Analysis of Time-Varying System Identification Using the Normalized Least Mean Square Algorithm in the Context of Data-Based Binaural Synthesis

Nara Hahn and Sascha Spors
Institute of Communications Engineering, University of Rostock, Germany
Email: {nara.hahn,sascha.spors}@uni-rostock.de

Introduction

The impulse response of a time-varying acoustic system can be measured by a continuous measurement technique. In such a measurement, a system is continuously excited by a periodic signal while the system changes over time, e.g. movement of source or receiver. The response of the system is captured by a microphone and the instantaneous impulse responses are computed from the microphone signal. Due to its time efficiency, continuous measurement methods are used for (i) the measurement of a large number of acoustic impulse responses e.g. spatial room impulse responses [1, 2] or head-related impulse responses [3]. During the measurement, the receiver (microphone or dummy head) moves on a predetermined trajectory. Depending on the required spatial resolution, an arbitrary number of impulse responses can be extracted and further used for sound field analysis or binaural synthesis [4, 5]. The second application of a continuous measurement technique is (ii) the auralization of dynamic auditory scenes [6]. For data-based binaural synthesis, for instance, binaural room impulse responses (BRIRs) are measured, and the ear signals are generated by filtering a dry source signal with the time-varying BRIRs. In this case, not only the accuracy of the individual impulse responses, but also the transient properties of the synthesized result have to be taken into account.

Several continuous measurement techniques have been proposed in earlier studies [1, 7, 8, 9]. Among others, adaptive system identification using the Normalized Least Mean Square (NLMS) algorithm is preferred due to its low computational complexity [10] and fast tracking ability when used in combination with a specific type of signal, called perfect sequence [8]. NLMS was often used for the measurement of HRIRs [3]. In data-based binaural synthesis of dynamic scenarios, however, the ear signals suffer from artifacts that are presumably attributed to system identification errors [6, 9]. It is unclear, though, why such errors occur, what kind of properties these errors have, and how they affect the performance of the auralization.

In this paper, it is shown that the output of a time-varying system carries limited information on the impulse response history. Therefore, system identification methods implicitly estimate the missing information by interpolation. The inherent interpolation filter of the NLMS algorithm is derived and its properties are discussed. It is proved that the transient properties of the identified system are responsible for the artifacts occurring in auralization.

System Model

In this paper, a finite impulse response (FIR) model is assumed,

$$y(n) = \sum_{k=0}^{N-1} h_k(n) x(n-k) \tag{1}$$

where $x(n)$ denotes the input, $y(n)$ the output, $h_k(n)$ the k-th filter coefficient at time n, and N the maximum length of $h_k(n)$.

For continuous measurement, the system is excited by a periodic perfect sequence $\psi(n) = \psi(n+N)$ [8]. The periodic autocorrelation of a perfect sequence yields,

$$\varphi_{\psi\psi}(n) = \sum_{n'=0}^{N-1} \psi(n+n')\psi(n) = E \cdot \delta_{\bar{n}0} \tag{2}$$

where $\bar{n} \equiv n \bmod N$, and δ_{mn} denotes the Kronecker delta. Without loss of generality it is assumed that $E = \sum_{n=0}^{N-1} |\psi(n)|^2 = 1$. Note that the period is longer than any impulse response of the system and thereby temporal aliasing is avoided [11].

According to (2), two different circular shifts of a given perfect sequence are orthogonal unless the shift difference is an integer multiple of N. Thus, an orthogonal basis set for $\mathbb{R}^N$ can be formed with N different circular shifts of $\psi(n)$. Here, the circular shifts of the time-reversal $\psi(-n)$ (which are also perfect sequences) are considered

$$\mathcal{S} = \Big\{ \psi(-n), \psi(1-n), \ldots, \psi(N-1-n) \Big\}. \tag{3}$$

By using these basis functions, an impulse response can be expanded into orthogonal components as

$$h_k(n) = \sum_{m=0}^{N-1} a_m(n) \psi(m-k), \tag{4}$$

for $k = 0, \ldots, N-1$, where $a_m(n)$ denotes the expansion coefficient corresponding to the m-th orthogonal component $\psi(m-k)$. The expansion coefficients can be computed from $h_k(n)$ by

$$a_m(n) = \sum_{k=0}^{N-1} h_k(n) \psi(m-k) \tag{5}$$

for $m = 0, \ldots, N-1$.

