IDENTIFICATION OF DYNAMIC ACOUSTIC SYSTEMS BY ORTHOGONAL EXPANSION OF TIME-VARIANT IMPULSE RESPONSES

Nara Hahn and Sascha Spors

Institute for Communications Engineering, University of Rostock,
Richard-Wagner-Strasse 31 (Haus 8), D-18119 Rostock
{nara.hahn,sascha.spors}@uni-rostock.de

ABSTRACT

We present a system identification method for dynamic acoustic systems. The system is excited by a periodic sequence the autocorrelation function of which is an impulse train. The circularly shifted versions of such sequence form an orthogonal basis for the corresponding vector space. As each orthogonal component of the system is sequentially excited, the output sample is the expansion coefficient of the instantaneous impulse response. Unfortunately, the expansion coefficients obtained from the output signal are undersampled, and thus, the time-varying impulse response cannot be fully determined. In this study, we estimate the missing expansion coefficients by interpolation. As a result, a full set of impulse responses is obtained that describes the history of the dynamic system. The performance of the proposed method is demonstrated by simulations and real measurements.

1. INTRODUCTION

Various system identification methods are employed to measure acoustic impulse responses, such as room impulse responses, head-related impulse responses (HRIRs), and binaural room impulse responses (BRIRs) [1]. Such measurements are used to simulate virtual acoustic scenes by filtering dry signals with the impulse responses. Although many acoustic scenes are time-dependent, existing system identification methods are more or less focused on static cases. The objective of this study is to continuously measure the impulse response, thereby fully identify the history of time-varying systems.

A series of research has been conducted with the same motivation [2, 3]. In [2], a massive number of head-related impulse responses were adaptively measured in a continuously changing setup. A so-called *perfect sequence*, which has the ideal correlation property, was used to excite the system [4]. Thanks to the relatively low changing rate and the anechoic environment, the system could be assumed to be piecewise time-invariant. In general, however, this is not the case in dynamic scenes.

In this study, we also use the perfect sequence as the excitation signal, but the impulse responses are computed in a different way. By exploiting the correlation property of perfect sequences, it is shown that the orthogonal components of the system are sequentially excited and its responses are temporally separated. Our focus is on estimating the orthogonal expansion coefficients. In this perspective, the relation between the correlation method and the least mean square algorithm is derived, and finally, an improved system identification method is presented.

2. LINEAR TIME-VARIANT SYSTEMS

2.1. System model

A linear time-varying system is represented by a finite impulse response (FIR) filter model,

$$y[n] = \sum_{k=0}^{N-1} h_k[n]x[n-k] = \mathbf{h}[n]^T\mathbf{x}[n], \tag{1}$$

where

$$\begin{aligned} \mathbf{h}[n] &= (h_0[n], h_1[n], \cdots, h_{N-1}[n])^T & (2)\\ \mathbf{x}[n] &= (x[n], x[n-1], \cdots, x[n-N+1])^T. & (3)\end{aligned}$$

The discrete-time sequence $y[n]$ is the output signal, $h_k[n]$ the time-dependent impulse response, $x[n]$ the input signal, and N the maximum length of the response of the system. It is assumed that the system is causal and noiseless.

2.2. Orthogonal expansion

If we choose a set of orthonormal basis vectors for $\mathbb{R}^N$

$$\begin{aligned} \mathcal{S} &= \{\mathbf{s}_0, \mathbf{s}_1, \cdots, \mathbf{s}_{N-1}\} & (4)\\ \text{with}\quad \mathbf{s}_i^T\mathbf{s}_j &= \delta_{ij} \quad (i, j = 0, \cdots, N-1), & (5)\end{aligned}$$

where δ_{ij} denotes the Kronecker delta, the vectors in (2) and (3) can be represented as the weighted sum of basis vectors

$$\mathbf{h}[n] = \sum_{\eta=0}^{N-1} a_\eta[n]\mathbf{s}_\eta = \mathbf{S}\mathbf{a}[n] \tag{6}$$

