NONLINEAR ACOUSTIC ECHO CANCELLATION USING ADAPTIVE ORTHOGONALIZED POWER FILTERS

Fabian Kuech, Andreas Mitnacht, and Walter Kellermann

Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg
Cauerstr. 7, D-91058 Erlangen, Germany
{kuech,wk}@LNT.de

ABSTRACT

In acoustic echo cancellation as, e.g., for mobile communication receivers, loudspeakers and their amplifiers cause significant nonlinear distortion in the echo path, resulting in a degradation of the performance of linear echo cancelers. In order to cope with this type of nonlinear echo paths, we discuss a orthogonalized version power filters that can be considered as a parallelized realization of the cascade of a memoryless polynomial followed by a linear filter. As, in the echo cancellation context, the statistics of the speech input are non-stationary and not known in advance, the orthogonalization follows the signal statistics. The performance of the resulting novel nonlinear structure is evaluated by experiments using real hardware.

1. INTRODUCTION

In the acoustic echo cancellation problem, illustrated in Fig. 1, the acoustic echo canceler (AEC) seeks to minimize contribution of the echo signal $r(k)$ to the power of the error signal $e(n)$ by subtracting an estimate of the echo signal $y(n)$ from the microphone signal $d(n)$. The performance of standard approaches for the cancellation of acoustic echos in telecommunication systems strongly depends on the assumption of a linear echo path. However, in applications such as echo cancellation for mobile communication receivers, non-negligible nonlinear distortion is introduced by loudspeakers and their amplifiers [1] [2]. With these nonlinear distortions, purely linear approaches are not able to provide a sufficient echo attenuation, making nonlinear approaches desirable.

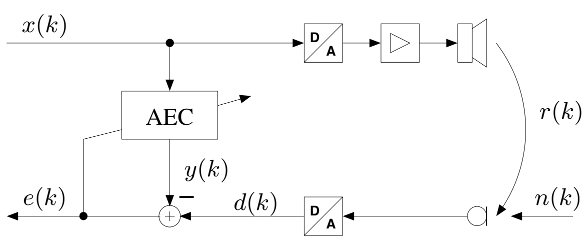

Fig. 1. General set-up of the acoustic echo cancellation problem.

In this contribution we present power filters as a practical parallelized model of the nonlinear echo path that can be considered as a memoryless saturation characteristic, approximating the nonlinear behaviour of loudspeaker and amplifier, followed by a linear room impulse response. The mapping of such a cascaded nonlinear structure to power filters is illustrated in Fig. 2, and is discussed in Section 2.1. In order to achieve a faster convergence of the adaptive implementation, an equivalent orthogonal structure is introduced. Orthogonalized power filters in complex baseband representation have already been applied in [3] for the predistortion of nonlinear power amplifiers with memory. The orthogonalized power filters presented in [3] are designed under the assumption of stationary, uniformly distributed inputs. However, besides the fact that the input signal distributions are in general not accessible in advance, the non-stationarity of the speech input has to be taken into account when considering the echo cancellation application. Therefore, the orthogonalization is performed in a time-variant manner. Simulations based on real measured data confirm that a significant increase in echo attenuation can be achieved with the proposed structure if the loudspeaker system introduces saturation-type nonlinearities in the echo path.

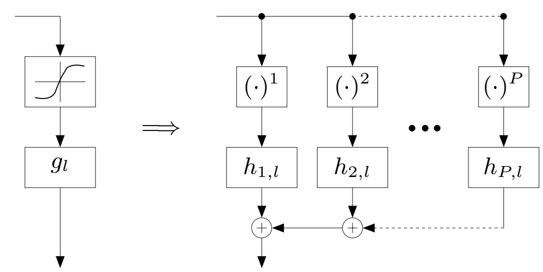

Fig. 2. Parallelized implementation of a memoryless nonlinearity followed by a linear filter (right) with power filters (left).

