Analysis and Comparison of different Adaptive Filtering Algorithms for Fast Continuous HRTF Measurement

Camilo Klinkert Correa, Song Li, Jürgen Peissig
Leibniz Universität Hannover, Institut für Kommunikationstechnik, 30167 Hannover, Germany
Email: caklinkert@gmail.com, {song.li, peissig}@ikt.uni-hannover.de

Abstract

Head-related transfer function (HRTF) is widely used for binaural sound reproduction over headphones. However, the acquisition of HRTFs using traditional measurement is usually a time-consuming task and can be only acquired at discrete directions. Recent work has shown, that the HRTF measurement can be speeded up and simplified via continuous acquisition by using an adaptive filter (identification of a time-variant system). With this method the traditional sampling and interpolation of many different positions can be avoided. There are many different adaptive filter types suitable for system identification, therefore this work analyses and compares various adaptive filter algorithms like Least-Mean-Squares (LMS), Normalized-LMS (NLMS), Recursive-Least-Squares (RLS), etc. The goal of this work is to achieve faster convergence speed and low steady-state mean squared error. Thus, the convergence and tracking properties of these algorithms are analyzed and compared. This implies that time-invariant systems (fixed dummy head) as well as time-variant systems (continuously rotated dummy head) are evaluated for convergence speed and tracking capability respectively. Furthermore, the noisy environment during measurements has to be considered. For this reason, it is necessary to simulate different additive noise levels to evaluate the algorithms' behavior under these adverse conditions.

Introduction

Head-related transfer function (HRTF) is fundamental for 3D audio reproduction via headphones. However, HRTFs depend on the listeners ears, head and torso. The perceptual externalization and the accuracy of localization are improved by using individual HRTFs for binaural synthesis [1]. Therefore, the individualized HRTFs are needed. Two different methods are widely known to obtain individual HRTFs. The first method is personalization HRTFs by using non-individualized HRTF database and anthropometric features [2]. The second method is direct measurement of individual HRTFs.

Individual HRTF measurements are usually time consuming and exhausting for the human subjects. Some fast HRTF measurement methods have been developed to improve the acquisition efficiency. For example, HRTF measurement by applying the principle of reciprocity [3] and multiple exponential sweep method (MESM) [4]. Enzner [5] introduced a measurement system for continuous azimuth acquisition of HRTFs based on normalized least mean square (NLMS) adaptive filter.

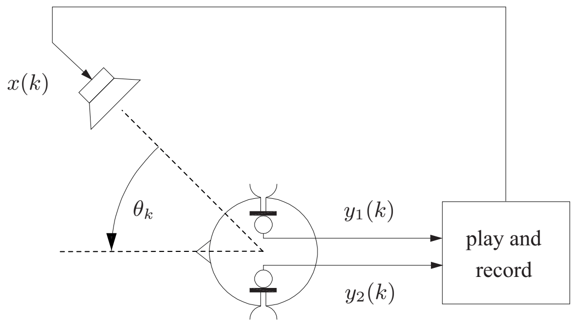

Figure 1: System for HRTF acquisition with continuous rotation [5].

The system illustrated in Fig. 1 was proposed in [5]. In this system, a (dummy) head is rotated continuously during the measurement; a reference signal $x(k)$ is played through a loudspeaker and the ear signals $y_{1,2}(k)$ are recorded with in-ear microphones during the rotation. The HRTFs at any azimuth can be obtained through an adaptive filter. Since the system is time-variant due to the rotation, the estimation performed by the filter algorithm must be fast enough.

In this paper, we concentrate on continuous HRTF acquisition system with different adaptive filtering algorithms. Different adaptive filter algorithms are implemented, analyzed and compared with respect to fast convergence and residual estimation error.

This paper is organized as follows. First, different adaptive estimation methods are presented. The following sections describe the simulation and measurement results. Finally, the conclusions and outlook are drawn.

