Measurement of Head Related Impulse Responses for Psychoacoustic Research

Florian Völk, Martin Straubinger, Luis Roalter, Hugo Fastl

AG Technische Akustik, MMK, TU München, Germany

Email: florian.voelk@mytum.de

Introduction

In recent years, the importance of Head Related Impulse Responses (HRIRs) for psychoacoustic experiments and especially for studies in directional hearing has been growing, mainly caused by the availability of fast computers and the resulting possibility of very realistic dynamic binaural synthesis (cf. Møller [1], Völk et al. [2]). A necessary prerequisite for all these methods is the correct measurement of HRIRs.

In this contribution a setup will be presented which is capable of measuring HRIRs for psychoacoustic research by two different techniques: Maximum Length Sequences (MLS, invented by Schroeder [3]) and Exponential Sine Sweep (ESS, early work by Craven et al. [4] and Griesinger [5]). To reduce the measurement time, an interpolation algorithm can be used. The resulting HRIRs are compared by typical physical parameters as well as judged in a localization experiment. Based on these data, parameters for high quality HRIR measurements for psychoacoustic research are proposed.

Measurement System

Figure 1 shows the setup used for all impulse response measurements presented in this paper.

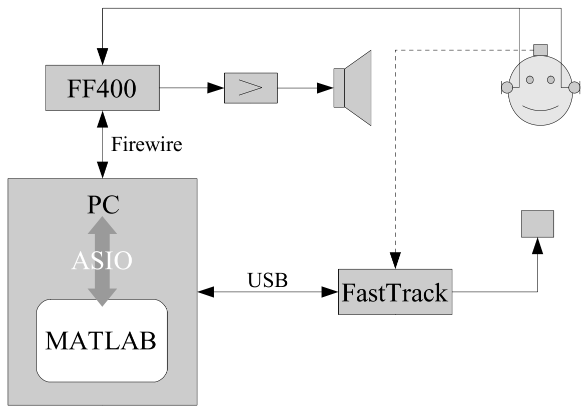

Figure 1: Setup used for measurements of Head Related Impulse Responses. Miniature microphones inserted in the blocked ear canals of a subject as well as a loudspeaker are connected via an audio interface (FF400) to the PC. In addition, the position and orientation of the head is recorded with a tracking device (FastTrack).

The digital signal processing is done at double precision word length (64 bit floating point) and 44.1 kHz sample rate on a Microsoft Windows XP based PC using the software package MATLAB by The MathWorks. In- and output of audio data is accomplished by a custom made MATLAB plug-in using the direct software-driver-communication specified by Steinberg in the ASIO (Audio Stream Input/Output) architecture. For the digital to analog and analog to digital conversion at 24 bit word length and 44.1 kHz sample rate, an RME Fireface 400 (FF400) is employed. Figure 2 shows the impulse response; Figure 3 the magnitude of the transfer function and the group delay of the used specimen, measured in the case of short circuit from in- to output (at 44.1 kHz sample rate).

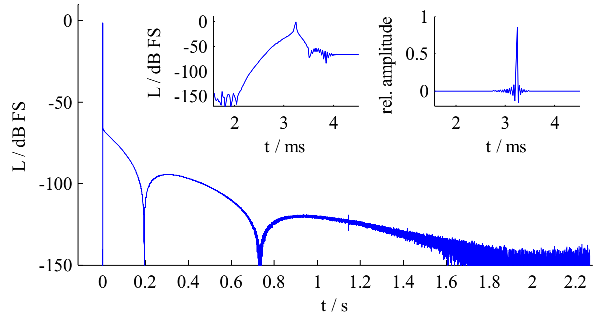

Figure 2: Impulse response of the used audio interface, measured in the case of short circuit from in- to output.