$$\mathbf{x}[n] = \sum_{\eta=0}^{N-1} b_\eta[n]\mathbf{s}_\eta = \mathbf{S}\mathbf{b}[n], \tag{7}$$

where

$$\begin{aligned} \mathbf{S} &= \begin{pmatrix}\mathbf{s}_0, & \mathbf{s}_1, & \cdots, & \mathbf{s}_{N-1}\end{pmatrix} & (8)\\ \mathbf{a}[n] &= (a_0[n], a_1[n], \cdots, a_{N-1}[n])^T & (9)\\ \mathbf{b}[n] &= (b_0[n], b_1[n], \cdots, b_{N-1}[n])^T. & (10)\end{aligned}$$

Due to (5), the transformation matrix $\mathbf{S}$ is an orthogonal matrix, i.e., $\mathbf{S}^T\mathbf{S} = \mathbf{S}\mathbf{S}^T = \mathbf{I}$. Thus, the expansion coefficients are

$$\begin{aligned} \mathbf{a}[n] &= \mathbf{S}^T\mathbf{h}[n] & (11)\\ \mathbf{b}[n] &= \mathbf{S}^T\mathbf{x}[n]. & (12)\end{aligned}$$

Substituting (6) and (7) into (1) gives

$$y[n] = \big(\mathbf{S}\mathbf{a}[n]\big)^T\big(\mathbf{S}\mathbf{b}[n]\big) = \mathbf{a}[n]^T \underbrace{\mathbf{S}^T\mathbf{S}}_{=\mathbf{I}}\mathbf{b}[n] \quad (13)$$

$$= \mathbf{a}[n]^T\mathbf{b}[n], \quad (14)$$

meaning that the output value of the system is the scalar product of the instantaneous expansion coefficients of $\mathbf{h}[n]$ and those of $\mathbf{x}[n]$. In the following section, a specific type of input signals and basis vectors are considered that simplifies (14).

3. EXCITATION SIGNALS WITH PERFECT AUTOCORRELATION

In system identification, sequences with rapidly decaying autocorrelation functions have gained great attention, and considerable efforts have been devoted to find such sequences [5]. In this context, a sequence with period M

$$\tilde{x}[n] = x_\nu \quad \text{where} \quad \nu \equiv n \bmod M \quad (15)$$

is referred to as a perfect sequence if its autocorrelation function is an impulse train, i.e.,

$$\varphi_{\tilde{x}\tilde{x}}[m] = \sum_{m=0}^{M-1}\tilde{x}[n]\tilde{x}[n+m] = E\sum_{\mu=-\infty}^{\infty}\delta[m+\mu M] \quad (16)$$

where $\delta[m]$ is the unit impulse function, and E is the energy of the sequence in one period, $E = \sum_{m=0}^{M-1}|\tilde{x}[m]|^2$. Without loss of generality, it is assumed that the sequence has normalized energy, $E = 1$. The period of the sequence is set equal to the maximum length of the impulse response, $M = N$, so that the impulse response can be measured without temporal aliasing [2].

If a system is excited with a perfect sequence, $x[n] = \tilde{x}[n]$, each entry of the input vector defined in (3) is also periodic,

$$\mathbf{x}[n] = (x_\nu, x_{\nu-1}, \cdots, x_0, x_{N-1}, \cdots, x_{\nu+1})^T \quad (17)$$

$$= \mathbf{x}[\nu] \equiv \mathbf{x}_\nu. \quad (18)$$

Equation (16) can be rewritten as $\mathbf{x}_i^T\mathbf{x}_j = \delta_{ij}(i,j = 0,\cdots,N-1)$, and thus, N consecutive input vectors form an orthonormal basis set for $\mathbb{R}^N$,

$$\mathbf{s}_0 = \mathbf{x}_0 = (x_0, x_{N-1}, \cdots, x_1)^T \quad (19)$$

$$\mathbf{s}_1 = \mathbf{x}_1 = (x_1, x_0, \cdots, x_2)^T \quad (20)$$

$$\vdots \quad (21)$$

$$\mathbf{s}_{N-1} = \mathbf{x}_{N-1} = (x_{N-1}, x_{N-2}, \cdots, x_0)^T. \quad (22)$$