Adaptive estimation methods

The system described by [5] models the ear signals $y_{1,2}(k)$ as

$$\begin{aligned} y_i(k) &= \sum_{n=0}^{N} x(k-n)h_i(n,\theta_k) + n_m(k) \\ &= \mathbf{h}_i(\theta_k)\mathbf{x}(k) + n_m(k) \end{aligned} \quad (1)$$

where $\mathbf{x}(k)$ is the reference signal, $\mathbf{h}_i(k)$ head-related impulse response (HRIR), $i \in \{1,2\}$ for left and right ears and $n_m(k)$ measurement noise. The estimation error is defined as

$$e_i(k) = y_i(k) - \hat{\mathbf{h}}_i^T(\theta_k)\mathbf{x}(k) \tag{2}$$

where $\hat{\mathbf{h}}_i(\theta_k)$ are the estimated HRIR. Usually, adaptive filtering algorithms can be categorized into two groups: stochastic methods and deterministic methods.

Stochastic methods

The algorithms in this group minimize the estimation mean square error (MSE) using the step-size μ under the assumption that MSE is a stochastic variable. In this group LMS [6], NLMS [5], VSSLMS [7], MVSS [8] and VSNLMS [7] are analyzed.

A. LMS: The Least Mean Squares algorithm, as described in [6], is the stochastic implementation of steepest-descent. Here, the filter coefficient vector update is given by

$$\hat{\mathbf{h}}_i(\theta_{k+1}) = \hat{\mathbf{h}}_i(\theta_k) + \mu e_i(k)\mathbf{x}(k). \tag{3}$$

B. NLMS: In contrast, the normalized LMS as used by [5] updates the coefficients as follows

$$\hat{\mathbf{h}}_i(\theta_{k+1}) = \hat{\mathbf{h}}_i(\theta_k) + \mu\frac{e_i(k)\mathbf{x}(k)}{||\mathbf{x}(k)||_2^2}. \tag{4}$$

C. VSSLMS: A variable step-size for the LMS algorithm is proposed by [7] as

$$\hat{\mathbf{h}}_i(\theta_{k+1}) = \hat{\mathbf{h}}_i(\theta_k) + \mu_i(k)e_i(k)\mathbf{x}(k), \tag{5}$$

whereby the variable step-size $\mu_i(k)$ for each ear is updated by using the squared instantaneous a priori estimation error

$$\mu_i(k+1) = \alpha\mu_i(k) + \gamma e_i^2(k), \tag{6}$$

where $0 < \alpha < 1$ and $\gamma > 0$. Fast convergence is ensured and instability is avoided by restricting $\mu_i(k)$ to $[\mu_{min}, \mu_{max}]$.

D. MVSS: The update of the variable step-size is modified by [8]. Instead of e_i^2 the square of the time-averaged estimate of the autocorrelation of e_i in adjacent time samples is used to update $\mu_i(k)$

$$\mu_i(k+1) = \alpha\mu_i(k) + \gamma p_i^2(k), \tag{7}$$

where

$$p_i(k) = \beta p_i(k-1) + (1-\beta)e_i(k)e_i(k-1) \tag{8}$$

and $0 < \beta < 1$.

E. VSNLMS: A variable step-size NLMS, described by [7], combines the fast convergence of NLMS and the low steady state MSE of VSSLMS. The filter coefficient vector update is given by (4) but the step-size $\mu_i(k)$ is calculated by (6).

Deterministic methods

The deterministic methods minimize a cost function using the existing past values.