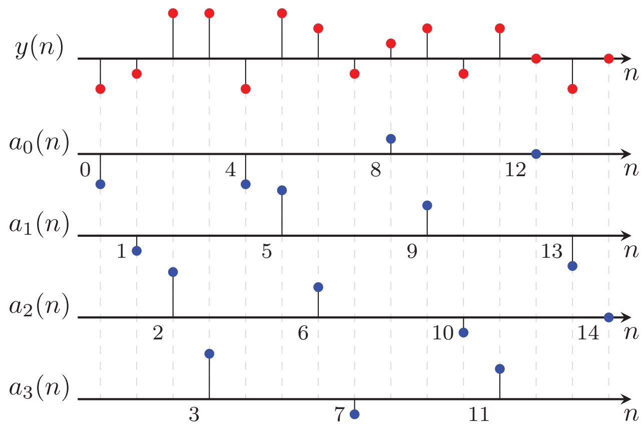


Figure 1: The relation of the captured signal $y(n)$ and the orthogonal expansion coefficients $a_m(n)$ for $N = 4$. See (6).

If the system is excited by $\psi(n)$, the output reads

$$\begin{aligned} y(n) &= \sum_{m=0}^{N-1} a_m(n) \sum_{k=0}^{N-1} \psi(m-k)\psi(n-k) \\ &= \sum_{m=0}^{N-1} a_m(n)\delta_{m\bar{n}} \\ &= a_{\bar{n}}(n) \end{aligned} \tag{6}$$

where (2) and (4) are exploited. Equation (6) states that the captured signal at time n corresponds to the $\bar{n}$-th orthogonal expansion coefficient of $h_k(n)$. This is illustrated in Fig. 1. For time-invariant systems, i.e. $h_k(n) = h_k$ and thus $y(n) = a_{\bar{n}}$, (4) constitutes the circular cross-correlation of $\psi(n)$ and $y(n)$,

$$h_k = \sum_{m=0}^{N-1} a_m \psi(m-k) = \sum_{m=0}^{N-1} y(m+\nu N)\psi(m-k) \tag{7}$$

which holds for $\forall \nu \in \mathbb{Z}$. If the system is piece-wise time-invariant for at least N consecutive samples, the instantaneous impulse response can be computed by (7). In general, however, the exact impulse responses cannot be obtained since only one orthogonal coefficient is observed for each $h_k(n)$. As shown in Fig. 1, the original coefficient $a_m(n)$ is decimated by a factor of N. Therefore, $a_m(n)$ can be recovered only if the bandwidth of $a_m(n)$ is less than $\frac{f_s}{N}$ with f_s denoting the sampling rate. In order to compute $h_k(n)$ for any n, the missing values of $a_m(n)$ have to be interpolated properly.

Normalized Least Mean Squre Algorithm

In this section, it is shown how the orthogonal expansion coefficients are interpolated by the NLMS algorithm. Here, the filter coefficients are adaptively updated as [10, Sec. 6.1.]

$$\hat{h}_k(n) = \hat{h}_k(n-1) + \frac{\mu\, \varepsilon(n)}{\sum_{k=0}^{N-1} |\psi(n-k)|^2} \psi(n-k), \tag{8}$$

for $k = 0, \ldots, N-1$, where $\hat{h}_k(n)$ denotes the estimate of $h_k(n)$ and $\mu > 0$ the step size (or adaptation constant).