Thus,

$$\mathbf{S} = \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{N-1} \\ x_{N-1} & x_0 & x_1 & \cdots & x_{N-2} \\ x_{N-2} & x_{N-1} & x_0 & \cdots & x_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & x_3 & \cdots & x_0 \end{pmatrix}. \quad (23)$$

Note that each column of $\mathbf{S}$ has circularly shifted entries of the previous column. Such matrices are called circulant matrices. Multiplication of a circulant matrix and a column vector is the circular convolution of the vector and the first column of the circulant matrix, $\mathbf{s}_0$ in this case , whereas the multiplication of the transpose of a circulant matrix and a column vector is the circular cross-correlation of the vector and the first column [6].

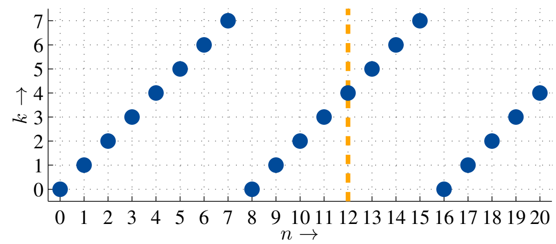


Fig. 1: An example showing the relation of the output samples (filled circles) and the orthogonal expansion coefficients. If a linear time-varying system is excited by a perfect sequence with the period of $N = 8$, then $y[12] = a_4[12]$. The other orthogonal components (grid points on the dashed line) are not excited.

Substituting (18) and (23) into (12) shows that the input vector has only one nonzero expansion coefficient, $\mathbf{b}[n] = \mathbf{S}^T\mathbf{s}_\nu = \mathbf{e}_{\nu+1}$, where $\mathbf{e}_i$ is the standard orthonormal basis vector that has only one nonzero entry, $\{\mathbf{e}_i\}_j = \delta_{ij}(i,j = 1,\cdots,N)$. Thus, from (14), the output sample of the system is

$$y[n] = \mathbf{a}[n]^T\mathbf{e}_{\nu+1} = a_\nu[n]. \quad (24)$$

In words, if a linear system is excited by a perfect sequence, the orthogonal components are sequentially excited, and thus, the output sequences are the expansion coefficients of the instantaneous impulse responses. An example is illustrated in Fig. 1, with $N = 8$ and $n = 12$. The filled circles indicate the output samples that are the exact values of the expansion coefficients. To fully identify $\mathbf{h}[n]$, N coefficients are required, that are the grid points on the dashed line. The available information is not sufficient as only one coefficient per sample is accessible. This problem is tackled in the next section where the missing expansion coefficients are estimated.

4. SYSTEM IDENTIFICATION

In this section, the problem of system identification is addressed with the focus on the estimation of the orthogonal expansion coefficients. We begin with the linear time-invariant case, and show that the solution coincides with the correlation method. The same approach is then applied to linear time-varying systems, and its limitations are discussed, which will be followed by our proposed method.

4.1. Linear time-invariant systems

The impulse response of a time-invariant system has constant impulse response coefficients, $\mathbf{h}[n] = \mathbf{h}$ and thus, $\mathbf{a}[n] = \mathbf{a}$. The true values of the latter can be directly obtained from N consecutive output values, $\mathbf{a} = (y[n-\nu],\cdots,y[n],y[n-N+1],\cdots,y[n-\nu-1])^T$. We define two vectors that have these values as its entries,

$$\mathbf{y}[n] = (y[n], y[n-1], \cdots, y[n-N+1])^T \quad (25)$$

$$\breve{\mathbf{y}}[n] = (y[n], y[n-N+1], \cdots, y[n-1])^T. \quad (26)$$

The exact expansion coefficient vector is then

$$\hat{\mathbf{a}} = \mathbf{a} = \mathbf{P}^\nu\breve{\mathbf{y}}[n] \quad (27)$$

where $\mathbf{P}$ is the cyclic permutation matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{28}$$

that circularly shifts the entries of a column vector by 1. Substituting (27) into (6) gives

$$\mathbf{h} = \mathbf{S}\big(\mathbf{P}^{\nu}\check{\mathbf{y}}[n]\big) = \big(\mathbf{S}\mathbf{P}^{\nu}\big)\check{\mathbf{y}}[n]. \tag{29}$$

