ANALYSIS AND OPTIMAL CONTROL OF LMS-TYPE ADAPTIVE FILTERING FOR CONTINUOUS-AZIMUTH ACQUISITION OF HEAD RELATED IMPULSE RESPONSES

Gerald Enzner

Institute of Communication Acoustics, Ruhr-University Bochum, D–44780 Bochum
Email: gerald.enzner@rub.de, Phone: +49-234-32-25392

ABSTRACT

Head related impulse responses (HRIRs) are the key to spatial realism in auditory virtual environments (AVEs). However, the measurement of discrete-azimuth HRIRs and their interpolation has been recognized as a tedious and delicate experimental procedure. We therefore suggest an adaptive filtering concept for continuous HRIR acquisition that completely avoids the traditional sampling and interpolation issue. Using an LMS-type adaptive algorithm, the HRIRs – at any azimuth – are extracted from a one-shot binaural recording. During data acquisition, the subject of interest is continuously rotated in the horizontal plane in order to capture the corresponding spatial information. In particular, the paper provides a profound theoretical and experimental analysis of the resulting HRIR inaccuracy in terms of the mean-square error. Furthermore, the optimal step-size parameter of the LMS-type adaptive algorithm is determined for which the minimum HRIR inaccuracy is attained.

Index Terms— acoustic filters, adaptive filters, virtual reality

1. INTRODUCTION

Research has proven that accurate sound localization in AVEs is difficult to achieve on the basis of non-individual HRIRs [1]. However, the acquisition of personalized HRIR data requires accelerated measurement procedures to reduce the cost of HRIR measurements and to increase the readiness of subjects for obtaining their individual HRIRs for the use in spatial sound systems.

HRIRs are generally measured in anechoic chambers, either using circular loudspeaker arrays or a single loudspeaker which is steered mechanically to discrete azimuth. In the center of the circle, the subject of interest is equipped with a binaural recording system [2]. For each individual loudspeaker position, the respective HRIR measurement is then carried out by probe noise reproduction and subsequent system identification based on the recorded microphone signals [3, 4, 5]. An azimuth-spacing of 5 degrees is necessary to allow for plenacoustic interpolation of the HRIRs without spatial aliasing [6, 7]. Unfortunately, most of the existing measurement procedures for sampled HRIRs are time-consuming and require considerable skill and experience.

Aiming at spatial realism and localization accuracy in AVEs, much work has also been devoted to the interpolation of HRIRs [8, 9, 10], but it remains a delicate issue. Especially in the virtualization of moving sound sources, the presence of interpolation errors potentially causes sound artifacts [11]. Better support of moving sources in virtual auditory space essentially requires the acquisition of plenacoustic (i.e., continuous) HRIRs.

In [12, 13], the acquisition of HRIRs for all azimuthal directions has been addressed by rotating the subject of interest during the recording. For the subsequent system identification, however, [12] still relies on the cross-correlation technique which is commonly used in the measurement of discrete-azimuth HRIRs. While [13] applies an elegant projection-slice theorem to reconstruct spatio-temporal sound fields and the respective HRIRs, it requires the loudspeaker signal to be designed carefully.

In order to overcome traditional HRIR sampling and interpolation, we adopt the continuous rotation of the subject and propose LMS-type (least mean-square) adaptive filtering to extract the time-varying HRIRs at any azimuth. Using adaptive filtering is motivated by the fact that it has been tailored for the estimation of unknown and possibly time-varying impulse responses from noisy measurement data [14]. The efficiency and simplicity of recursively operating LMS-type adaptive filters basically enables the extraction of plenacoustic HRIRs at runtime of spatial sound systems. Naturally, the binaural recordings and the loudspeaker signal have to be stored in the AVE instead of sampled HRIRs.

The proposed concept for continuous HRIR acquisition on the circle is described in Sec. 2. An analytic prediction of the achievable HRIR accuracy using LMS-type adaptive filtering is derived in Sec. 3 and experimental results, presented in Sec. 4, verify our theory. The conclusions of this work are finally drawn in Sec. 5.

2. CONTINUOUS-AZIMUTH ACQUISITION OF HEAD RELATED IMPULSE RESPONSES

A continuous 360° rotation of the loudspeaker or the subject of interest, as illustrated by Fig. 1, yields a "one-shot" binaural recording of the microphone signals $y_i(k)$, $i \in \{1, 2\}$. This recording represents the acoustical transfer of the known loudspeaker signal through all possible HRIRs on the circle. For system identification, the loudspeaker signal could be chosen as a maximum length or perfect sequence [15], but, for the sake of simplicity and mathematical tractability, we simply choose white noise $x(k)$.

