FAST MEASUREMENT SYSTEM FOR SPATIALLY CONTINUOUS INDIVIDUAL HRTFS

B. MASIERO, M. POLLOW, P. DIETRICH, J. FELS AND M. VORLÄNDER

Institute of Technical Acoustics, RWTH Aachen University, Germany

mpo@akustik.rwth-aachen.de

The head-related transfer functions (HRTFs) play a major role for the auralization of virtual sources around the listener as well as for crosstalk cancellation systems. Generic HRTFs of artificial heads are often used, as measuring individuals using a high spatial resolution is in general a tedious and time-consuming task. A fast measurement system for HRTFs is presented, consisting of a circular arc where up to 40 broadband loudspeakers can be distributed at will. By rotating the subject horizontally in discrete steps HRTFs can be acquired on a spherical surface. Using an optimized version of the multiple exponential sweep technique, thousands of discrete points can be measured within a few minutes making the use of individual HRTFs well feasible in practice. This measurement data can be used to obtain a spatially continuous representation of the HRTFs by using a reciprocal formulation as modal components of an outgoing spherical wave. This results in a setup independent and compact description of individual HRTFs, allowing the evaluation of any binaural transfer functions at any point in near-field or far-field.

INTRODUCTION

The head-related transfer function (HRTF) describes the individual perception of sound for different angles of incidence of a sound wave. The directional cues contained in the HRTF are introduced by reflection and diffraction of the sound wave at head and torso and are highly individual [1]. When hearing a binaural signal synthesized via a generic HRTF – deviating from the listener's HRTF – localization is degraded, especially for the height impression as these deviations are is strongly related to the individual geometry of the pinna. As lateral impression relies on binaural cues, hearing with a generic HRTF will produce an overshooting or undershooting of the localization.

Individually measured or individualized HRTFs play an important role in the overall quality of acoustic virtual reality systems, together with other aspects such as adequate tracking of head motion and orientation and acoustic environment simulation [2]. To obtain a dataset of be used in such systems, various HRTF measurement systems have been constructed following with various design approaches leading to shorter or longer total measurement time. These systems can be classified in three different categories according to the number of required sound:

- Dense array: have as many loudspeakers as directions to be measured.
- Hybrid array: have a group of loudspeakers placed on an arc, where either the arc of the subject is rotated.
- Sparse array: have only one loudspeaker moved to each direction to be measured.

A measurement method based on the principle of reciprocity was proposed by ZOTKIN ET AL. [3]. A miniature sound source and 32 microphones distributed in a spherical array of 0.7 m radius are used. This is a very fast method, as the excitation signal has to be played only once for each ear. The drawback is however that the miniature source delivers a considerably small signal-to-noise ratio (SNR) and restricts the measurement frequency range to frequencies above approx. 1 kHz. ZOTKIN used a model-based extension of the HRTF for frequencies below this limit, losing the individual character of the HRTF measurement. If a high spatial resolution is desired, many microphones positions have to be measured, increasing hardware costs.

Direct measurement systems need two miniature microphones, placed ideally in the entrance of the blocked ear canal [4], and at least one loudspeaker. Just as with the reciprocal measurement, if high spatial resolution is desired, a dense set-up will require a large number of hardware channels, dramatically increasing the costs of the system.

A sparse measurement system will always be the slowest of all methods as no parallelization in the measurement procedure is possible. For example the system built at TNO in the Netherlands took 2.5 h to measure 976 directions [5]. Subjects wore a head tracking device to verify that the head position was kept constant throughout the complete measurement duration.

The hybrid array is a trade-off between speed and hardware complexity and is therefore most commonly found, e.g. the renowned system at the CIPIC Labs that measured the HRTF at 1250 directions in approximately 1.5 h [6]. The arc's size also varies considerably; e.g. the CIPIC's system has an arc with 1 m radius while the system build in Aalborg University has a radius of 1.95 m [1], allowing the subjects to be measured in standing position. Until recently, all multichannel direct HRTF measurement systems used pseudo-random sequences, such as Golay codes, as excitation signals to allow measurement parallelization and thus achieve a considerable speed-up. However, current state-of-the-art acoustic measurements use sine sweeps or chirps for excitation [7,8]. MAJDAK ET AL. proposed the interleaved sweep technique to speed-up HRTF measurements with sine sweeps in their 22 loudspeaker hybrid set-up [9] allowing the measurement of circa 1500 directions in less than 30 min. This approach is more robust to artifacts due to non-linearity in the loudspeakers and is therefore preferable.