The matrix $\mathbf{SP}^{\nu}$ is a circulant matrix where the first column is $\mathbf{s}_{\nu} = \mathbf{x}[n]$. Therefore, $\mathbf{h}$ in (29) is the circular convolution of $\mathbf{x}[n]$ and $\check{\mathbf{y}}[n]$, which is equivalent to the circular cross-correlation of $\mathbf{x}[n]$ and $\mathbf{y}[n]$. Although we began the derivation with the focus on the expansion coefficients, we ended up with the well-known correlation method [7].

4.2. Linear time-varying systems

Under the assumption that the system is piecewise time-invariant, the correlation method is often used for time-varying systems. The estimated instantaneous impulse response $\hat{\mathbf{h}}[n]$ is then $\hat{\mathbf{h}}[n] = \big(\mathbf{SP}^{\nu}\big)\big(\check{\mathbf{y}}[n]\big)$. Once an output sample $y[n]$ is captured, the vector $\check{\mathbf{y}}[n-1]$ is replaced by $\check{\mathbf{y}}[n]$, meaning that only $\hat{a}_{\nu}[n]$ is updated and the other $N-1$ orthogonal components remain unchanged,

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + \big(\hat{a}_{\nu}[n] - \hat{a}_{\nu}[n-1]\big)\mathbf{s}_{\nu}. \tag{30}$$

Recall that $y[n] = a_{\nu}[n]$. Prior to the update, the corresponding expansion coefficient $\hat{a}_{\nu}[n-1]$ can be interpreted as the estimate of the output $\hat{y}[n]$. Thus,

$$\hat{\mathbf{h}}[n] = \hat{\mathbf{h}}[n-1] + e[n]\mathbf{x}[n] \tag{31}$$

where $e[n] = y[n] - \hat{y}[n]$ is the estimation error. We have used the fact that $\mathbf{s}_{\nu} = \mathbf{x}[n]$. Interestingly, (31) coincides with the least mean square (LMS) algorithm with unit step size [8]. This agrees with the observation in [9], where the impulse responses are measured with the normalized LMS algorithm in combination with a perfect sequence. It is reported that the step size of 1 shows the best performance for noiseless systems.

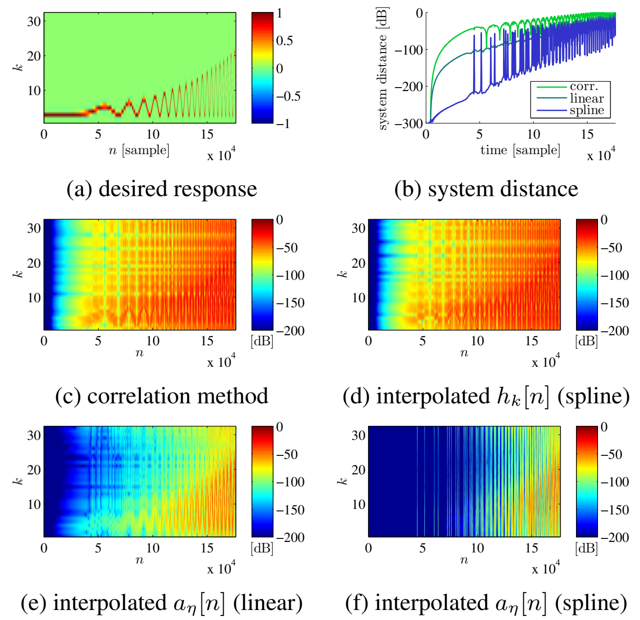
Note that this approach updates each orthogonal expansion coefficient once per N samples. Once an orthogonal component is updated, the value remains for $N-1$ samples until the next update. This results in an abrupt change of each coefficient. This is a well-known property of LMS-based algorithms used in combination with perfect sequences [10].