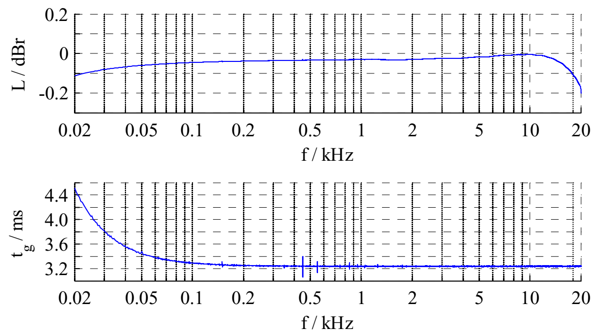

Figure 3: Magnitude of the transfer function and group delay of the used audio interface, measured in the case of short circuit from in- to output (at 44.1 kHz sample rate).

Following additional specifications are given by the manufacturer: Channel separation more than 110 dB, THD+N smaller than -96 dB, SNR on input side 105 dB, dynamic range on output side 110 dB (cf. [6]). These technical data justify the application of the FF400 for our purposes. The loudspeaker and the microphones are in general arbitrary and can therefore be selected for each specific use case. The measurements discussed in the present paper were carried out using miniature microphones Sennheiser KE-4-211-2 (electret condenser microphones, diameter 4.75 mm, height 4.2 mm sensitivity 10 mV/Pa, SNR 58 dB¸ frequency response within ±1 dB between 100 Hz and 20 kHz, cf. [7]), connected to the preamplifiers of the FF400 and inserted in the blocked ear canals of a male subject. The excitation signals were played back by a Klein & Hummel O200 active two way studio monitor loudspeaker box (frequency response within ±2 dB between 60 Hz and 20 kHz). Figure 4 shows a right frontal HRIR, measured in a laboratory environment with the discussed setup. Figure 5 shows the corresponding transfer function's magnitude and group delay.

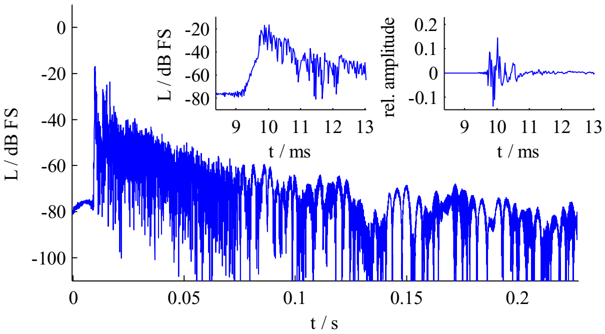

Figure 4: Right frontal HRIR captured with the proposed measurement setup in a laboratory environment in the blocked ear canal of a 28 year old male subject.

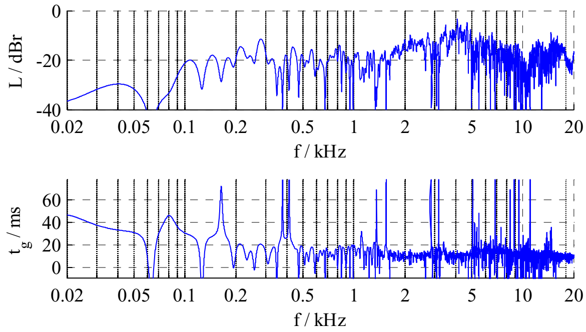

Figure 5: Magnitude and group delay of the head related transfer function, measured in a laboratory environment in the blocked ear canal of a 28 year old male subject.

Measurement Methods

All impulse response measurement techniques strive for one goal: the determination of the response from a linear and time invariant system to impulsive excitation. As the generation of a sufficiently short although powerful impulse is difficult especially if the measurement is conducted with a digital source, a number of different measurement procedures were developed through the years, each having advantages and disadvantages. To select a proper method for HRIR measurements, we will define the specific requirements for that use case. It will not be distinguished explicitly between recordings under anechoic or reflective conditions. The resulting impulse responses will be called HRIRs in any case.