Recently, ENZNER proposed a new technique based on adaptive filters for the spatially continuous measurement of HRTFs on a circle. This technique allows a very high azimuth resolution in a fast manner, i.e. 20 s per elevation angle [10]. This technique is not directly applicable for weakly non-linear systems.

This paper introduces a new HRTF measurement setup, designed and manufactured to deliver an optimal trade-off between measurement duration and accuracy, as shorter measurement times will increase the comfort of the subject at the same time that it will reduce displacement error during the measurement which can cause severe errors, especially at higher frequencies. We first describe the design concept and then the hardware characteristics of the system. Next, a time optimized measurement strategy is described and we conclude showing measurements did with the new system compared with reference measurements of the same artificial head.

DESIGN CONCEPT

The HRTF measurement system presented in this paper was designed with the goal of allowing fast measurements and flexible source positioning. Regarding the fact that a broad-band measurement system is desired, the reciprocal method was discarded. As the best compromise between cost and speed, a hybrid set-up has been chosen.

HRTF databases are composed of many measurements made at many discrete source orientations. If a signal should be presented at a direction that is not contained in the database, then either the nearest neighboring direction is used or a new HRTF has to be interpolated out of its neighboring points. Several techniques have been investigated on how to accomplish this interpolation without generating audible artifacts [11,12,13,14]. The use of a flexible source positioning system allow for distinct measurement grids optimized for different interpolation strategies.

A very important contribution in the field of HRTF measurement is the range extrapolation technique that describes "a way to obtain the range dependence of the HRTF from existing measurements conducted at a single range!" [11]. Because this is a holography problem, the closer to the source the measurement is made, the better the extrapolation behaves numerically. For this reason the measurement arc was planned to be of relatively small size and later constructed with a radius of one meter.

Sound Source

In order to guarantee good SNR during measurement, many other HRTF measurement set-ups use relatively large loudspeakers. Sometimes even two-band systems are used. At close ranges these large speakers will show near-field effects in their directivity pattern. Furthermore, multi-band systems possess an unstable acoustic center. Hence, near-field measurements require the sound source to radiate as similar as possible to an ideal point source for the audible frequency range.

An ideal loudspeaker for a HRTF measurement setup should meet the following design criteria:

1. Broad-band reproduction
2. Low non-linearity
3. Smooth directivity in frontal direction

A real loudspeaker cannot fulfill all these criteria perfectly, so some trade-off must be accepted. The desired frequency range had to be reduced from the complete audible range to the range between 300 Hz and 16 kHz, since below 300 Hz HRTFs show very little individual variation. Still, criterion number one can only be fulfilled by specific broad-band drivers.

Three loudspeaker drivers were tested regarding their frequency range, maximal sound pressure level (SPL), distortion and directivity. The driver with the highest maximal SPL and consequently lowest non-linear distortion had relatively large dimensions and, as could be expected, a bundled directivity; being discarded for these reasons. The other two drivers showed equivalent characteristics, with lower maximal SPL, higher non-linear distortion and directivity pattern closer to omni-directional (consequence of the smaller membrane diameter). Since the loudspeakers on the arc will be relatively close to the microphones, maximal SPL is not as critical and so the smaller driver with 32 mm diameter was chosen, as it also allowed an easier fixation at the enclosure. To radiate in the low frequency region the chosen driver needs an enclosure. According to its Thiele-Small parameters, a volume of at least 100 ml is required. Note that even such a small enclosure can influence the radiated sound field due to its edge diffraction. An optimization of the enclosure was carried out to minimize these effects.

First the driver's membrane velocity was measured with a laser Doppler vibrometer at 154 points and these values were used as input data for the loudspeaker simulation. The vibrometry results showed the presence of eigenmodes only at very high frequencies. Three forms of enclosure were simulated: a cylinder with rounded front edge, a cylinder with both front and back edges rounded and a drop-like enclosure. All forms avoid, in varying degrees, sharp edges responsible for

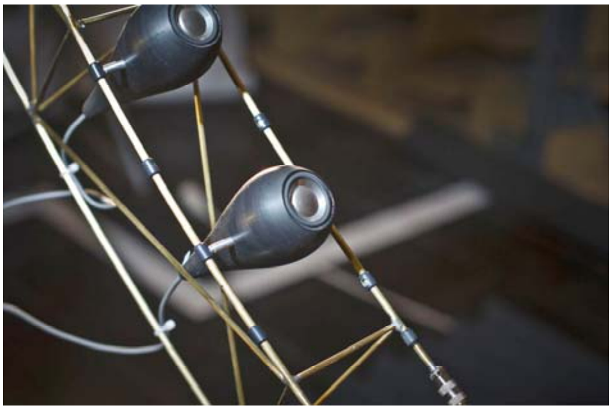

Figure 1 – Developed drop-like loudspeaker with mounting to the arc.

diffraction. Simulation results showed that the drop-like enclosure has the least influence in the loudspeakers frequency response and directivity for a point 1 m away from the membrane.