The estimation error $\varepsilon(n)$ is the difference between the desired output $y(n)$ and its estimate $\hat{y}(n)$,

$$\varepsilon(n) = y(n) - \hat{y}(n) \tag{9}$$

where

$$\hat{y}(n) = \sum_{k=0}^{N-1} \hat{h}_k(n-1)\psi(n-k). \tag{10}$$

As in the previous section, $\sum_{k=0}^{N-1} |\psi(n-k)|^2 = 1$ is assumed, and thus (8) is simplified as

$$\hat{h}_k(n) = \hat{h}_k(n-1) + \mu\, \varepsilon(n)\psi(n-k). \tag{11}$$

Since $h_k(n)$ is estimated by using the current output $y(n)$, (11) constitutes an off-line processing which is different from the traditional NLMS [10, Eq. (7,10)]. In the context of this paper, the entire output $y(n)$ is first recorded and $\hat{h}_k(n)$ is computed by post processing. Real-time processing is thus not a requirement, and the above modification is acceptable.

The first thing to note from (11) is that only one orthogonal component is updated at a time. Therefore

$$\begin{aligned} \hat{h}_k(n) = &\underbrace{\sum_{m=0}^{N-1} \hat{a}_m(n-1)\psi(m-k)}_{\hat{h}_k(n-1)} \\ &+ \mu \underbrace{\left(a_{\bar{n}}(n) - \hat{a}_{\bar{n}}(n-1)\right)}_{\varepsilon(n)} \underbrace{\psi(n-k)}_{\psi(\bar{n}-k)} \end{aligned} \tag{12}$$

where $\hat{a}_m(n)$ denotes the estimate of $a_m(n)$. At time n, only the $\bar{n}$-th coefficient is updated,

$$\begin{aligned} \hat{a}_{\bar{n}}(n) &= \hat{a}_{\bar{n}}(n-1) + \mu\big(a_{\bar{n}}(n) - \hat{a}_{\bar{n}}(n-1)\big) \\ &= \mu a_{\bar{n}}(n) + (1-\mu)\hat{a}_{\bar{n}}(n-1). \end{aligned} \tag{13}$$

Once an orthogonal coefficient is updated, the value remains unchanged for the subsequent $N-1$ samples,

$$\hat{a}_{\bar{n}}(n+l) = \hat{a}_{\bar{n}}(n), \quad l = 0, \ldots, N-1. \tag{14}$$

The estimate $\hat{a}_{\bar{n}}(n)$ is thus a weighted sum of the decimated sequence of $a_m(n)$,

$$\hat{a}_{\bar{n}}(n+l) = \mu \sum_{\nu=0}^{\lfloor \frac{n}{N} \rfloor - 1} (1-\mu)^{\nu} a_{\bar{n}}(n-\nu N) \tag{15}$$

for $l = 0, \ldots, N-1$, where $\lfloor \cdot \rfloor$ denotes the floor operation. The initial estimate is assumed to $\hat{a}_{\bar{n}}(0) = 0$, for convenience. Equation (15) constitutes an interpolator where its FIR coefficients are given as

$$g_\nu = \mu(1-\mu)^{\nu}, \quad \nu = 0, \ldots, \lfloor \tfrac{n}{N} \rfloor - 1. \tag{16}$$

The filter coefficients are shown in Fig. 2 for different values of μ. The step-like behavior is attributed to the fact that $\hat{a}_m(n)$ is updated at every N-th sample. Although not shown here, this is also the case if a time-varying step size $\mu(n)$ is used.

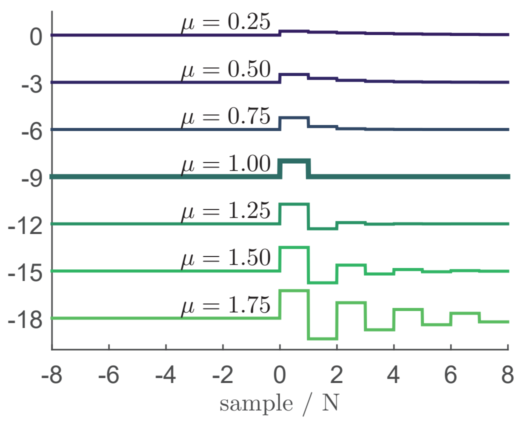


Figure 2: Equivalent interpolation filters for NLMS with different μ (see (15)). The abscissa is normalized by N.

