Technische Universität Berlin
Fachgebiet Audiokommunikation und –technologie

Simulation and Analysis of fast individual head-related transfer function measurement techniques
(Proposal)

Prof. Dr. Stefan Weinzierl
Fabian Brinkmann M.A.

Mina Fallahi (335211)
April 2013

Abstract
The need for the acquisition of individual head-related transfer functions with high spatial resolution requires techniques to reduce the measurement duration. This thesis studies two of these techniques and simulates the measurements to analyse the different parameters influencing the efficiency of the methods. The goal is to specify the optimal parameters. Within another thesis a real HRTF measuring setup will be constructed for which the results of the present thesis could be used.

Introduction and motivation
The head-related transfer functions (HRTFs) as the basis of the binaural technology describe the free-field sound propagation between the sound source and the listener's ears. According to Møller [Møl92, Møl95] this propagation can be divided in to two parts, a direction dependent part and a part independent of the direction. The direction dependent part is defined as the ratio between the pressure at the entrance to the (blocked) ear canal to the one measured at the centre position of the head with listener absent and includes the spatial cues which the listener uses to perceive a sound's direction. Despite the possibility of interpolating between measured HRTFs, still a large number of source positions have to be measured for a correct representation of the binaural space [Lin08, Min05]. In most practical applications the time consuming measurement of HRTFs is carried out using recordings of one listener or, more often, of an artificial head. But since HRTFs also include the highly individual filtering information due to reflection and diffraction of the listener's head, torso and pinna, listening to the recordings acquired from another listener or an artificial head will cause coloration and localization errors [Møl96, Wen93]. Therefore for applications with high demands on fidelity individual HRTFs gain more importance. There have been attempts to acquire personalized HRTFs using anthropometric measurements of the pinna and approximating head and torso by models [Zot03], simulations with ray tracing methods [Röb06] or computation using multipole accelerated boundary element method [Gum10], but the most accurate are real measurements with individual subjects although this is not without its own restrictions: during the long measurement time subjects must keep still to avoid the artefacts caused by head movements. Thus it is of interest to decrease the measurement duration as far as possible to make it comfortable for subjects and at the same time

reduce the effects of time varying elements in the measurement system such as temperature change or unavoidable head movements.

In this thesis two methods for fast measurement of HRTFs [Enz09, Maj07] are studied and simulated. As described in these methods speeding up the measurement of HRTFs is limited by required accuracy and technical effort. The goal is to acquire optimized parameter for each method regarding measuring time, accuracy and signal to noise ratio (SNR).

State of Art

The continuous azimuth measurement of HRTFs using Normalized Least Mean Square (NLMS) algorithm and adaptive filtering introduced by [Enz09] is based on the idea of simultaneous measurement with loudspeakers arranged at discrete elevations. The listener is placed at the centre of the loudspeaker array and is continuously rotated above the vertical axis. During a single rotation of 360° all loudspeakers produce the excitation signal simultaneously and the recordings in the left and right ear together with the known excitation signal lead to the complete set of the HRIRs as a function of continuous azimuth angle. Simultaneous system identification is achieved using multichannel adaptive filtering [Ben01] and the NLMS algorithm [Hay02]. The advantage of continuous azimuth measurement is that there will be no need of interpolations between discrete measurements. Fukudome et al [Fuk07] also had the idea of continuous azimuth measurement using Maximum Length Sequences. According to the fact that the maximum convergence speed of the NLMS algorithm is achieved by perfect sequences these should be chosen to overcome the problem of the time variant acoustic transmission path. Sweeps have shown to come off well as excitation signals having high energy and reducing the limitations due to nonlinear distortions [Mül01]. Telle et al [Tel10] introduced the perfect sweep as excitation signal and Antweiler et al [Ant12] show the rapid tracking ability of NLMS algorithm when using perfect sweeps in comparison to white noise.