Furthermore, influences due to possible sound reflections by neighboring loudspeakers have to be considered. In order to verify which form delivers the best result, a simulation was made with three identical loudspeakers placed at an arc of 1 m radius placed 10° apart. The central loudspeaker was set as the sound source while the other two loudspeakers were left inactive, as mere diffraction bodies. Again, the drop-like enclosure showed a slightly lower influence in the radiated sound field and was therefore chosen as the

Figure 2 – Frequency response of the drop-like loudspeaker measured at 1 m and 1 V.

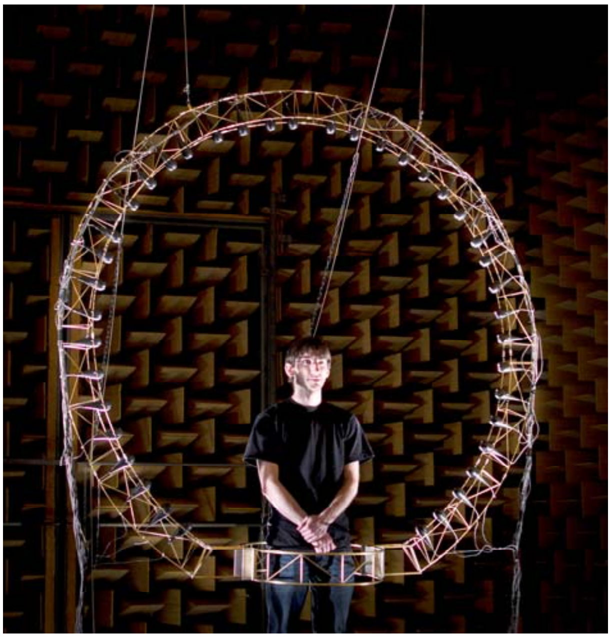

Figure 3 – Measurement of the individual HRTF of a subject.

form for the loudspeaker enclosure. The loudspeaker designed is shown in Figure 1 and its frequency response in Figure 2.

Supporting Arc

The design of the arc to support the loudspeakers is also focused on minimizing its influence on the radiated sound field, i.e. to avoid reflection and diffraction effects. Although easier to manufacture, bulky structures have a very high influence on the sound field and should thus be avoided. On the other hand, a thin metal rod can be considered as acoustically transparent if its diameter is much smaller than the wavelength of the impinging sound wave. The supporting arc was therefore designed with thin metal rods in a trellis structure seeking to minimize disturbing scattering effects while providing sufficient stability.

A first prototype consisting of a 30° arc section was built to verify its influence on the sound field radiated by the drop-like loudspeaker. The frequency response of the drop-like loudspeaker was measured at an anechoic chamber with and without the supporting arc. Variation of less than 1 dB was observed in the frequency response and differences that could be obviously traced back to the supporting arc were verified in the impulse response [15].

For virtual reality applications the use of near-field HRTFs is highly interesting, so range extrapolation of measured HRTFs becomes of importance. As already mentioned, outward range extrapolation (i.e. from near-field measurements to the far-field) are theoretically more stable than inward extrapolation, so a relatively small radius of 1 m was chosen for the arc.

As a person has to stay in the middle of the loudspeaker array, the use of a complete circle is not feasible, hence an arc of 300° was chosen, allowing measurements of elevation angle from -60° to 90°.

Due to its lightly built structure, the supporting arc has an under-damped oscillatory behavior. If the supporting arc where thus to be rotated, a long settling time would be necessary until the arc reaches again its rest position. For this reason it was decided to keep the arc stationary and to rotate the subjects inside the arc with the help of the turntable and a head-rest device to help the test subjects to keep still during measurement.

Data Acquisition and Amplifiers

In order to drive all loudspeakers independently a multi-channel measurement setup has been designed. The computer is connected via IEEE 1396 Firewire 400 to a *Presonus Lightpipe* in cascade with a *Presonus FireStudio*. From there five *Behringer ADA 800* AD/DA to ADAT converters are connected via TOSLINK. Two 20-channel, low noise and low distortion amplifiers with a maximum power of 10 *W* per channel have been designed. The condenser microphones are directly connected to the *Presonus FireStudio*.