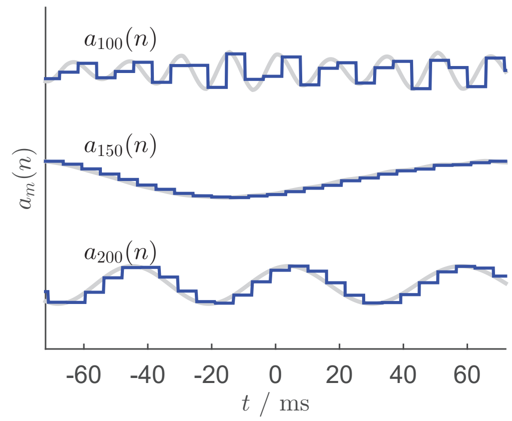


Figure 3: Estimates of the orthogonal expansion coefficients ($m = 100, 150, 200$) by using NLMS for $\mu = 1$ and $N = 256$. $\hat{a}_m(n)$ is indicated by blue lines whereas the true value $a_m(n)$ is indicated by gray lines.

The step size of $\mu = 1$ is a special case where the equivalent interpolation filter is a rectangular window, as shown in Fig. 2. In this case, the estimation error (9) reads

$$\varepsilon(n) = a_{\bar{n}}(n) - a_{\bar{n}}(n - N). \tag{17}$$

It corresponds to the $\bar{n}$-th orthogonal component of the accumulated system changes between $n - N$ and n. The resulting filter $\hat{h}_k(n)$ is identical to the circular convolution of $y(n)$ and $\psi(n)$ [12, 9]. As discussed in the previous section, this is useful only if the system is time-invariant.

Note that the filters in Fig. 2 exhibit discontinuities which results in abrupt changes of $a_m(n)$. Moreover, since NLMS exploits only the values from $y(0)$ to $y(n)$, the interpolation filter is asymmetric. The influence of these properties on time-varying system identification is examined in the following section.

Evaluation

A dynamic system was simulated where the impulse response is a time-varying delay

$$h(t, \tau) = \begin{cases} \delta(\tau + \tau_0), & t < -T \\ \delta\left(\tau - \tau_0 \sin(\frac{\pi}{2}\frac{\tau}{T})\right), & -T \le t \le T \\ \delta(\tau - \tau_0), & t > T, \end{cases} \tag{18}$$

where $\delta(t)$ denotes the Dirac delta function. For $t < -T$, the system is in a static state for a long time so that $\hat{h}_k(n)$ has converged to $h_k(n)$. Between $-T \le t \le T$, the time-of-arrival changes continuously from $-\tau_0$ to τ_0. For $t > T$, the system reaches a static state. The discrete-time impulse responses corresponding to (18) were simulated for $T = 1$ s, $\tau_0 = 1$ ms, and $f_s = 44.1$ kHz. The non-integer delays were implemented by using fractional delay filters of order 23 [13]. A perfect sequence with a period of $N = 256$ was used to excite the system. The instantaneous impulse responses were computed by the NLMS algorithm where the step size was set to $\mu = 1$ unless otherwise specified.

The orthogonal expansion coefficients $\hat{a}_m(n)$ were computed from $\hat{h}_k(n)$ by using (5). Three selected coefficients are shown in Fig. 3 along with the true values indicated by gray lines. As anticipated, $\hat{a}_m(n)$ suffers from discontinuities appearing at $n = m + \nu N, \nu \in \mathbb{Z}$. This is the case even for slowly changing coefficients like $\hat{a}_{150}(n)$. Due to the asymmetric interpolation, the tracking of $a_m(n)$ is lagged, which is observable for $\hat{a}_{100}(n)$ and $\hat{a}_{200}(n)$.