2. POWER FILTERS

In this section we discuss a certain class of polynomial systems with memory, i.e., the so-called power filters. A P-th order power filter, depicted on the left hand side in Fig. 2, is defined by its input/output relation as follows

$$y(k) = \sum_{p=1}^{P} \sum_{l=0}^{N_p-1} h_{p,l}\, x^p(k-l). \tag{1}$$

From (1) we notice that power filters can be considered as linear multiple input/single output systems, where the input of the p-th channel is given by the p-th power of $x(k)$. The input of each channel is then filtered by an associated linear filter $h_{p,l}$ with memory length N_p. For compactness, we write (1) in matrix notation:

$$y(k) = \sum_{p=1}^{P} \mathbf{h}_p^T \mathbf{x}_p(k), \tag{2}$$

with the vectors

$$\begin{aligned} \mathbf{x}_p(k) &= [x^p(k),\, x^p(k-1), \ldots,\, x^p(k-N_p+1)]^T, & (3) \\ \mathbf{h}_p &= \left[h_{p,0},\, h_{p,1}, \ldots,\, h_{p,N_p-1}\right]^T. & (4) \end{aligned}$$

Note that the input signals of each channel are in general not mutually orthogonal, i.e., $\mathrm{E}\{x^i(k)x^j(k)\} \neq 0$. Thus, a direct adaptive implementation of the non-orthogonalized power filter suffers from slow convergence, as verified in Section 5.

2.1. Model of the nonlinear echo path

For modeling the echo path of mobile communication receivers, mainly two sources of nonlinearities have to be considered. On the one hand, amplifiers show a memoryless saturation characteristic, e.g., due to low battery voltage [1], and, on the other hand, small loudspeakers driven at high volume cause non-negligible nonlinear distortion, too. While the nonlinear behaviour of common electrodynamic loudspeakers can be modeled by Volterra filters [4], i.e., a nonlinearity with memory, miniaturized loudspeakers, e.g., used in mobile phones, exhibit a saturation-type nonlinearity. As the propagation path between loudspeaker and microphone, including the microphone, can be modeled by a linear filter, the overall structure of the echo path consists of the cascade of a memoryless nonlinearity and a linear filter, as depicted on the right hand side of Fig. 2. In the following we assume that the memoryless nonlinearity can be approximated by a truncated Taylor series expansion, i.e., we exclude nonlinearities such as hard-clipping characteristics. Then, the output of the memoryless nonlinearity can be written as

$$s(k) = \sum_{p=1}^{P} a_p x^p(k), \tag{5}$$

where a_p denotes the coefficients of the Taylor series expansion. The output of the subsequent linear filter with coefficients g_l, with $s(k)$ as input, reads

$$y(k) = \sum_{l=0}^{N_g-1} g_l\, s(k-l). \tag{6}$$

Introducing (5) into (6), and reversing the order of the summation yields the mapping of the cascaded structure to the parallel structure, illustrated in Fig. 2:

$$y(k) = \sum_{p=1}^{P} \sum_{l=0}^{N_g-1} a_p g_l x^p(k-l). \tag{7}$$

Comparing (7) and (1), the power filter model of the considered cascaded structure is immediately obtained by equating

$$h_{p,l} = a_p g_l. \tag{8}$$

Note, however, that the number of parameters is increased from $P + N_g$ for the cascaded structure to PN_g for the parallel structure. On the one hand, this increase of the number of parameters may disqualify the parallelized implementation, but, on the other hand, for an adaptive realization of cascaded structures, as proposed in [1], [2], it is often challenging to assure convergence to the optimum solution or even assure a stable adaptation behaviour.

3. ORTHOGONALIZATION OF THE INPUT SIGNALS

Referring to the multi-channel interpretation of power filters, we recall that the input signals of the different channels, i.e., $x(k)$, $x^2(k), \ldots, x^P(k)$ are in general correlated. This implies that the convergence speed of a respective adaptive implementation is rather slow compared to a corresponding orthogonalized version. Therefore, we introduce a new set of mutually orthogonal input signals:

$$\begin{aligned} x_{\mathrm{o},1}(k) &= x(k), & (9) \\ x_{\mathrm{o},p}(k) &= x^p(k) + \sum_{i=1}^{p-1} q_{p,i}\, x^i(k), & (10) \end{aligned}$$

for $1 < p \leq P$. The orthogonalization coefficients $q_{p,i}$ are chosen such that

$$\mathrm{E}\{x_{\mathrm{o},i}(k)\, x_{\mathrm{o},j}(k)\} = 0, \quad \text{for } i \neq j. \tag{11}$$