Sampling Grid

In general the loudspeakers can be placed at almost any place on the arc only restrained by the supporting trellis structure. There are several sampling strategies on the sphere that allow more or less efficient conversions into spherical harmonics for further calculation [16]. Even for incomplete spheres the quadrature sampling schemes show advantages compared to randomly distributed loudspeakers. As the measurement system uses only horizontal rotation, the choice of sampling is limited to longitude-latitude grids, i.e., grids with a set of elevations that are used for several azimuthal rotation angles. For the measurements showed later in this paper, a Gaussian grid of order 23 was chosen, with the four lowest loudspeakers missing due to the open lower end of the arc. The placement of the loudspeakers however does have some minor deviations from the exact positions of the Gaussian sampling, caused by structural restrictions. The exact locations of the loudspeakers were determined by acoustic measurement and were used for further data processing.

EXCITATION SIGNAL

The design of the excitation signal has a great influence on the quality and the duration of the measurement. Since new measurement strategies are available for MIMO systems the used optimizations are summarized.

Multiple-exponential sweep method

Commonly systems in acoustics are assumed to be linear time invariant (LTI). The well-known LTI system theory along with various excitation signals for correlation measurement techniques is applicable. Exponential sweeps are known to have advantages when it comes to non-linear systems [7]. Non-linear behavior of $t_{harm,i}$ the system is observed as anti-causal impulse responses for different harmonic orders $\boldsymbol{i}$ separately. The multiple-exponential sweep method (MESM) proposed by MAJDAK is applicable for weakly non-linear systems [9]. By this method the measurement duration can be significantly reduced when the number or sources or loudspeakers L is high. MAJDAK introduced two different strategies. One called *overlapping*, where the harmonic impulse responses appear between the impulse responses of interest; and another strategy called *interleaving*, where the impulse responses of interest are grouped together and then a group of all non-linear impulse responses follows. A combination of both strategies is given with an optimization algorithm by MAJDAK as well.

The measurement duration of N multiple sweep measurements with a sweep of length t_{sweep} and a silence in the end of length $t_{stopMargin}$ is given by

$$t_{MESM}(L) = (L-1)t_{wait} + t_{sweep} + t_{stopMargin}, \quad (1)$$

compared to the duration of a conservative, separate measurement

$$t_{separate} = L(t_{sweep} + t_{stopMargin}). \quad (2)$$

For large values of L the measurement duration becomes proportional to t_{wait} instead of the length of the sweep.

This MESM method can yield impulse responses that have the same quality as separately and consecutively measured impulse responses. The signal to noise ratio and the temporal and spectral structure of both results remain the same if the following requirements are fulfilled: The system has to be at most weakly nonlinear, i.e. the number of harmonic impulse responses has to be small. In case non-linearity is observed, the level has to be kept constant during both measurement and calibration – note that this constrained has not been stated in the original paper. The length of the impulse response should be much smaller than the smallest time t_{wait} between two subsequent sweeps. Once the weakly non-linear loudspeakers play back the MESM signal no further weak non-linearity are allowed, i.e., the microphones and preamplifiers have to be driven in a straight linear range only.

Optimization Strategy

A new optimization strategy is used taking into account the expected structure of the theoretical impulse response of the system measured with an exponential sweep as shown in Figure 4. The impulse response of the loudspeaker and the HRTF is very short with an approximate duration of 4 ms. This part of the impulse