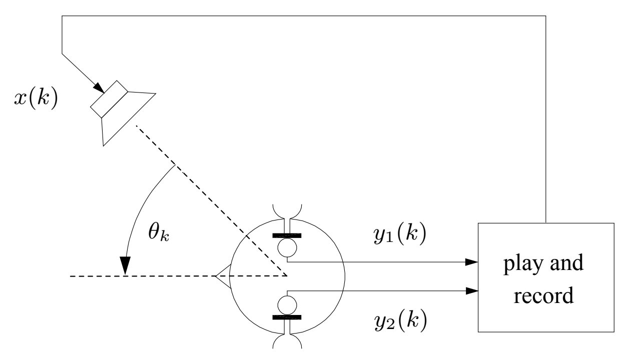

Fig. 1. Measurement setup based on continuous rotation.

2.1. System Model

The propagation of soundwaves from fixed sources in space to fixed receivers is usually described by linear time-invariant systems. Prerequisite for the measurement of such acoustical systems is the employment of sufficiently linear electroacoustic transducers with broadband reproduction and recording capabilities.

Also in the case of rotating HRIRs, according to Fig. 1, we may describe the recorded signals $y_i(k)$ at discrete time k by means of a linear convolution model for the probe noise $x(k)$ and time-varying HRIRs $h_i(\kappa, \theta_k)$, $\theta_k = \omega k T_s$, where $\omega = 2\pi/T_{360}$ is the angular frequency of the subject, T_{360} the duration of a 360° revolution, and $T_s = 1/f_s$ the temporal sampling interval:

$$y_i(k) = \sum_{\kappa=0}^{N} x(k-\kappa) h_i(\kappa, \theta_k) + n_i(k) \; . \qquad (1)$$

In this discrete-time model, the symbol N denotes the effective length of the HRIRs and $n_i(k)$ represents independent observation noise at the two binaural recording positions.

In the case of time-varying systems, as given by rotating HRIRs, it should be noted that the validity of the linear convolution model yet rests upon the assumption that the time-constant of impulse response changes is significantly larger than the HRIR memory N.

2.2. Extraction of Plenacoustic HRIRs on the Circle

First, let

$$\mathbf{x}(k) = (x(k), x(k-1), \ldots, x(k-N+1))^T \qquad (2)$$

denote an assembly of the most recent HRIR input samples and let

$$\mathbf{h}_i(\theta_k) = (h_i(0,\theta_k), h_i(1,\theta_k), \ldots h_i(N-1,\theta_k))^T \qquad (3)$$

denote a vector representation of the HRIR coefficients.

Aiming at the minimum mean-square output error, $\mathrm{E}\{e_i^2(k)\}$, with $e_i(k)$ as defined below, the sample-based normalized least mean-square (NLMS) algorithm for the recursive calculation of the estimate $\widehat{\mathbf{h}}_i(\theta_k)$ of the acoustical paths $\mathbf{h}_i(\theta_k)$ reads [14]:

$$\widehat{\mathbf{h}}_i(\theta_{k+1}) \;=\; \widehat{\mathbf{h}}_i(\theta_k) + \mu_0 \frac{e_i(k)\mathbf{x}(k)}{||\mathbf{x}(k)||_2^2} \qquad (4)$$

$$e_i(k) \;=\; y_i(k) - \widehat{\mathbf{h}}_i^T(\theta_k)\mathbf{x}(k) \; . \qquad (5)$$

Decorrelating adaptive filters were not considered for this paper, because of the assumed white noise input to the HRIRs.

The stepsize μ_0 is the key parameter in balancing the tracking behavior and the noise rejection of the NLMS algorithm, while the angular frequency $\omega = 2\pi/T_{360}$ is the corresponding key parameter of the rotating measurement setup. Both parameters have to be adjusted with respect to each other in order to minimize the system identification error $e_i(k)$. Faster rotation of the recording equipment requires larger μ_0. The optimal choice of μ_0, for certain recording conditions, is determined analytically in Sec. 3.2.