RLS: The Recursive Least Squares with forgetting factor is described by [6]. Here, the filter coefficient vector update is given by

$$\hat{\mathbf{h}}_i(\theta_k) = \hat{\mathbf{h}}_i(\theta_{k-1}) + \mathbf{K}(k)e_i(k) \tag{9}$$

where $\mathbf{K}(k)$ is the gain vector.

$$\mathbf{K}(k) = \frac{\mathbf{\Psi}^{-1}(k-1)\mathbf{x}(k)}{\lambda + \mathbf{x}^T(k)\mathbf{\Psi}^{-1}(k-1)\mathbf{x}(k)}, \tag{10}$$

where the forgetting factor $\lambda < 1$ and the $\mathbf{\Psi}^{-1}(k)$ matrix is updated as

$$\mathbf{\Psi}^{-1}(k) = \lambda^{-1}(\mathbf{\Psi}^{-1}(k-1) - \mathbf{K}(k)[\mathbf{x}^T(k)\mathbf{\Psi}^{-1}(k-1)]). \tag{11}$$

Performance evaluation

Three scenarios are defined in order to evaluate the characteristics of the mentioned algorithms:

1. Static at $\theta = 0°$: The loudspeaker is fixed direct in front of the subject.
2. Dynamic steps $\theta = 0° \rightarrow 45° \rightarrow 90°$: the head is turned suddenly twice, each time after two seconds.
3. Continuous rotation $\theta = 0° \rightarrow 360°$: The head is rotated with an angular speed of $6\,°/s$.

The following criteria are verified for the different algorithms: Convergence time, tracking capability, normalized mean square error (NMSE) and Matlab computation time.

Simulation

The CIPIC HRTF Database [9] is used to simulate the described scenarios according to (1). Since the database provides HRTFs for certain angles, the missing angles are linear interpolated. As a reference signal $x(k)$, white gaussian noise is used. Furthermore, microphone simulating noise n_m is added into the ear signals with a signal-to-noise ratio (SNR) of 30 dB.

Measurement

The measurement serves to validate the simulation results. The measurement of scenarios 1 and 2 was performed in an anechoic chamber. For scenario 3 the reference and measured ear signals for continuous HRTF measurement are used from the auditory modeling toolbox (AMT) [10, 11].

Results

The better the estimated $\hat{\mathbf{h}}_i(\theta_k)$ matches the real HRIR $\mathbf{h}_i(\theta_k)$, the smaller $e_i(k)$ is. Therefore it can be used as indicator for quality. However, as [5, 12, 13] mentioned, a small value of $E[e_i^2(k)]$ does not always mean a small system distance. Therefore, the quality index is given by the NMSE as

$$NMSE = \frac{E[e_i^2(k)]}{E[y_i^2(k)]}. \tag{12}$$

1. *Static at* $0°$*:* In this scenario, the results are shown in the interval $[0\,s, 2\,s]$ of figures 2, 3 and 4. RLS has the fastest convergence with and without microphone simulating noise, but shows the highest NMSE. VSSLMS and MVSS are the algorithms with the slowest convergence. VSNLMS and NLMS show the best NMSE without noise. VSNLMS converges as fast as NLMS but reaches better NMSE with noise. These algorithms show similar behavior for measurement and simulation with noise. On average VSNMLS shows the best performance.

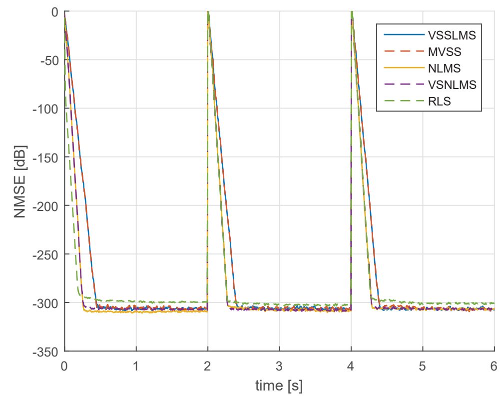

Figure 2: NMSE results for adaptive filtering algorithms simulation without noise in scenarios 1 and 2.