4.3. Interpolation of the expansion coefficient

As pointed out in Sec. 3, only one true value out of N expansion coefficients can be observed. The remaining parts of the impulse response are unknown, and thus, have to be estimated. In this study, each coefficient is estimated from the sequence of the same coefficient,

$$\hat{a}_{\nu}[n] = f(\mathcal{Y}_{\nu}), \tag{32}$$

where $f(\cdot)$ is an interpolation function and $\mathcal{Y}_{\nu} = \{y[\nu + mN], m \in \mathbb{Z}\}$. In this context, the correlation method corresponds to a piecewise constant interpolator, the simplest but poorest one. It is expected that the system identification performance can be improved when using higher order interpolation methods, e.g., linear interpolation or spline interpolation. The interpolator has to be chosen depending on the requirements, constraints, and time-varying properties of the system of interest.



(a) desired response (b) system distance

(c) correlation method (d) interpolated $h_k[n]$ (spline)

(e) interpolated $a_{\eta}[n]$ (linear) (f) interpolated $a_{\eta}[n]$ (spline)

Fig. 2: Simulation results. The impulse response of the dynamic system is shown in (a). The performances of the employed methods are shown in (b) in terms of system distance, $20\log_{10}\frac{\|\mathbf{h}[n]-\hat{\mathbf{h}}[n]\|}{\|\mathbf{h}[n]\|}$. The filter coefficient errors $|h_k[n] - \hat{h}_k[n]|$ are in (c)-(f) for different methods. Results in (d) is not included in (b).

Each expansion coefficient is updated with a frequency of $\frac{f_s}{N}$, where f_s is the sampling frequency of the discrete-time system. For perfect reconstruction, the bandwidth of each expansion coefficient has to be less than $\frac{f_s}{N}$ [11]. Otherwise, the estimated expansion coefficients suffer from aliasing artifacts. It can be deduced that the accuracy of the estimated impulse response depends on the length of the system's response and on the changing rate of the system. The effect of aliasing artifacts on the estimated impulse responses and its perception is not clear so far.

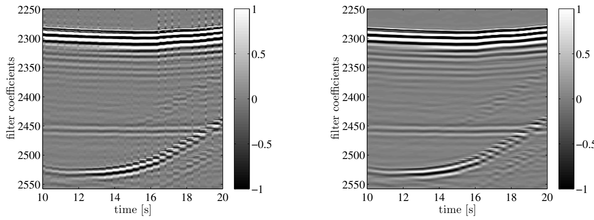
Finally, it is worth noting that the interpolation of expansion coefficients is by no means equivalent to the interpolation of impulse response coefficients. The latter ignores (24) which states that the expansion coefficients belong to different impulse responses.

5. RESULTS

To demonstrate improvement by interpolating the expansion coefficients, the identification of a dynamic virtual system and a real system is considered. The systems are excited by *perfect sweeps* [12], a type of perfect sequence that was reported to be more immune to nonlinearity of the system [13]. The sampling frequency is $f_s = 44.1$kHz.

5.1. Simulation

The output of a virtual system is simulated where the impulse response is a time-varying delay. The desired impulse response is shown in Fig. 2a, where each vertical slice of the plot show the im-



(a) correlation method (b) linear interpolation

Fig. 3: BRIRs (left ear) of a moving sound source, obtained by (a) the correlation method and (b) interpolated expansion coefficients.

pulse response coefficients. In the beginning, $n \leq 4096$, the delay remains constant. Then it begins to vary sinusoidally with increasing frequency, thus, the changing rate of the system increases with time. The fractional part of the delay is implemented by a 4th-order *Largrangian* interpolator [14]. The period of the perfect sweep was 32 samples. As proposed in Sec. 4.3, the expansion coefficients were interpolated and used to compute $\hat{\mathbf{h}}[n]$ using (6). Two well-known methods, linear interpolation and cubic spline interpolation were used. For comparison, results using the correlation method and interpolated impulse response coefficients are also shown.