The orthogonalization coefficients $q_{p,i}$ can be determined using the Gram-Schmidt orthogonalization method [5], i.e., the $p-1$ coefficients $q_{p,i}$ for the p-th order channel are obtained by solving

$$\begin{bmatrix} m_x^{(2)} & \ldots & m_x^{(p)} \\ \vdots & \ddots & \vdots \\ m_x^{(p)} & \ldots & m_x^{(2p-2)} \end{bmatrix} \begin{bmatrix} q_{p,1} \\ \vdots \\ q_{p,p-1} \end{bmatrix} = - \begin{bmatrix} m_x^{(p+1)} \\ \vdots \\ m_x^{(2p-1)} \end{bmatrix}, \tag{12}$$

where $m_x^{(n)} = \mathrm{E}\{x^n(k)\}$ denotes the n-th order moment of $x(k)$. Obviously, the orthogonalization coefficients $q_{p,i}$ are constant for stationary input $x(k)$. However, in practical implementations, $m_x^{(n)}$ has to be replaced by estimates $\widehat{m}_x^{(n)}(k)$ that can be obtained, e.g., by a first order recursion according to

$$\widehat{m}_x^{(n)}(k) = \lambda \widehat{m}_x^{(n)}(k-1) + (1-\lambda) x^n(k). \tag{13}$$

The forgetting factor λ can be adjusted in order to adapt the estimation to the statistics of the excitation signal $x(k)$. Obviously, the orthogonalization coefficients $q_{p,i}(k)$, resulting from (12) by replacing $m_x^{(n)}$ by $\widehat{m}_x^{(n)}(k)$, also depend on time. Analogously to (3), we introduce the orthogonalized signal vectors

$$\mathbf{x}_{\mathrm{o},p}(k) = [x_{\mathrm{o},p}(k), \ldots, x_{\mathrm{o},p}(k-N_p+1)]^T, \tag{14}$$

where for presentational convenience $N_p = N, \forall p$, is assumed in the following. Regarding (9), (10), the signal vectors (14) can be expressed by

$$\begin{aligned} \mathbf{x}_{\mathrm{o},1}(k) &= \mathbf{x}_1(k), & (15) \\ \mathbf{x}_{\mathrm{o},p}(k) &= \mathbf{x}_p(k) + \sum_{i=1}^{p-1} \mathbf{Q}_{p,i}(k)\, \mathbf{x}_i(k), & (16) \end{aligned}$$

for $1 < p \leq P$. Here, $\mathbf{Q}_{p,i}(k)$ represent the diagonal orthogonalization matrices accounting for the time variance of the orthogonalization coefficients $q_{p,i}(k)$:

$$\mathbf{Q}_{p,i}(k) = \mathrm{diag}\left\{[q_{p,i}(k), \ldots, q_{p,i}(k-N+1)]\right\}. \tag{17}$$

If $x(k)$ is a white noise process, we obtain a orthogonality property of the signal vectors $\mathbf{x}_{\mathrm{o},p}(k)$ that corresponds to (11), i.e.,

$$\mathrm{E}\left\{\mathbf{x}_{\mathrm{o},i}(k)\mathbf{x}_{\mathrm{o},j}(k)^T\right\} = \mathbf{0}, \quad \text{for } i \neq j. \tag{18}$$

Note that (18) does in general not hold for correlated $x(k)$, although it can easily be verified that (18) is still valid for $P = 3$, if $x(k)$ is a zero-mean, correlated Gaussian process. For non-Gaussian processes, (18) does not hold in theory, however, for typical speech signals $x(k)$ it can be assumed to be approximately fulfilled. To give an example, we consider the case that $x(k)$ represents a zero-mean, first-order Laplacian Markov process with an autocorrelation function $r_{xx}(n) = \mathrm{E}\{x(k)x(k-n)\} = 0.9^{|n|}$. In Fig. 3 the normalized crosscorrelation function

$$c_{13}(n) = \frac{\mathrm{E}\{x(k)x^3(k-n)\}}{\sqrt{\mathrm{E}\{x^2(k)\}\,\mathrm{E}\{x^6(k)\}}} \tag{19}$$

between $x(k)$ and $x^3(k)$ is shown together with the normalized crosscorrelation function

$$c_{\mathrm{o},13}(n) = \frac{\mathrm{E}\{x_{\mathrm{o},1}(k)x_{\mathrm{o},3}(k-n)\}}{\sqrt{\mathrm{E}\{x^2(k)\}\,\mathrm{E}\{x_{\mathrm{o},3}^2(k)\}}} \tag{20}$$

between $x_{\mathrm{o},1}(k) = x(k)$ and $x_{\mathrm{o},3}(k)$ according to (10). From

