responses with a separated pure delay component is found to be the best option to save the HRIRs, because it allows an easy manipulation of both, without taking phase into account. Interpolation can then be performed in time domain without needing to align impulse responses. The openDAFF format allows a universal way of saving audio data. Many of its new functionalities have been developed in cooperation with this project, so it is still in its alpha phase. The MATLAB scripts written for this work will be of future use in the pre-processing of FABIAN measured HRIRs.

Bibliography

[1] D. Begault. 3-d sound for virtual reality and multimedia. Technical report, NASA, 2000.

[2] J. Blauert. *The Psychophysics of Human Sound Localization*. MIT Press, 1996.

[3] Z. A. Constan and W. M. Hartmann. On the detection of dispersion in the head-related transfer function. *JASA*, 2003.

[4] K. Hartung, J. Braasch, and S. J. Sterbing. Comparison of different methods for the interpolation of head-related transfer functions. *AES 16th International Conference*, 1999.

[5] A. Lindau. Ein instrument zur softwaregestützten messung binauraler raumimpulsantworten in mehreren freiheitsgraden. Master's thesis, TU Berlin, 2006.

[6] A. Lindau, J. Estrella, and S. Weinzierl. Individualization of dynamic binaural synthesis by real time manipulation of the itd. *AES 128th Convention*, 2010.

[7] A. Lindau and S. Weinzierl. Fabian - an instrument for software-based measurement of binaural room impulse responses in multiple degrees of freedom. *VDT International Convention*, 2006.

[8] A. W. Mills. On the minimum audible angle. *JASA*, 1958.

[9] C. Moldrzyk, W. Ahnert, S. Feistel, T. Lentz, and S. Weinzierl. Head-tracked auralization of acoustical simulation. *AES 117th Convention*, October 2004.

[10] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 1992.

[11] J. Nam, J. S. Abel, and J. O. Smith. A method for estimating interaural time difference for binaural synthesis. *AES 125th Convention*, October 2008.

[12] R. Nicol. Binaural technology. *AES*, 2010.

[13] OpenDAFF. http://www.opendaff.org/.

[14] M. Vorländer. *Auralization*. Springer, 2008.

[15] F. Wefers. Opendaff. *DAGA*, 2010.