Requirements

HRIR sets for data based binaural synthesis are usually measured in a resolution adapted to the just noticeable differences in spatial location of sound sources. According to Blauert [8], the maximum sensitivity occurs in the frontal region with horizontal head rotations and lies in the order of one degree. Therefore, a typical synthesis resolution for the azimuth in the horizontal plane (for the definition cf. Blauert [8]) is one degree. Even if the synthesis is restricted to the horizontal plane, the measurement of 360 pairs of impulse responses is necessary. If we assume a common duration for the recording of one impulse response pair of about five seconds, and the time for repositioning of the head and/or the loudspeaker of at least 10 seconds, the measurement of the considered data set would last more than one hour. There are different ways to shorten this time, for example measuring more loudspeakers sequentially or more than one loudspeaker at a time (cf. Majdak [9] or Xiang [10]), but the duration of a single measurement remains nevertheless a crucial factor. So, a method to acquire HRIRs should get along with excitation signals and recording times as short as possible. In addition, the achievable signal to noise ratio (SNR) plays a major role since it limits the dynamic range of the binaural synthesis.

Possible Solutions

In the following, two possible measurement procedures will be discussed in relation to our purpose: measurement with Exponential Sine Sweeps (ESS, cf. Farina [11] and [12]) or Maximum Length Sequences (MLS, cf. Rife and Vanderkooy [13]). Müller and Massarani [14] and Stan et al. [15] presented a comprehensive and detailed comparison of the mentioned (and some other) measurement procedures. Therefore we will discuss only the factors that are important for the specific application presented in this paper. Both methods have in common that the room answer is captured while the excitation signal is played back. Afterwards, the impulse response is computed from these signals. The duration of the excitation signal and the recording time has to be selected in any case at least as long as the impulse response to be captured (cf. Müller and Massarani [14]). For measurements with ESS, the only further restriction is the extension of the recording time beyond the ESS duration so that the room answer has decayed to the noise floor at all frequencies before the recording is stopped. This means usually only a little overhead on the ESS duration, because the sweep starts at low frequencies, where commonly the highest reverberation times occur. In contrast, with MLS measurements, the duration of the excitation signal needs in general not to exceed the impulse response duration, but directly influences the achievable SNR. The longer the excitation signal lasts, the more energy is fed into the room and therefore, due to the decorrelation used as post processing in the MLS case to derive the impulse response, the higher the SNR becomes.

For measurements in small and medium sized halls, with our procedure, ESS duration of five seconds has proven useful. The MLS with a similar duration is of order n = 18, resulting in a duration of about 6 seconds at 44.1 kHz sampling rate. Figure 6 shows the SNR of impulse responses measured with ESS (red circles) and MLS (blue asterisks) at different sound pressure levels (SPL) using the described setup.

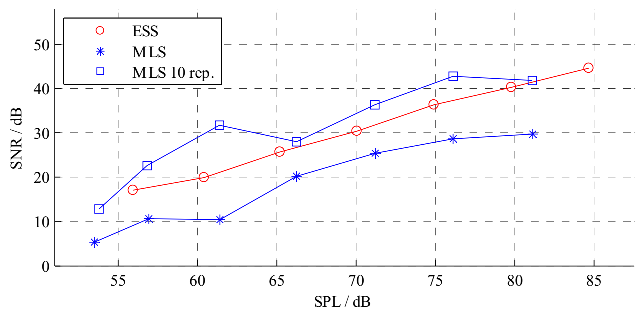

Figure 6: SNR of impulse responses measured at different sound pressure levels (SPL). Red circles indicate ESS as measurement method, blue asterisks MLS. Blue squares show the results for the mean over 10 MLS repetitions.

It is evident that for excitation signals of about the same length and SPL, the ESS method yields higher SNRs than the MLS method. For the latter it is possible to increase the achievable SNR by averaging over multiple recordings of the room response to the MLS before the impulse response extraction. The blue squares in Figure 6 represent the SNR of the impulse responses derived from the average over ten repetitively captured responses of the room to the MLS. Because of the averaging, the SNR increases to or at some excitation levels (depending on the background noise) above the values reached with ESS but at the cost of extending the signal to one minute. Another important difference between the ESS and MLS methods is the treatment of nonlinearities in the measurement chain. Impulse responses measured with MLS are unavoidably distorted by occurring nonlinear effects. If a special deconvolution procedure is used with ESS measurements, it is possible to eliminate nonlinear effects by a simple windowing procedure in the time domain (cf. Farina [11]). It should be mentioned that there might occur other influence factors like background noise in or time variance of the measurement environment that should be considered in selecting a measurement procedure. In the case discussed here, they are negligible.