Sample impulse responses $|h_i(\kappa, \theta_k)|$ for the azimuth $\theta_k = 90^\circ$ are depicted in logarithmic scale in Fig. 2. The expected interaural time and level differences between the two ear positions are clearly visible and we observe a noise floor (in the left ear diagram) which is at least 50 dB below the direct sound amplitude at the right ear. The depicted HRIRs were determined with a revolution time of $T_{360} = 20\,s$, using an artificial head and torso simulator, and a stepsize of $\mu_0 = 0.5$ of the NLMS algorithm. Furthermore, the recommended

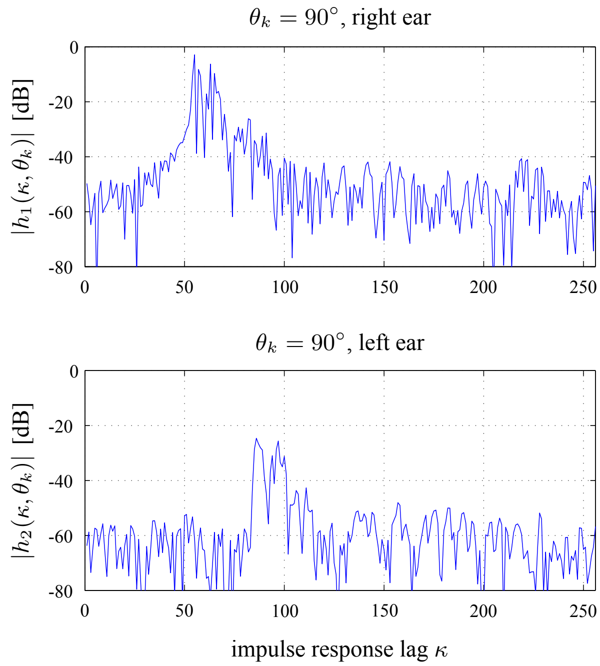

Fig. 2. Snapshot of rotating HRIR measurements. $T_{360} = 20\,s$.

HRIR length of $N = 256$ for the underlying sampling frequency $f_s = 44.1$ kHz has been used [16]. In this case, the time-constant of the NLMS algorithm, which is $N/(2\mu_0)$ samples [14], corresponds to an angular interval of 0.1°. Since the required angular spacing for HRIR sampling is $\Delta\theta = 5^\circ$, this setup causes hardly any spatial smoothing of the estimated HRIRs.

3. PREDICTED HRIR INACCURACY USING NLMS

In order to determine an analytic prediction of the resulting HRIR accuracy using NLMS, we first have to develop a dynamical model process to describe the physical behavior of the rotating HRIRs. In the end, we have to map our prediction of the HRIR accuracy to a measurable quantity that allows for experimental verification.

3.1. Dynamical Model of the Rotating HRIRs

Consider the first-order statistical Markov process as an intuitive model of the gradual changes of the rotating HRIRs:

$$\mathbf{h}_i(\theta_{k+1}) = a \cdot \mathbf{h}_i(\theta_k) + \Delta\mathbf{h}_i(\theta_k) \; , \quad a < 1 \; . \qquad (6)$$

With $a^{k_0} = 1/\mathrm{e}$, let $\tau = k_0 T_s$ denote the time-constant of this process, and assume that the variability of the physically rotating HRIR is governed by the same time-constant than the discrete model process. The model parameter a is then obviously related to τ through the expression $a = \exp(-T_s/\tau)$.

Further assume that the bandwidth ω_g of the HRIR variability is related to the time-constant τ through a certain time-bandwidth product $\tau \cdot \omega_g = \rho$, where ρ is an appropriate constant. This bandwidth can be quantified on the basis of the minimum angular sampling interval $\Delta\theta = 5^\circ$ needed for the plenacoustic function on the circle [6, 7]. With a revolution time T_{360} of the measurement setup, the corresponding minimum time interval for lossless HRIR sampling is $\Delta t = T_{360} \cdot \Delta\theta/360^\circ$. According to Nyquist's sampling theorem, this corresponds to the bandwidth $\omega_g = 2\pi \cdot \frac{1}{2} \cdot \frac{1}{\Delta t} = (\pi \cdot 360°)/(T_{360} \cdot \Delta\theta)$. Thus, we have the model parameter

$$a = \exp\left(-\frac{\pi \cdot T_s \cdot 360°}{\rho \cdot T_{360} \cdot \Delta\theta}\right) \qquad (7)$$

as a function of the parameters T_{360} and T_s of the recording.