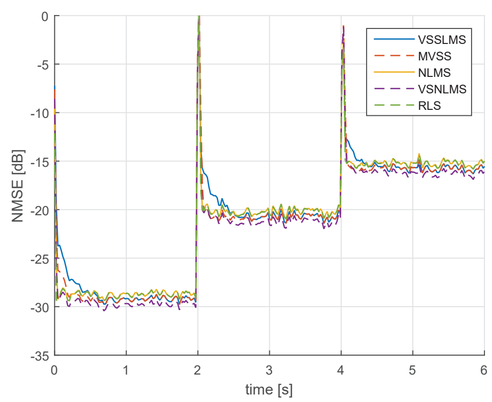

Figure 3: NMSE results for adaptive filtering algorithms simulation with 30 dB SNR in scenarios 1 and 2.

2. *Dynamic steps:* In this scenario, the results are shown in Fig. 2, 3 and 4. The peaks after second 2 ($45°$) and 4 ($90°$) correspond to the sudden change of head direction. In both cases of simulation, VSSLMS shows the slowest and RLS the fastest convergence, but RLS shows the highest NMSE. VSNLMS reaches the best NMSE with noise, even after tracking, so all these algorithms are tracking-capable. Here, the simulations results with noise can also be confirmed after measurement.

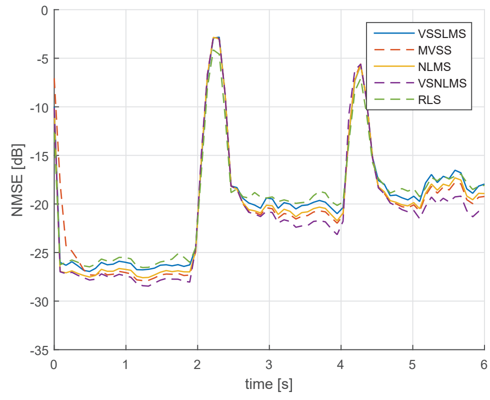

Figure 4: HRTF estimation results in scenarios 1 and 2. while measurement in anechoic chamber.

3. *Continuous rotation:* In this case, fast convergence as well as low NMSE are important since it is a time varying system. Results are shown in Fig. 5, 6 and 7. Results show that VSSLMS has the slowest convergence with and without noise. NLMS and VSNLMS show almost the same NMSE without noise but VSNLMS is better with noise. So on average the best performance is reached by VSNLMS. For measurement the algorithms show similar behavior as simulated with noise. Finally, the Matlab CPU computation time is listed in Table 1.

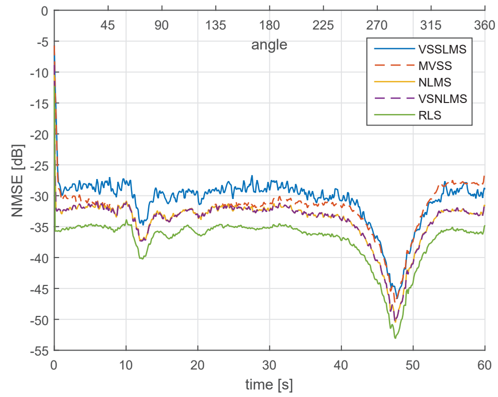

Figure 5: NMSE results for adaptive filtering algorithms simulation without noise for continuous rotation (scenario 3).

Table 1: Matlab computation time for 60 s rotation signal.

VSSLMS	MVSS	NLMS	VSNLMS	RLS
23 s	27 s	29 s	36 s	8807 s

Conclusions

The results of simulation and measurement are consistent. NLMS and VSNLMS are equally good without noise, whereas with microphone simulating noise, VSNLMS is as fast as NLMS but achieves lower NMSE. It was shown that all illustrated algorithms support tracking feature. Furthermore the performance of the algorithms crucially depends on parameter settings

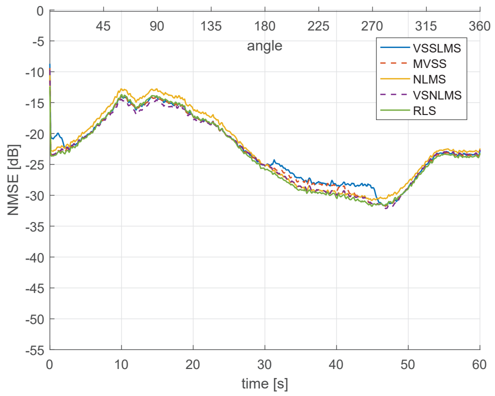

Figure 6: NMSE results for adaptive filtering algorithms simulation with 30 dB SNR for continuous rotation (Scenario 3).