Majdak et al [Maj07] introduced the multiple exponential sweep method (MESM) for shortening the measurement time by simultaneous measurement with loudspeakers which are arranged in a vertical arc. Since the measurement equipment consisting of loudspeakers or power amplifiers introduces nonlinear distortions the measurement system can be described as weakly nonlinear, which will result in higher order harmonics in the impulse response. By exciting the system with logarithmic sweeps linear and non linear parts of the response can be separated. MESM takes advantage of this property and interleaves logarithmic sweeps, by starting each sweep one after another so that all impulse responses are located between the linear and the first harmonic impulse response of the last sweep. Dividing the existing channels into groups which are interleaved one can furthermore overlap these groups by starting the next group before the last sweep of the present group has reached its end. In order to maintain the desired SNR the number of channels to be interleaved should be optimized. Weinzierl et al [Wei09] also introduced a more generalized multiple sweep measurement with sweeps which are spectrally adapted to the noise floor to improve the SNR. Masiero et al [Mas11] described a setup for a rapid measurement of HRTFs also using interleaved sweeps as excitation signal with subject standing on a turntable inside a vertical arc of up to 40 loudspeakers. In contrast to Majdak et al [Maj07] who used discrete azimuth measurements Masiero et al [Mas11] discuss the electro-acoustical and mechanical aspects to be considered for a correct quasi continuous acquisition. Dietrich et al [Die13] also introduced the optimized MESM, which goes further than overlapping and interleaving and takes advantage of temporal structure of the linear impulse response to place the single harmonics among arbitrary fundamentals to yield even shorter measurement times.

Methods

The path to be simulated consists of loudspeakers and amplifiers, sound propagation path through air and the HRTF as recorded by the blocked ear canal microphones. To have an estimation of the recorded signals at microphones the HRTF dataset of the audio communication group (ground proof data) can be used. For both methods which are to be simulated [Enz09, Maj07], sweeps are generated [Mül01] as excitation signal. Using the ground proof data and the excitation signal HRTF measurements can be simulated and HRTFs obtained from these simulations can then be compared to the ground proof dataset. To verify the implementation of the measurement algorithms simulations can be run for the ideal transfer path, with only one channel system, a fixed azimuth angle, without noise, non-harmonic distortion and air sound attenuation. Second, the transfer path as shown in Fig. 1 can be modelled and lastly, a simulation of continuous azimuth measurement can be included.