3.2. HRIR Deviation Using NLMS

Generally, the error signal $e_i(k)$ of adaptive filters, as formulated in Sec. 2.2, consists of a system identification residual $b_i(k)$ and the measurement noise $n_i(k)$. In vector notation, we have:

$$\begin{aligned} e_i(k) &= b_i(k) + n_i(k) & (8)\\ &= \left(\mathbf{h}_i(\theta_k) - \widehat{\mathbf{h}}_i(\theta_k)\right)^T \mathbf{x}(k) + n_i(k) & (9)\\ &= \boldsymbol{\epsilon}_i^T(\theta_k)\,\mathbf{x}(k) + n_i(k)\ . & (10)\end{aligned}$$

In the following, we wish to determine the relative inaccuracy of our HRIR estimation, D/σ_h^2, in terms of the mean-square HRIR deviation $D = \mathrm{E}||\boldsymbol{\epsilon}_i(\theta_k)||_2^2/N$ with respect to the variance $\sigma_h^2 = \mathrm{E}||\mathbf{h}_i(\theta_k)||_2^2/N$ of the original HRIR coefficients, given the relevant parameters of the measurement setup.

For the underlying dynamical model in (6), the distortion D consists of an estimation error variance due to the measurement noise $n_i(k)$ and a lag variance due to the process noise $\Delta\mathbf{h}_i(\theta_k)$. According to [14], the steady-state mean-square deviation obtained from the LMS algorithm, assuming small stepsize μ and white noise $x(k)$, is given by

$$\mathrm{E}||\epsilon_i^T(\theta_k)||_2^2 = \frac{\mu}{2} N\sigma_n^2 + \frac{1}{2\mu}\frac{\mathrm{E}||\Delta\mathbf{h}_i(\theta_k)||_2^2}{\sigma_x^2}\ , \qquad (11)$$

where σ_n^2 and σ_x^2 denote the variances of the observation noise and the loudspeaker signal, respectively.

Based on the identity $\mu = \mu_0/||\mathbf{x}(k)||_2^2$, the proposed NLMS algorithm for HRIR acquisition is inherited from that more generic LMS algorithm. Additionally using $S = N \cdot \sigma_h^2 \cdot \sigma_x^2$ and taking the approximation $||\mathbf{x}(k)||_2^2 \approx N\sigma_x^2$ into account, we can deduce a formula for our HRIR inaccuracy using NLMS:

$$\frac{D}{\sigma_h^2} = \frac{\mu_0}{2}\frac{\sigma_n^2}{S} + \frac{1}{2\mu_0}\frac{\mathrm{E}||\Delta\mathbf{h}_i(\theta_k)||_2^2}{\sigma_h^2}\ . \qquad (12)$$

Applying the mean-square (inner product) on both sides of the equality in (6), the following expression for the process noise variance $\sigma_\Delta^2 = \mathrm{E}||\Delta\mathbf{h}_i(\theta_k)||_2^2/N$ is established:

$$\sigma_\Delta^2 = (1-a^2)\sigma_h^2\ . \qquad (13)$$

Thus, we have

$$\frac{D}{\sigma_h^2} = \frac{\mu_0}{2}\mathrm{SNR}_y^{-1} + \frac{1}{2\mu_0}N(1-a^2)\ , \qquad (14)$$

where $\mathrm{SNR}_y = S/\sigma_n^2$ defines the global signal-to-noise ratio at the recording microphones.

Again, the first contribution to the entire HRIR distortion in (14) is essentially caused by the observation noise $n_i(k)$, while the second contribution represents the impact of the variability of the dynamical HRIR model, which in turn depends on the revolution time T_{360} according to (7).

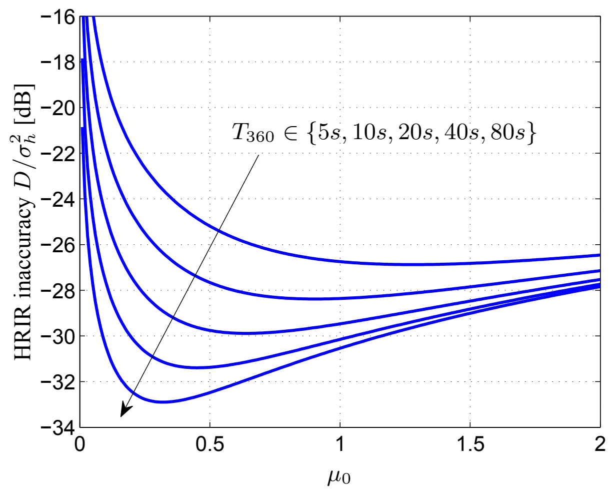

Fig. 3. Predicted HRIR inaccuracy of NLMS. $\mathrm{SNR}_y = 28\,\mathrm{dB}$.