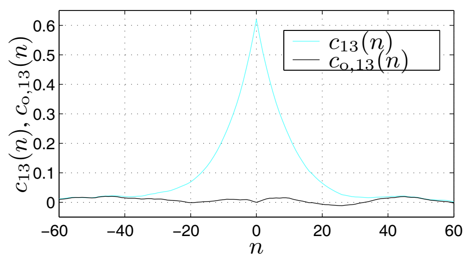

Fig. 3. Normalized crosscorrelation functions $c_{13}(n)$ and $c_{\mathrm{o},13}(n)$ between $x(k)$, $x^3(k)$ and $x(k)$, $x_{\mathrm{o},3}(k)$, respectively.

Fig. 3 we conclude that (18) is also valid in this case. Thus, we assume in the following that (18) at least approximately holds also for speech signals $x(k)$.

4. EQUIVALENT ORTHOGONALIZED STRUCTURE

In this section we discuss the so-called equivalent orthogonal structure (EOS) of the power filter which results from using the orthogonalized input signals instead of the original input signals of each channel.

The output of the power filter according to (2) using the orthogonalized signal vectors can be expressed by

$$y(k) = \sum_{p=1}^{P} \mathbf{h}_{\mathrm{o},p}(k)^T \mathbf{x}_{\mathrm{o},p}(k). \tag{21}$$

Introducing the definition of the orthogonalized input vectors (15), (16) into (21) leads to the EOS of the corresponding power filter. It is straightforward to verify that the relation between the original coefficients $\mathbf{h}_p$ and the filter coefficients $\mathbf{h}_{\mathrm{o},p}$ of the EOS is given by

$$\mathbf{h}_{\mathrm{o},P} = \mathbf{h}_P, \tag{22}$$

$$\mathbf{h}_{\mathrm{o},p}(k) = \mathbf{h}_p - \sum_{i=p+1}^{P} \mathbf{Q}_{i,p}(k)\mathbf{h}_{\mathrm{o},i}(k), \tag{23}$$

for $1 \le p < P$. Interestingly, (23) implies that for time varying orthogonalization matrices $\mathbf{Q}_{i,p}(k)$, the coefficients of the EOS $\mathbf{h}_{\mathrm{o},p}(k)$ also have to be time-variant, although (or because) the coefficients $\mathbf{h}_p$ may be constant in time.

4.1. Bias correction for time-variant orthogonalization

Next, we consider the influence of time-varying orthogonalization matrices $\mathbf{Q}_{i,p}(k)$ on the EOS. For the following discussion we assume that the physical echo path to be modeled corresponds to a P-th order power filter, where all linear filters $\mathbf{c}_p$ have memory length N. Using the notation of Fig. 1, the desired signal $d(k)$ can then be expressed by

$$d(k) = \sum_{p=1}^{P} \mathbf{c}_p^T \mathbf{x}_p(k) + n(k), \tag{24}$$

where $n(k)$ denotes a distortion that is zero-mean and independent of the input signal $x(k)$. Replacing the input vectors $\mathbf{x}_p(k)$ in (24) by the orthogonalized input vectors $\mathbf{x}_{\mathrm{o},p}(k)$, used for the computation of the output $y(k)$ of the EOS of the echo canceller according to (21), we obtain the EOS that corresponds to the original filter coefficients $\mathbf{c}_p(k)$ of the unknown echo path:

$$d(k) = \sum_{p=1}^{P} \mathbf{c}_{\mathrm{o},p}(k)^T \mathbf{x}_{\mathrm{o},p}(k) + n(k). \tag{25}$$