This behavior clearly influences the technical performance of system identification. The time-varying filter coefficients $h_k(n)$ and $\hat{h}_k(n)$ are shown in Fig. 4(a) and 4(b), respectively. For $t < -T$ and $t > T$, $\hat{h}_k(n)$ converges to $h_k(n)$. However, the performance is strongly degraded in the time-varying phase, $-T \le t \le T$. This is also evident in Fig. 5 where the normalized system distances,

$$D(n) = \frac{\sum_{k=0}^{N-1} |h_k(n) - \hat{h}_k(n)|^2}{\sum_{k=0}^{N-1} |h_k(n)|^2}, \tag{19}$$

are plotted for different μ. As is well known, NLMS with larger step size shows faster convergence but suffers from misalignment [10]. Around $t = 0$ where the changing rate of the system is at maximum, the performance does not differ significantly.

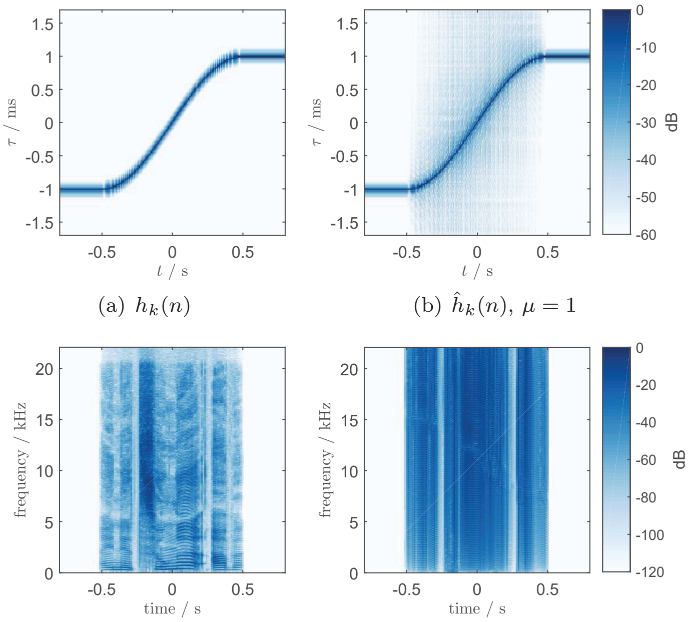
If the individual impulse responses are used to filter a source signal $s(n)$, which simulates a time-invariant system, the difference between $h_k(n)$ and $\hat{h}_k(n)$ is barely perceptible. However, if a dynamic scenario is simulated by time-varying filtering,

$$y_s(n) = \sum_{k=0}^{N-1} \hat{h}_k(n) s(n - k) \tag{20}$$

strong artifacts are audible for $\hat{h}_k(n)$. The transient response seems to be important from a perceptual point of view. The transient signal is defined as the additional component caused by the update of $\hat{h}_k(n)$,

$$y_{tr}(n) = \sum_{k=0}^{N-1} \Delta\hat{h}_k(n) s(n - k) \tag{21}$$

where $\Delta\hat{h}_k(n) = \hat{h}_k(n) - \hat{h}_k(n - 1)$. The spectrograms of $y_{tr}(n)$ and $\hat{y}_{tr}(n)$ are shown in Fig. 4(c) and 4(d),



(a) $h_k(n)$ (b) $\hat{h}_k(n)$, $\mu = 1$

(c) $y_{\mathrm{tr}}(n)$ (d) $\hat{y}_{\mathrm{tr}}(n)$, $\mu = 1$

Figure 4: The original and estimated impulse responses are shown in (a) and (b), respectively. The spectrograms of the resulting transient signals (see (21)) are shown in (d) and (e) for speech signal.