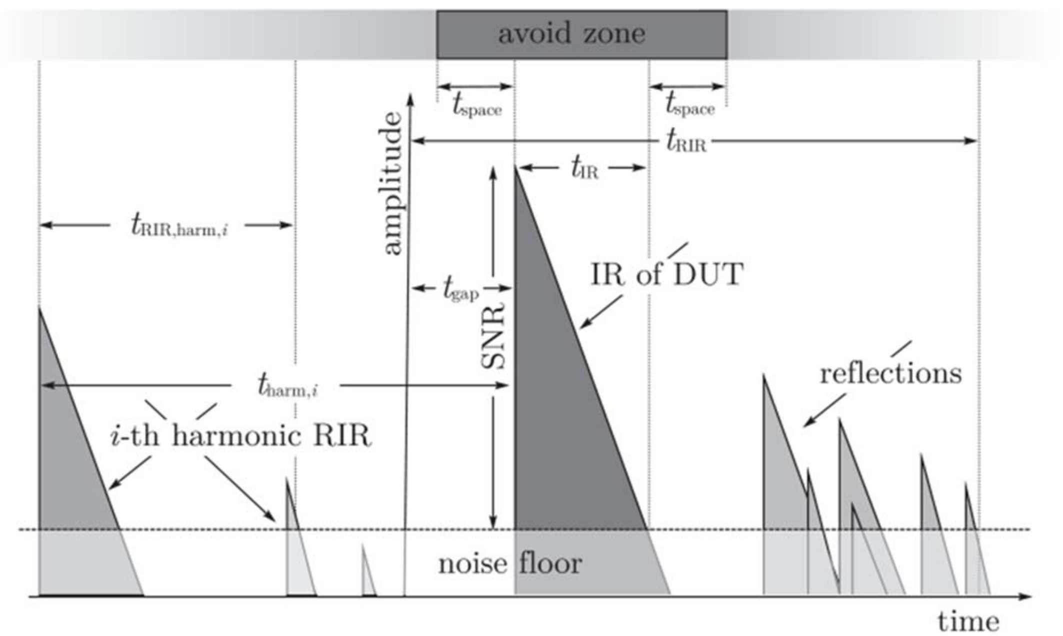

Figure 4 – Temporal structure of an impulse response measured with an exponential sweep.

response carries the important information and shall never be superposed by harmonic impulse responses. It is therefore called *avoid zone*. On the other hand, the impulse response of the hemi-anechoic chamber of the institute shows reflections from the floor, supports and mounts and doors in the order of 40 ms. The remaining impulse response of the room does not carry any useful information an can therefore be used as a place holder for the harmonic impulse response. This little trick enables us to find optimized sweep parameters to fulfill this requirement by using t_{wait} very close to these 40 ms. As this time is usually much smaller than the measurement signal the MESM reduces the measurement time dramatically. Since the subject has to be measured under N different orientation angles the measurement time of the optimized MESM method reads as $N \cdot t_{MESM}(L)$.

MEASUREMENT RESULTS AND POST-PROCESSING

As a proof of concept the HRTFs of an artificial head were measured in order to compare it with existing measurements results of the same head[1].

In Figure 6 two HRTFs for the same direction (on median plane, elevation of 86°) obtained with the new measurement system and the earlier system are depicted. Despite the fact that the HRTFs were measured at different distances (1 m and 2 m) the resulting magnitude spectra show a good agreement.

Having obtained all signals as measured with the MESM method, the impulse responses of all directions can be extracted by truncation and is ready for further processing. To gain a spatially continuous representation of the HRTFs the set of functions can be decomposed into a set of multipoles, a product of spherical harmonics for the angular part and the outgoing Hankel function as the radial part.

As the data coverage does not include the lower part of the sphere, an exact spherical harmonic transform is not given. Applying suitable regularization algorithms, however, still allows the transformation of the measured HRTFs into the spherical harmonic domain [17].

The obtained data can be used to apply both range extrapolation and interpolation methods in order to derive the spatially continuous representation of the head-related transfer function at any point in space around the listener [11].

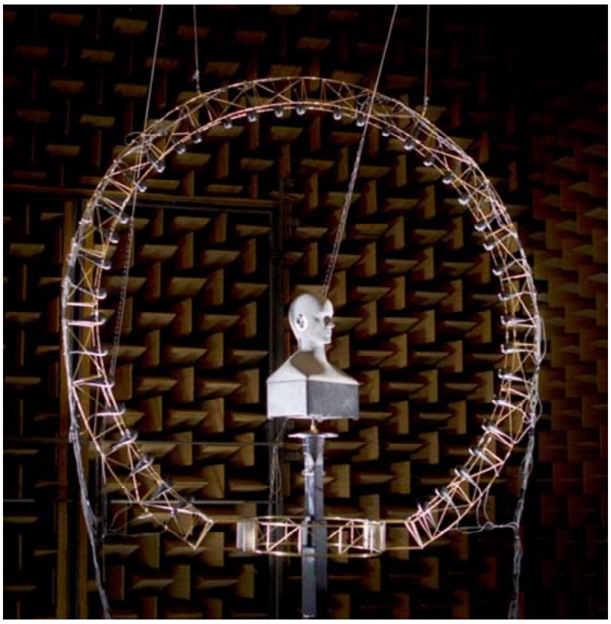

Figure 5 – Photo of the HRTF measurement set-up while measuring the HRTFs of a dummy head.