The Wiener solution for $\mathbf{h}_{\mathrm{o},p}(k)$ can be found by applying the so-called orthogonality theorem [6], i.e., for the optimum filter coefficients $\mathbf{h}_{\mathrm{o},p}(k) = \mathbf{h}_{\mathrm{o},p,\mathrm{opt}}(k)$,

$$\mathrm{E}\{e(k)\mathbf{x}_{\mathrm{o},p}(k)\}\,\big|_{\mathbf{h}_{\mathrm{o},p}(k)=\mathbf{h}_{\mathrm{o},p,\mathrm{opt}}(k)} = \mathbf{0} \tag{26}$$

holds. With (18), and regarding that $\mathrm{E}\{n(k)\mathbf{x}_{\mathrm{o},p}(k)\} = \mathbf{0}$, it follows from (26) that for white noise input, the optimum filter coefficients of the EOS are given by

$$\mathbf{h}_{\mathrm{o},p,\mathrm{opt}}(k) = \mathbf{c}_{\mathrm{o},p}(k), \tag{27}$$

As it follows from (23), the coefficients $\mathbf{c}_{\mathrm{o},p}(k)$ are time-varying for $p < P$, even for time-invariant $\mathbf{c}_p$ and, thus, the optimum solution for $\mathbf{h}_{\mathrm{o},p}(k)$ is also changing in time. In order to account for fluctuations of $\mathbf{h}_{\mathrm{o},p}(k)$ due to time-varying orthogonalization matrices $\mathbf{Q}_{i,p}(k)$, it is important to readjust the filter coefficients $\mathbf{h}_{\mathrm{o},p}(k)$ after each change of the orthogonalization matrices $\mathbf{Q}_{i,p}(k)$. Regarding (22) and (23), it follows that a bias correction is only required for the channels of order $p < P$, and can be performed, starting with $p = P-1$, by recursively computing

$$\mathbf{h}_{\mathrm{o},p}(k) = \mathbf{h}_{\mathrm{o},p}(k-1) + \sum_{i=p+1}^{P} \left[\mathbf{Q}_{i,p}(k-1)\mathbf{h}_{\mathrm{o},i}(k-1) - \mathbf{Q}_{i,p}(k)\mathbf{h}_{\mathrm{o},i}(k)\right], \tag{28}$$

for $1 \le p < P$. The importance of the above bias correction is pointed out by a simulation presented in Section 5.

4.2. Adaptation of the equivalent orthogonalized structure

Applying standard gradient descent techniques [6], it directly follows from (21) that the normalized least mean square (NLMS) update of the coefficients of the EOS is given by

$$\mathbf{h}_{\mathrm{o},p}(k+1) = \mathbf{h}_{\mathrm{o},p}(k) + \alpha_p \frac{\mathbf{x}_{\mathrm{o},p}(k)e(k)}{\mathbf{x}_{\mathrm{o},p}(k)^T\mathbf{x}_{\mathrm{o},p}(k)}, \tag{29}$$

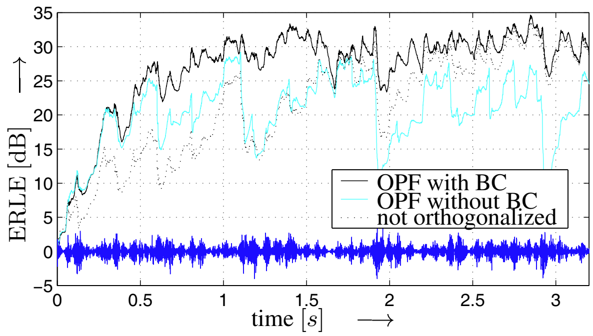

Fig. 4. Comparison of different realizations of adaptive power filters for a nonlinear AEC scenario together with the input signal.

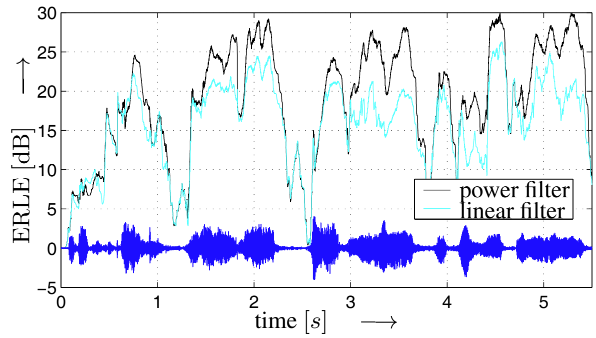

Fig. 5. Comparison of the adaptive orthogonalized power filter with bias correction and a linear AEC for recorded speech data.

where the normalization of the step-size α_p can be performed separately for each channel, as the input vectors $\mathbf{x}_{o,p}(k)$ are mutually orthogonal according to (18). Note, however, that the error introduced by a misadjusted filter in the i-th channel of the EOS acts as a distortion for the adaptation of the filters in the other channels and vice versa, rendering control of the adaptation a crucial issue. This topic is subject to current research and is not discussed here. It should be emphasized that the update (29) is performed with respect to the bias corrected filter coefficients $\mathbf{h}_{o,p}(k)$, i.e., the bias correction according to (28) is carried out first, and then the EOS of the power filter is adapted subsequently by applying (29).