Used Procedure

As shown in the previous paragraphs, the ESS method requires significantly less measurement time than the MLS procedure while providing, especially at high playback levels around 85 dB SPL, an SNR of the resulting impulse responses in the order of the SNR provided by the used microphones. Therefore the measurements presented in the following were conducted using ESS excitation. For measurement of whole HRIR-sets, a tracking-system (Polhemus FastTrack) was used to control and record the position of the head during the measurement.

A further reduction of measurement time could be achieved if the impulse responses are measured at less spatial resolution than required for the synthesis. The full resolution could be reached afterwards by spatial interpolation between measured impulse responses. From the variety of proposed interpolation procedures an interpolation in the time domain using cubical splines after time shift of the impulse responses was checked for the just noticeable azimuthal sample point density in the horizontal plane. For that purpose, 12 subjects rated in a paired comparison if there is a difference between the interpolated and the originally measured HRIR pair corresponding to the same spatial position if they are convolved with broadband noise pulses. It can be concluded that at a sample point density of five degree, less than 20 % of the subjects are able to differentiate between the original and interpolated data set, regardless if the recording was carried out using an artificial or a real head.

Data

With the presented setup, an HRIR set covering the azimuths in the horizontal plane was measured with a 28 year old male subject at 5° spatial resolution and afterwards interpolated to 1° resolution. These and some other data are available on the internet (http://www.mmk.ei.tum.de/~vol/hrtfs).

Physical Analysis

Figure 7 shows the frequency dependent interaural level differences (ILDs) for different source directions (180° to 360°) computed from the first three milliseconds of the HRIRs measured in the blocked ear canals. The images differ in the used frequency analysis method and the scaling of the ordinate. Figure 7 a) and b) are computed using the Fast Fourier Transformation (FFT) whereupon the ordinate is scaled linearly in a) and logarithmically in b). Figure 7 c) is computed as mean of the Fourier-t-Transformation (FTT, cf. Terhardt [16]) and displayed on an ordinate divided according to Critical Bands (cf. Fastl and Zwicker [17]), but labeled with the corresponding physical frequencies.

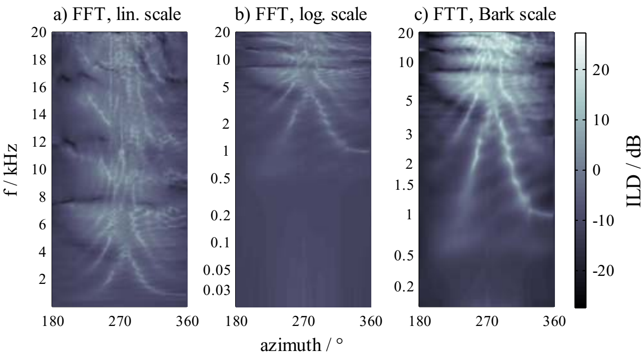

Figure 7: Interaural Level Differences (ILDs) of a HRIR set acquired in a laboratory environment.

Figure 8 is structured in the same way as Figure 7 but shows the Interaural Time Differences (ITDs), computed as differences in group delays.