Fig. 3 indicates a significant dependency of the resulting HRIR inaccuracy D/σ_h^2 on the stepsize parameter μ_0 of the adaptive algorithm. On the basis of (14), we can immediately say that the HRIR inaccuracy attains its minimum for the optimum stepsize

$$\mu_{0,\mathrm{opt}} = \sqrt{N \cdot \mathrm{SNR}_y \cdot (1-a^2)} \qquad (15)$$

for which the estimation error variance due to observation noise and the lag variance contribute equally to the distortion D.

3.3. A Measurable Quantity for the HRIR Deviation

Unfortunately, the predicted HRIR inaccuracy D/σ_h^2 in (14) cannot be verified experimentally in the case of noisy HRIR measurements. Neither the distortion D nor the underlying distance $\boldsymbol{\epsilon}_i(\theta_k)$ or the system identification residual $b_i(k)$ are accessible alone.

We therefore consider the measurable quantity σ_e^2/σ_y^2 which can be calculated from the available signals $e_i(k)$ and $y_i(k)$. This quantity can be traced back uniquely to the HRIR inaccuracy. With $\sigma_b^2 = N \cdot D \cdot \sigma_x^2$, and thus $\sigma_b^2 = S \cdot D/\sigma_h^2$, we readily find:

$$\begin{aligned} \frac{\sigma_e^2}{\sigma_y^2} &= \frac{\sigma_b^2}{\sigma_y^2} + \frac{\sigma_n^2}{\sigma_y^2}\\ &= \left(\frac{D}{\sigma_h^2}\right)\cdot\frac{S}{S+\sigma_n^2} + \frac{\sigma_n^2}{S+\sigma_n^2}\\ &= \left[\left(\frac{D}{\sigma_h^2}\right)\cdot SNR_y + 1\right]\cdot\left[SNR_y + 1\right]^{-1}\ . \qquad (16)\end{aligned}$$

Given the $\mathrm{SNR}_y = S/\sigma_n^2$, the predicted HRIR inaccuracy according to (14) is thus mapped onto a prediction of σ_e^2/σ_y^2 that can be compared to measured values of this ratio. In this way, the underlying HRIR inaccuracy D/σ_h^2 can be verified at least indirectly.

4. RESULTS

The solid line in Fig. 4 represents our prediction of the HRIR inaccuracy D/σ_h^2 according to (14) and (7) − using the optimal stepsize in (15) − as a function of the revolution time T_{360}. Essentially, we observe that most of the achievable accuracy is obtained for revolution times below one minute. For very fast rotation, i.e., $T_{360} < 10\,s$, the resulting accuracy however deteriorates drastically.

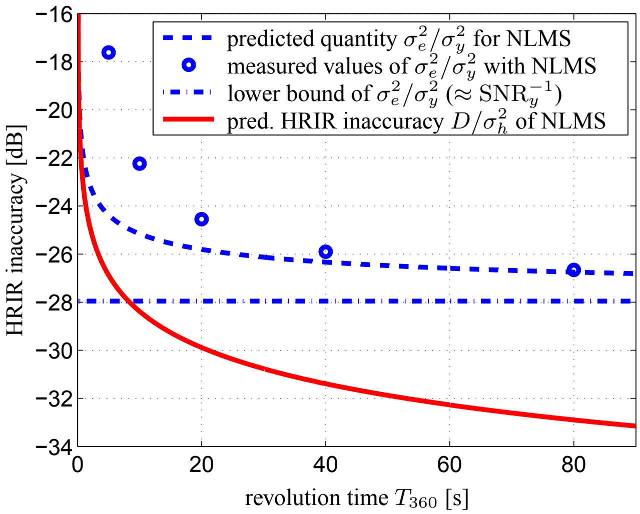

Fig. 4. Prediction and measurement of the HRIR inaccuracy using NLMS with $N = 256$ and $\mu_0 = \mu_{0,\mathrm{opt}}$. $\mathrm{SNR}_y = 28\,\mathrm{dB}$. $\rho = 200$.