5. SIMULATION RESULTS

To evaluate the performance of the proposed adaptive orthogonalized power filter, we present simulation results obtained for nonlinear acoustic echo cancellation. In the first experiment, the unknown echo path has been modeled by the cascade of a third-order memoryless polynomial, followed by a linear filter of length $N = 200$. The adaptive echo canceler has been realized as a third-order power filter, where the memory length of each channel is also $N_p = 200$, for $p \in \{1, 2, 3\}$. The forgetting factor for the recursive estimation of the moments according to (13) has been chosen to $\lambda = 0.97$. The input signal has been a zero-mean, uncorrelated, non-stationary Laplacian process [7], and a signal-to-noise ratio of 35 dB w.r.t $n(k)$ has been preset. As evaluation criterion we use the echo return loss enhancement (ERLE) defined as

$$\mathrm{ERLE} = 10 \log \frac{\mathrm{E}\{d^2(k)\}}{\mathrm{E}\{e^2(k)\}} \text{ [dB]}. \qquad (30)$$

The ERLE graphs that have been obtained for an orthogonalized power filter (OPF) with bias correction (BC) according to (28), an orthogonalized power filter without bias correction, and the corresponding non-orthogonalized power filter are shown in Fig. 4. As can be seen from Fig. 4, the achievable echo attenuation of the adaptive EOS without bias correction is limited due to time-variant orthogonalization matrices. Furthermore, we notice that the EOS with bias control and the non-orthogonalized power filter lead to the same final echo attenuation, where the EOS provides a significantly faster convergence speed. The second experiment is based on recorded speech data from a loudspeaker of a mobile phone receiver placed in an enclosure with low-reverberation. For the recording, the loudspeaker has been mounted in the handset, while the microphone has been separated from it to avoid undesired vibration effects. The parameters for the third-order power filter has been the same as for the previous experiment. The ERLE graphs resulting from the proposed orthogonalized power filter with bias correction and a linear echo canceller with filter length $N = 200$ are presented in Fig. 5. We notice that the nonlinear approach leads to a significant increase in echo attenuation compared to a linear echo canceller if the loudspeaker causes non-negligible nonlinear distortion. Although not presented here, simulations including nonlinear amplifiers yield similar results with respect to performance improvement compared to a linear echo canceler.

6. CONCLUSION

We presented orthogonalized adaptive power filters including a bias correction for time-variant orthogonalization of non-stationary input signals. It has been shown that power filters are well suited to model the nonlinear echo path of mobile communication receivers, i.e., the cascade of a memoryless saturation characteristic followed by a linear filter. The simulation results confirm the capability of the proposed approach to cope with nonlinear distortion in the echo path, introduced by nonlinear loudspeaker systems.

7. REFERENCES

[1] A. Stenger and W. Kellermann, “Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling,” *Signal Processing*, vol. 80, pp. 1741–1760, Sept. 2000.

[2] B. S. Nollett and D. L. Jones, “Nonlinear echo cancellation for hands-free speakerphones,” in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing (NSIP)*, Sept. 1997.

[3] R. Raich, H. Qian, and G.T. Zhou, “Digital baseband predistortion of nonlinear power amplifiers using orthogonal polynomials,” in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2003.

[4] W. Klippel, “Dynamic measurement and interpretation of the nonlinear parameters of electrodynamic loudspeakers,” *J. Audio Eng. Soc.*, vol. 38, no. 12, pp. 944–955, Dec. 1990.

[5] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, New Jersey: Prentice Hall, 2000.

[6] S. Haykin, *Adaptive Filter Theory*, New Jersey: Prentice Hall, 2001.

[7] H. Brehm and W. Stammler, “Description and generation of spherically invariant speech-model signals,” *Signal Processing*, vol. 12, no. 2, pp. 119–141, March 1987.