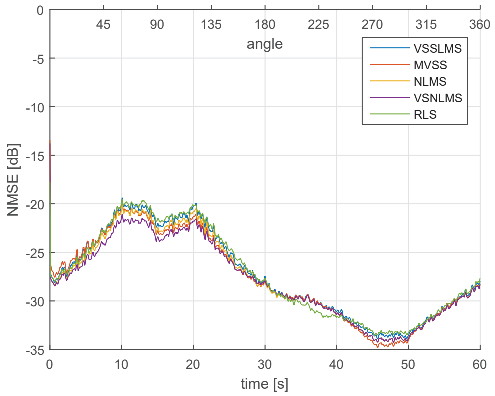

Figure 7: HRTF estimation results for continuous rotation (Scenario 3) for the measurement in an anechoic chamber.

(λ, μ, α, γ) which depend on the measurement noise. The Matlab computation time is significant longer with RLS than with LMS-based algorithms.

Outlook

More continuous HRTF measurements with different rotation speeds have to be carried out in order to ensure a good performance by unconstrained movements. And hence variation of filter parameters with respect to different rotation speeds have to be examined. In order to exploit the fast convergence and overcome the long calculation time of RLS, frequency-based RLS may be implemented. Furthermore a psychoacoustic evaluation of measured HRTFs (e.g. ABX test) has to be performed. Finally our results must be optimized for 2D continuous unconstrained measurement in azimuth and elevation with headtracker.

References

[1] E. Wenzel, M. Arruda, D. Kistler, and F. Wightman, "Localization using non-individualized head-related transfer functions," *Journal of the Acoustical Society of America*, 1993.

[2] D. Y. N. Zotkin, J. Hwang, R. Duraiswaini, and L. S. Davis, "Hrtf personalization using anthropometric measurements," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, Oct 2003, pp. 157–160.

[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, pp. 2202–2215, 2006.

[4] A. Fuß, "Entwicklung eines vollsphärischen multikanalmesssystems zur erfassung individueller kopfbezogener Ãœbertragungsfunktionen," Master's thesis, Technische Universität Berlin, 2014.

[5] G. Enzner, "Analysis and optimal control of lms-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 393–396.

[6] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*. Wiley, 2013.

[7] R. H. Kwong and E. W. Johnston, "A variable step size lms algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1633–1642, Jul 1992.

[8] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: analysis and simulations," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 631–639, Mar 1997.

[9] (2016) The cipic hrtf database. CIPIC- Center for Image Processing and Integrated Computing University of California. [Online]. Available: http://interface.cipic.ucdavis.edu/sound/hrtf.html

[10] (2016) Auxiliary data auditory modeling toolbox. [Online]. Available: https://sourceforge.net/projects/amtoolbox/files

[11] P. Søndergaard and P. Majdak, "The auditory modeling toolbox," in *The Technology of Binaural Listening*, J. Blauert, Ed. Berlin, Heidelberg: Springer, 2013, pp. 33–56.

[12] C. Antweiler and G. Enzner, "Perfect sequence lms for rapid acquisition of continuous-azimuth head related impulse responses," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct 2009, pp. 281–284.

[13] C. Antweiler, A. Telle, P. Vary, and G. Enzner, "Perfect-sweep nlms for time-variant acoustic system identification," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 517–520.