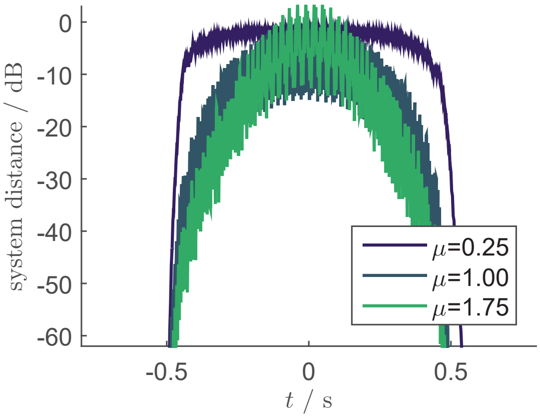


Figure 5: Normalized system distances (19) for different μ.

respectively. For the ideal case (Fig. 4(c)), the transient signal resembles the time-frequency structure of the source signal (speech in this case). For $\hat{h}_k(n)$, on the other hand, the transient signal has a much larger amplitude and the time-frequency structure is quite different from the source signal. In informal listening, it rather sounds like the perfect sequence $\psi(n)$ used in the measurement. This can be explained by the fact that $\Delta\hat{h}_k(n) = \mu\,\varepsilon(n)\psi(n-k)$ and thus

$$\hat{y}_{\mathrm{tr}}(n) = \mu\,\varepsilon(n) \sum_{k=0}^{N-1} \psi(n-k)s(n-k). \tag{22}$$

The summation on the right-hand side corresponds to the N-tap cross-correlation of $s(n)$ and $\psi(n)$. It is modulated by $\varepsilon(n)$ which characterizes the changing rate of the system.

Discussion and Conclusion

The NLMS algorithm was investigated in the context of off-line time-varying system identification. It was shown that the output of a system excited by a perfect sequence can be interpreted as a decimated version of the orthogonal expansion coefficients of the time-varying impulse response. NLMS turned out to inherently perform a coarse interpolation where the equivalent interpolation filter has asymmetric and step-like coefficients, as shown in Fig 2. NLMS should be used only if piece-wise time-invariance can be assured. Otherwise, it is suggested to employ linear or higher-order interpolation filters [9]. This is particularly the case if the impulse responses are intended to be used for the auralization of dynamic auditory scenes.

References

[1] E. M. Hulsebos, "Auralization using wave field synthesis," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, 2004.

[2] N. Hahn and S. Spors, "Continuous measurement of impulse responses on a circle using a uniformly moving microphone," in *Proc. of 23rd European Signal Processing Conference (EUSIPCO)*. Nice, France: IEEE, Aug. 2015, pp. 2536–2540.

[3] G. Enzner, C. Antweiler, and S. Spors, "Acquisition and representation of head-related transfer functions," in *The Technology of Binaural Listening*. Springer, 2013, pp. 57–92.

[4] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 135–143, 2005.

[5] R. Duraiswami, E. Grassi, N. A. Gumerov, Z. Li, D. N. Zotkin, and L. S. Davis, "High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues," in *Proc. of 119th Audio Engineering Society (AES) Convention*. Audio Engineering Society, 2005.

[6] N. Hahn and S. Spors, "Measurement of time-variant binaural room impulse responses for data-based synthesis of dynamic auditory scenes," in *Proc. of 40th German Annual Conference on Acoustics (DAGA)*, Oldenburg, Germany, Mar. 2014.

[7] T. Ajdler, L. Sbaiz, and M. Vetterli, "Dynamic measurement of room impulse responses using a moving microphone," *The Journal of the Acoustical Society of America (JASA)*, vol. 122, no. 3, pp. 1636–1645, 2007.

[8] C. Antweiler, "Multi-channel system identification with perfect sequences." Chichester, UK: John Wiley & Sons, 2008, pp. 171–198.

[9] N. Hahn and S. Spors, "Identification of dynamic acoustic systems by orthogonal expansion of time-variant impulse responses," in *Proc. of 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*. Athens, Greece: IEEE, May 2014, pp. 161–164.

[10] S. S. Haykin, *Adaptive Filter Theory*. Pearson, 2005.

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

[12] C. Antweiler, S. Kuehl, B. Sauert, and P. Vary, "System identification with perfect sequence excitation-efficient NLMS vs. inverse cyclic convolution," in *Proc. of 11th ITG Symposium: Speech Communication*. Erlangen, Germany: VDE, 2014, pp. 1–4.

[13] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30—-60, 1996.