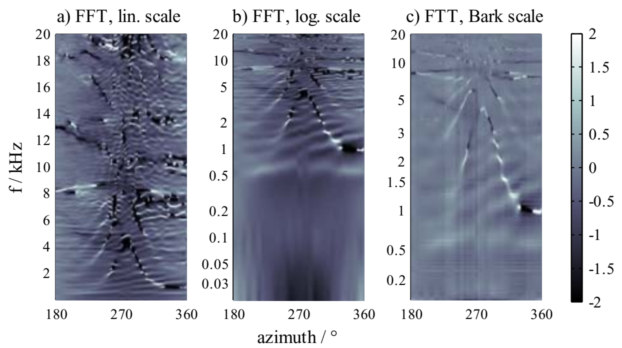

Figure 8: Interaural Time Differences (ITDs, computed as group delay differences) of the considered HRIR set.

Psychoacoustical Analysis

With the measured HRIR set, a hearing experiment was conducted to determine the resulting auditory event positions. A virtual sound source created by dynamic binaural synthesis using the discussed HRIR-set was placed at 18 different azimuth angles in the horizontal plane. The system was operated at 44.1 kHz; all involved (digital) signals were originally captured or synthesized at this sample rate. All filters were realized as FIR filters. The impulse responses were recorded in the reflective laboratory environment where the trial took place and shortened to a filter-length of 8192 sample-points. Only impulse responses measured in the horizontal plane were used in the binaural synthesis. For that reason, vertical head movements and rotations had no impact on the rendered ear signals. The resulting data were filtered with a linear phase FIR filter to equalize the undesired distortion of the headphones, microphones, amplifiers, and other playback as well as recording equipment. All calculations in the audio-processing chain were performed at a block-size of 512 samples, which made partitioning of the audio signal as well as of the impulse responses necessary. This was realized by an adaptive overlap and save procedure. As acoustic stimulus, a German male speech signal recorded in an anechoic chamber was used. To exclude a possible influence of any visual stimuli besides darkness (which may also have an influence on auditory perception), the experiments were conducted under very dark conditions and the listeners did not see the listening room at any instant of time (to ensure this, they were blindfolded before entering and leaving the room). Stimuli were presented to each person in an individual random order from the different spatial source locations; each stimulus occurred three times. Subjects could enter the direction and distance at which they perceived the auditory events via a graphical interface. They were sitting on a chair in the middle of a laboratory and could move their heads freely but were not allowed to stand up or walk around. The whole trial was automated using a software program running on a PC that was connected to a touch-screen which served as input device. The subjects could give their answers by pencil-click directly on the screen. For this purpose, they saw first a top down view and afterwards a side view of the shape of a head. They had to mark up the position corresponding to the auditory event's location on the completely black sketch (except for the grey head shape) in the respective representation and had the possibility to correct a given answer as many times as they wanted. 15 subjects (2f, 13m) participated in the experiment. All involved subjects were naive regarding the presentation system, four of them had no experience with listening experiments at all, and the others were familiar with listening in virtual auditory displays and with localization experiments. There were no significant differences between the results of these groups. Four subjects (1f, 3m) were excluded from further analysis because of in-consistent judgments for identical stimuli. So the results of eleven subjects (1f, 10m, mean age 25.3 years, mean duration 18.7 minutes) were used for computation of quartiles. Figure 9 shows the deviation between the denoted auditory event azimuths and the intended azimuths.

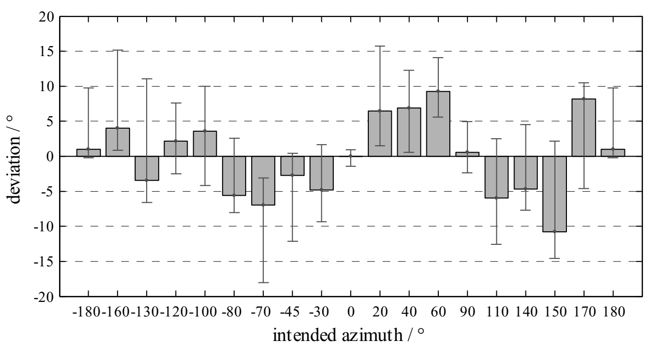

Figure 9: Deviations between denoted auditory event azimuths and intended azimuths for dynamic binaural synthesis using the presented HRIR set.