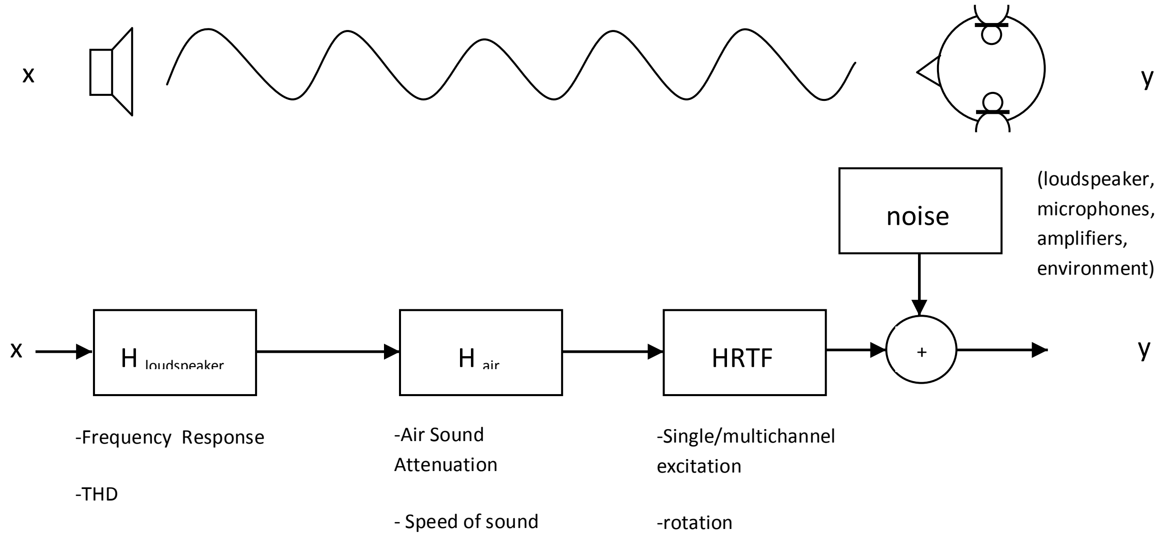

Figure 1. Transfer path model

In Fig.1 x represents the excitation signal. Using Volterra series [Nov10] harmonic distortions can be supplied to a typical loudspeaker transfer function modelled by a second order butterworth high-pass. H_{air} represents air attenuations with negligible spectral coloration. Speed of sound is to be modelled as fractional delay. The continuous rotation can be modelled by time-variant convolution using interpolated HRTFs from the ground proof data set. The noise which will be added to the recorded signal has to be modelled with spectral adaptation to the noise which is present in the measurement room. For sweep generation and calculation of simulated HRTFs the preliminary works of audio communication group can be used. The goal of the current work is to optimize the measurement time of both methods [Enz09, Maj07] while maintaining a desired SNR and HRTF length. The minimum HRTF length is given by the time of the arrival of the latest considerable reflection (shoulder or legs) and the decay of the modes in the pinna and will be determined by exemplary measurements. The dynamic range of HRTF during this time gives the minimum required SNR.

For the method introduced by Enzner [Enz09] (NLMS adaptive filtering method) T_{360} (duration of a complete rotation) and μ (step size) are two key parameters which play an important role in the behaviour of the NLMS adaptive filtering method. Faster rotations require larger μ, since μ decides for the convergence behaviour of the NLMS algorithm [Hay02]. But larger μ also means less accuracy. Therefore an optimized μ should be found to have the best T_{360} with minimum error. [Enz08] introduces calculable and measurable quantities for the HRIR inaccuracy as functions of μ and T_{360}. Extending the single channel system to multichannel, for a perfect simultaneous excitation of all N channels Antweiler et al [Ant08] suggest an N-time extended excitation signal to be supplied to the first channel with phase shifted versions for other channels. Again the optimization parameters should be acquired since the number of channels also influences the accuracy.

For the method suggested by Majdak et al and Masiero et al [Maj07, Mas11] (Multiple Exponential Sweep Method - MESM) one of the most important questions is the number of channels (η) to be interleaved. Since the interleaving occurs in the time between the linear impulse response and the first higher order harmonic response the sweep duration should be adapted according to the length of the linear impulse response (L_1), the length of second harmonic response (L_2) and the number of channels to be interleaved (η). Furthermore overlapping N/η groups of interleaved channels requires the number of all existing higher order harmonics (K) in the impulse response. For the method introduced by Dietrich et al [Die13] (Optimized MESM) there may also be the need to have a measurement of the length of all higher order impulse responses. It should also be noticed that the MESM algorithm will not always lead to shorter measurement time. The efficiency of the method with respect to the measurement time can be compared to the normal single excitation method according to a time compression or acceleration factor as a function of channel number, L_1, L_2 and K [Wei09].

Both methods should be compared to each other and to the ground proof dataset with respect to SNR, measurement duration, technical effort and accuracy. The most important parameters to be determined by simulations are the rotation speed (T_{360}), the number of channels, the number of interleaved channels (η) and the exact excitation time for each sweep signal.

Bibliography

[Ant08] C. Antweiler, A. Telle and P. Vary (2008): "NLMS system type identification of MISO systems with shifted perfect sequences". In: *Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, Washington, USA, *September 2008*

[Ant12] C. Antweiler, A. Telle, P. Vary and G. Enzner (2012): "Perfect sweep NLMS for time variant acoustic system identification". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *Kyoto, Japan, March 2012,* pp. 517-520

[Ben01] J. Benesty, T. Gänsler, D.R. Morgan, M.M.Sondhi and S.L. Gay (2001): *Advances in Network and Acoustic Echo Cancellation*, Springer

[Die13] P. Dietrich, B.Masiero and M. Vorländer (2013): "The optimization of the Multiple Exponential Sweep Method". In: *J. Audio Eng. Soc.*, source in print

[Enz08] G. Enzner (2008): "Analysis and Optimal control of LMS type adaptive filtering for continuous azimuth acquisition of Head Related impulse responses". In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, *Las Vegas, Nevada, USA, March-April 2008*, pp. 393-396