The estimation errors of impulse response coefficients are shown in Fig. 2c–2f. All methods work very well in the static phase. As soon as the system varies with time, the performance of all are degraded. The errors are somewhat concentrated on the nonzero coefficients. As can be seen in Fig. 2c, the correlation method is very vulnerable to system variance. Figure 2d show that using interpolated filter coefficients is not so helpful. On the contrary, the benefits of using the interpolated expansion coefficients are clearly visible in Fig. 2e and 2f. The system distance of the employed methods is shown in Fig. 2b. Although spline interpolation is superior to linear interpolation, both methods become similar when the system is extremely time-variant. Even though, both methods still outperform the correlation method.

Each set of impulse responses was used to filter a speech signal to generate a simulated output signal. In informal listening, chirp-like artifacts were audible when the correlation method was used, whereas the signals using the proposed methods are indistinguishable from the desired signal. Interestingly, despite its poor system distance shown in Fig. 2d, the signal using interpolated impulse response coefficients has comparable quality with the proposed methods. The underlying perceptual mechanism is not clear at this stage.

5.2. Dynamic BRIRs

The same experiment was performed for a real acoustic system. A loudspeaker was placed on a rolling table and was moved in a room in the Institute of Communications Engineering, University of Rostock. The period of the perfect sweep was $N = 2^{13}$.The reproduced sound was captured with a dummy head.

The extracted BRIRs for the left ear are shown in Fig. 3. Only the direct sound and a couple of early reflections are shown. In Fig. 3a, where the correlation method was used, the impulse response is contaminated by artifacts that are widely spread over all coefficients. It is attributed to the fact that each orthogonal component is updated only once per N samples which causes an abrupt change in the FIR coefficients. It is likely that such artifacts cause pre-echo when used for syntheses. Apparently, this problem is resolved when the expansion coefficients are linearly interpolated (See Fig. 3b). In this case, all orthogonal components are updated in a sample-by-sample fashion. Informal listening to the synthesized audio signal also confirms the improvement.

6. CONCLUSION

The problem of dynamic system identification is addressed in which the systems are excited with perfect sequences. By exploiting the specific autocorrelation property of the excitation signal, it was shown that the output of the system is the orthogonal expansion coefficient of the instantaneous impulse responses. With the aid of this interpretation, the proposed method estimates the expansion coefficients by interpolation and computes the impulse responses in a sample-by-sample fashion. Only the general framework is presented in this paper, and thus, the interpolation function in 32 can be chosen that is best suited for the application. The proposed method is compared with the correlation method in simulations and real measurements. It was shown that a considerable improvement can be achieved even with linear interpolation. The results suggest that the presented method can be used for the data-based auralization of dynamic acoustic scenes.

It is yet not clear how to quantify the time-dependent variability of a dynamic system. The physical and perceptual effects of the aliased expansion coefficients on the reconstructed impulse responses need to be examined.

7. REFERENCES

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

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

[3] 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*, vol. 122, pp. 1636, 2007.

[4] H. D. Lüke, *Korrelationssignale*, Springer, 1992.

[5] H. D. Lüke, "Sequences and arrays with perfect periodic correlation," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 287–294, 1988.

[6] P. J. Davis, "Circulant matrices," *New York*, 1979.

[7] R. Isermann, *Identifikation dynamischer Systeme*, Springer-Verlag, 1992.

[8] S. S. Haykin, *Adaptive Filter Theory*, Pearson Education India, 2005.

[9] C. Antweiler, A. Telle, and P. Vary, "NLMS-type system identification of MISO systems with shifted perfect sequences," *Proc. IWAENC*, 2008.

[10] C. Antweiler and M. Dörbecker, "Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control," in *Annales des Telecommunications*. 1994, vol. 49, pp. 386–397, Springer.

[11] B. Girod, R. Rabenstein, and A. Stenger, *Signals and systems*, Wiley New York, 2001.

[12] N. Aoshima, "Computer-generated pulse signal applied for sound measurement," *The Journal of the Acoustical Society of America*, vol. 69, pp. 1484, 1981.

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

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