When rating the values of the HRIR inaccuracy D/σ_h^2, it should be noted that the visible noise floor of the corresponding HRIR coefficients lies a factor of about $N \cdot \sigma_h^2/D$ below the direct sound amplitude. This is due to the extremely unequal energy distribution of the HRIRs across the impulse response lag. For $T_{360} = 20\,s$, e.g., we have $\sigma_h^2/D \approx 30\,\mathrm{dB}$ while the factor $N = 256$ contributes another 25 dB. This is confirmed by the HRIR example in Fig. 2, where the noise floor is indeed more than 50 dB below the direct sound.

An experimental verification of the theoretical results has been carried out using the dynamic measurement setup of Fig. 1 in an anechoic chamber. An artificial head and torso simulator, equipped with in-the-ear microphones, was rotated by means of a turntable system with variable revolution time. The fixed loudspeaker, at a distance of 1.2 m from the artificial head, reproduced white noise during the 360° rotation. The HRIRs were then extracted using NLMS and the ratio σ_e^2/σ_y^2 was calculated globally.

The dashed line in Fig. 4 represents the prediction of the measurable HRIR accuracy, σ_e^2/σ_y^2, obtained from D/σ_h^2 through the mapping in (16). For large T_{360}, it coincides well with our discrete measurements of this quantity, but for fast rotations, demanding larger stepsize μ, the predicted accuracy diverges from the measurements as expected [14]. The coincidence of measurements and prediction, for large T_{360}, has been achieved on the basis of the time-bandwidth product $\rho = 200$ obtained experimentally.

The finite $\mathrm{SNR}_y = 28\,\mathrm{dB}$ in the measurements done here owes to the turntable engine and to the HRIR tail. The latter is due to unwanted reflections in the non-ideal measurement setup, which are not modeled by the adaptive filter of length $N = 256$. With higher SNR_y, the resulting HRIR inaccuracy can be lowered directly.

5. CONCLUSIONS

In contrast to the delicate interpolation of discrete HRIRs, the suggested adaptive filtering method efficiently allows for HRIR acquisition at any azimuth. The proposed technique thus enables a most natural and exact way of creating virtual auditory spaces in realtime, including an improved support of moving sources. In particular, we showed that the error signal of the adaptive algorithm is an inherent measure of the accuracy of the resulting continuous-azimuth HRIRs.

6. REFERENCES

[1] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman, "Localization using nonindividualized head-related transfer functions," *Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, July 1993.

[2] Jens Blauert, *Spatial Hearing*, The MIT Press, Cambridge, MA, 1997.

[3] Daniele Pralong and Simon Carlile, "Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature in-ear recording system," *Journal of the Acoustical Society of America*, vol. 95, no. 6, pp. 3435, June 1994.

[4] Bill Gardner and Keith Martin, "HRTF measurements of a KEMAR dummy-head microphone," MIT Media Lab Perceptual Computing Technical Report No. 280, May 1994.

[5] Dorte Hammershoi and Henrik Moller, "Sound transmission to and within the human ear canal," *Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408, July 1996.

[6] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "Plenacoustic function on the circle with application to HRTF interpolation," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, March 2005.

[7] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, October 2006.

[8] M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for HRTFs using a logarithmic error criterion," *IEEE Trans. on Speech and Audio Processing*, vol. 5, pp. 278–287, 1997.

[9] S. Carlile, C. Jin, and V. Van Raad, "Continuous virtual auditory space using HRTF interpolation: Acoustic and psychophysical errors," in *Intl. Symp. on Multimedia Information Processing*, Sidney, Australia, December 2000.

[10] Ramani Duraiswami, Dmitry N. Zotkin, and Nail A. Gumerov, "Interpolation and range extrapolation of HRTFs," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004.

[11] Jens Blauert, "Implementation issues of AVEs," Personal Communication, August 2007.

[12] Kimitoshi Fukudome, Toshimitsu Suetsugu, Takuya Ueshin, Ryo Idegami, and Kazuki Takeya, "The fast measurement of head related impulse responses for all azimuthal directions using the continuous measurement method with a servo-swiveled chair," *Applied Acoustics*, vol. 68, no. 8, pp. 864–884, August 2007.

[13] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1636–1645, September 2007.

[14] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, 4th edition, 2002.

[15] Peter Vary and Rainer Martin, *Digital Speech Transmission*, John Wiley & Sons Ltd., Chichester, England, 2006.

[16] Andreas Silzle, "Länge von Außenohr-Impulsantworten und Auswahl von Reflexionen für interaktive auditive virtuelle Umgebungen," in *Proc. of Deutsche Jahrestagung für Akustik (DAGA)*, Braunschweig, Germany, March 2006.