[Enz09] G. Enzner (2009): "3D continuous-azimuth acquisition of Head Related Impulse Responses using Multi-channel Adaptive Filtering". In: *IEEE Workshop on Applications of Signal Porcessing to Audio and Acoustics*, *New Paltz NY, October 2009*

[Fuk07] K. Fukudome, T. Suetsugu, T. Ueshin, R. Idegami and K. Takeya (2007): "The fast measurement of head related impulse response for all azimuthal directions using the continuous measurement method with a servo swivelled chair". In: *Applied Acoustics*, *Vol.68*, pp.864-884

[Gum10] N.A. Gumerov, A.E. O'Donovan, A. Duraiswami and D.N. Zotkin (2010): "Computation of head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation". In: *J. Acoust. Soc. Am, Vol.127, No.1,* pp. 370-386

[Hay02] S.Haykin (2002): *Adaptive Filtering Theory*, 4th ed. Upper Saddle River NJ.: Prentice-Hall

[Lin08] A. Lindau, H.-J. Maempel and S. Weinzierl (2008): "Minimum BRIR grid resolution for dynamic binaural synthesis". In: *Acoustics 08 Paris*, *2008*, pp.3851-3856

[Maj07] P. Majdak, P. Balazs and B. Laback (2007): "Multiple Exponential Sweep Method for fast measurement of Head Related Transfer Functions". In: *J. Audio Eng. Soc. Vol 55, No. 7/8*, pp.623-637

[Mas11] B. Masiero, M. Pollow and J. Fels (2011): "Design of fast broadband individual Head Related Transfer function measurement system". In: *Forum Acusticum*, *Aalborg, Denmark 2011*, pp. 2197-2202

[Min05] P. Minnaar, J. Plogsties and F. Christensen (2005): "Directional resolution of Head Related Transfer Functions required in binaural synthesis". In: *J. Audio Eng. Soc. Vol. 53, No. 10*, pp.919-929

[Møl92] H. Møller (1992): "Fundamentals of Binaural Technology". In: *Applied Acoustics*, *Vol.36,* pp. 171-218

[Møl95] H. Møller, M.F. Sørensen, D. Hammershøi and C. B. Jensen (1995): "Head Related Transfer Functions of Human Subjects". In: *J. Audio Eng. Soc. Vol.43, No.5,* pp.300-321

[Møl96] H. Møller, M.F. Sørensen, C. B. Jensen and C. B. Jensen (1996): "Binaural technique: Do we need individual recordings?". In: *J. Audio Eng. Soc. Vol.44, No.6,* pp. 451-469

[Mül01] S. Müller, P. Massarani (2001): "Transfer Function measurement with sweeps". In: *J. Audio Eng. Soc. Vol.49,* pp. 443-479

[Nov10] A. Novak, L.Simon (2010): "Nonlinear System Identification Using Exponential Swept-Sine Signal". In: *IEEE Transactions on instrumentation and measurement, Vol. 59, No.8,* pp. 2220-2229

[Röb06] N. Röber, A. Andres and M. Masuch (2006): *HRTF Simulations through acoustic ray tracing,* Department of Simulation and Graphics, Otto-von-Guericke University Magdeburg, Germany

[Tel10] A.Telle, C. Antweiler and P. Vary (2010): „Der Perfekte Sweep – Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme". In: *DAGA 2010*, *Berlin*, pp. 341-342

[Wen93] E. M. Wenzel, M. Arruda, D.J.Kistler and F.L.Wightman (1993): " Localization using nonindividual Head Related Transfer Functions". In: *J. Acoust. Soc. Am*. *Vol.94, No. 1,* pp. 111-123

[Wei09] S. Weinzierl, A. Giese and A. Lindau (2009): "Generalized multiple sweep measurement". In: *126th AES Convention*, *May 7-10, Munich, Germany*

[Zot03] D.N.Zotkin, J. Hwang, R. Duraiswami and L.S. Davis (2003): "HRTF personalization using anthropometric measurements*"*. In: *IEEE Workshop on Applications of Signal Porcessing to Audio and Acoustics, New Paltz NY, October 2003*, pp.157-160