Figure 10 depicts the denoted heights and distances, normalized to the respective maximum value. These psychoacoustical measurements justify in our opinion the proposed measurement parameters as the results are in line with known values for binaural synthesis (e.g. Begault [18]).

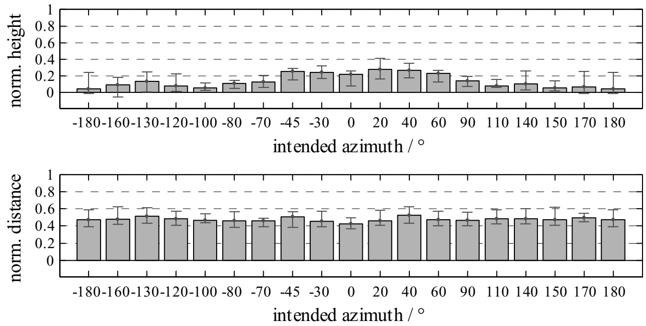

Figure 10: Denoted relative height and distance of auditory events created by dynamic binaural synthesis.

Acknowledgements

Part of this work was supported by grant FA 140/4 of the Deutsche Forschungsgemeinschaft (DFG). The authors gratefully acknowledge the support of Sennheiser electronic GmbH & Co. KG, who provided the miniature microphones.

References

[1] Møller, H.: Fundamentals of Binaural Technology. Appl. Acoustics **36**, 171-218 (1992)

[2] Völk, F.; Kerber, S.; Fastl, H.; Reifinger, S.: Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor. *Fortschritte der Akustik, DAGA '07*, DEGA e. V., Berlin (2007)

[3] Schroeder, M. R.: Integrated-impulse method measuring sound decay without using impulses. J. Acoust. Soc. Am. **66**, 497-500 (1979)

[4] Craven, P. G..; Gerzon, M. A.: Practical Adaptive Room And Loudspeaker Equaliser for Hi-Fi Use. *92nd AES Conv.* (1992)

[5] Griesinger, D.: Beyond MLS - Occupied Hall Measurement with FFT Techniques. *101st AES Conv.* (1996)

[6] Carstens, M.: Fireface 400 User's Guide. Version 1.6, RME Intelligent Audio Solutions (2008)

[7] Industry Information: Electret Microphone Capsule KE 4. Sennheiser electronic KG (1986)

[8] Blauert, J.: *Spatial Hearing. The Psychophysics of Human Sound Localization*. The MIT Press, Cambridge, Massachusetts, London (1997)

[9] Majdak, P.; Balazs, P.; Laback, B.: Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. J. Audio Eng. Soc. **55**, 623-637 (2007)

[10] Xiang, N.; Li, S.: A pseudo-inverse algorithm for simultaneous measurements using multiple acoustical sources (L). J. Acoust. Soc. Am. **121**, 1299-1302 (2007)

[11] Farina, A.: Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. *108th AES Conv.* (2000)

[12] Farina, A.: Advancements in impulse response measurements by sine sweeps. *122nd AES Conv.* (2007)

[13] Rife, D. D.; Vanderkooy, J.: Transfer-Function Measurement with Maximum-Length Sequences. J. Audio Eng. Soc. **37**, 419-444 (1989)

[14] Müller, S.; Massarani, P.: Transfer-Function Measurement with Sweeps, J. Audio Eng. Soc. **49**, 443-471 (2001)

[15] Stan, G.-B.; Embrechts, J.-J.; Archambeau, D.: Comparison of Different Impulse Response Measurement Techniques. J. Audio Eng. Soc. **50**, 249-262 (2002)

[16] Terhardt, E.: Fourier transformation of time signals: Conceptual revision. ACUSTICA 57, 242-256 (1985)

[17] Fastl, H.; Zwicker, E.: *Psychoacoustics – Facts and Models*. 3rd ed., Springer, Berlin, Heidelberg (2007)

[18] Begault, D. R.: Perceptual effects of synthetic reverberation on three-dimensional audio systems, J. Audio Eng. Soc. **40**, 895-904 (1992)