CONCLUSIONS

A new HRTF measurement set-up was developed in order to allow the fast acquisition of individual HRTFs with satisfactory signal quality. The system is composed of 40 loudspeakers, a supporting fixed arc of a radius of 1 m and a turntable with a head fixation for rotation of the test subjects.

A measurement loudspeaker for this specific task has been designed with a broad-band frequency response and a defined and stable acoustical center. The first measurement conducted with the new system showed good agreement with reference measurements of the same dummy head.

A complete measurement can be concluded in less than six minutes. In a next step the presented method can be enhanced similar to the method described by ENZNER towards a multichannel continuous measurement setup. This results in a continuous rotation of the subject, leading to frequency dependent measurement angles that can be easily compensated by the presented modal approach.

[1] The earlier system was of the sparse array type and took 8 h to measure an equiangular grid with 2° by 2° resolution at a distance of 2 m from the head.

Figure 6 – HRTF measured at 86° elevation and 0° azimuth.

(Right channel shifted by -20 dB)

ACKNOWLEDGMENTS

The authors would like to thank Piotr Majdak for valuable discussions, Johannes Klein for first testing of the routines and photography and Benedikt Krechel for ongoing work towards a continuous measurement. The ITA-Toolbox for MATLAB has been used for data acquisition and post processing (www.ita-toolbox.org).

REFERENCES

[1] H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, “Head-Related Transfer Functions of Human Subjects,” *Journal of the Audio Engineering Society*, vol. 43, no. 5, pp. 300–310, 1995.

[2] D. R. Begault, E. M. Wenzel, and M. R. Anderson, “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source,” *Journal of the Audio Engineering Society*. *Audio Engineering Society*, vol. 49, no. 10, pp. 904–16, Oct. 2001

[3] D. N. Zotkin, R. Duraiswami, E. Grassi, and N. A. Gumerov, “Fast head-related transfer function measurement via reciprocity,” *The Journal of the Acoustical Society of America*, vol. 120, no. 4, p. 2202, 2006.

[4] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, “Transfer Characteristics of Headphones Measured on Human Ears,” *Journal of the Audio Engineering Society*, vol. 43, no. 4, pp. 203–217, 1995.

[5] A. Bronkhorst, “Localization of real and virtual sound sources,” *The Journal of the Acoustical Society of America*, vol. 98, no. 5, p. 2542, 1995.

[6] V. R. Algazi, R. Duda, D. Thompson, and C. Avendano, “The CIPIC HRTF database,” in *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*. 2001, pp. 99–102.

[7] S. Müller and P. Massarani, “Transfer-function measurement with sweeps,” *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[8] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” 2000, pp. 18–22.

[9] P. Majdak, P. Balazs, and B. Laback, “Multiple exponential sweep method for fast measurement of head-related transfer functions,” *Journal of the Audio Engineering Society*, vol. 55, no. 7/8, p. 623, 2007.

[10] G. Enzner, “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*, vol. 2, 2008, pp. 393–396.

[11] R. Duraiswami, D. N. Zotkin, and N. A. Gumerov, "Interpolation and range extrapolation of hrtfs," in *IEEE Conference on Acoustics, Speech and Signal Processing*, 2004.

[12] R. C.-M. Hom, V. R. Algazi, and R. O. Duda, "High-Frequency Interpolation for Motion-Tracked Binaural Sound," in *the 121th AES Convention*, San Francisco, USA, 2006.

[13] L. Wang, F. Yin, and Z. Chen, "Head-related transfer function interpolation through multivariate polynomial fitting of principal component weights," *Acoustical Science and Technology*, vol. 30, no. 6, pp. 395–403, 2009.

[14] W. Zhang, T. D. Abhayapala, R. a. Kennedy, and R. Duraiswami, "Insights into head-related transfer function: Spatial dimensionality and continuous representation." *The Journal of the Acoustical Society of America*, vol. 127, no. 4, pp. 2347–57, Apr. 2010.

[15] B. Masiero, M. Pollow, and J. Fels, "Design of a Fast Broadband Individual Head-Related Transfer Function Measurement System," in *Proceeding of Forum Acusticum*, Aalborg, Denmark, Jun. 2011.

[16] F. Zotter, "Analysis and synthesis of sound-radiation with spherical arrays," PhD Thesis, KU Graz, Austria, 2009.

[17] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, and M. Noisternig, "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition," *Acta Acustica united with Acustica*, vol. 98, no. 1, pp. 72–82, 2012.