FREQUENCY DOMAIN IMPROVED PRACTICAL VARIABLE STEP-SIZE FOR ADAPTIVE FEEDBACK CANCELLATION USING PRE-FILTERS

Linh T. T. Tran¹, Henning Schepker², Simon Doclo², Hai H. Dam¹, and Sven Nordholm¹

¹Faculty of Science and Engineering, Curtin University, Perth, Australia
{t.tran57@postgrad, H.Dam@exchange}curtin.edu.au, S.Nordholm@curtin.edu.au
²Signal Processing Group, Department of Medical Physics and Acoustics and the Cluster of Excellence "Hearing4All", University of Oldenburg, Oldenburg, Germany
{henning.schepker, simon.doclo}@uni-oldenburg.de

ABSTRACT

In adaptive feedback cancellation (AFC) methods, the step-size plays an important role in controlling the convergence speed of an adaptive filter in the feedback canceller path. The selection of this step-size provides a compromise between a low steady-state error and a fast convergence rate. The use of a variable step-size (VSS) is a potential solution to achieve both fast convergence and low steady-state error. In this paper, we propose a frequency-domain AFC method which integrates an improved practical VSS (IPVSS) algorithm into a partitioned-block frequency-domain (PBFD) implementation of the prediction error method (PEM) for hearing aid applications. The proposed method derives benefit from the IPVSS algorithm, e.g., a better compromise solution between convergence and steady-state error, and from the PBFD-PEM, e.g., a low numerical complexity and an improved convergence. The proposed method is evaluated for different types of speech incoming signals as well as for a sudden change of the acoustic feedback path. Simulation results show that the proposed method provides a significant performance improvement compared to the PBFD affine combination approach as well as the PBFD using either the upper or the lower step-sizes which are utilised as boundaries in the IPVSS algorithm.

Index Terms— Acoustic feedback cancellation, hearing aid, improved practical variable step-size, PBFD, PEM.

1. INTRODUCTION

In hearing aids (HAs), there always exists an acoustic feedback signal leaking from the loudspeaker to the microphone. This feedback signal deteriorates the sound quality as well as limits the achievable maximum amplification of HAs. In certain conditions, this acoustic feedback can render the system unstable, i.e., howling may be perceived.

Among many existing methods to reduce acoustic feedback considered in literature, see [1–8], adaptive feedback cancellation (AFC) is one of the most efficient methods. In the AFC method, typically the impulse response (IR) of the true feedback path is estimated by an adaptive filter. The least mean squares (LMS) algorithm and the normalized LMS (NLMS) algorithm are commonly used for adaptive filtering. However, there often exists a bias in the IR estimate due to the correlation between the incoming signal and the loudspeaker signal.

To reduce the bias, the prediction error method (PEM) is one of the most successful AFC methods, which has been implemented in both time domain [3, 9–11] and frequency domain [12–16]. In the PEM, the speech incoming signal is assumed to be modelled by passing white Gaussian noise through an autoregressive (AR) filter. The inputs of the adaptive filter used to estimate the feedback path will be pre-whitened by utilizing an estimate of the inverse of the AR-filter. As a result, the PEM can significantly lower this bias, especially for speech incoming signals.

To further improve the system performance, some variants of the PEM have been introduced. For instance, the affine projection algorithm (APA) or the proportionate NLMS (PNLMS) algorithm was used in [17, 18], whereas a Kalman filter was used in [12], and subband filters as well as a frequency shift were used in [14]. A partitioned-block frequency-domain (PBFD) affine combination of two adaptive filters using the NLMS algorithm with two different step-sizes for the PEM, PBFD-PEM-AffComb, has been proposed in [13]. In [19] a hybrid AFC (H-AFC) scheme which used a switched combination adaptive filter controlled by a soft-clipping-based stability detector has been investigated.

Moreover, variable step-size (VSS) algorithms are promising techniques in order to significantly improve the trade-off between a fast convergence rate and a low steady-state error. Although VSS algorithms have been commonly used in acoustic echo cancellation (AEC), their application for acoustic feedback cancellation in HAs is a challenge because of the strong correlation between the loudspeaker and incoming signals. To reduce the correlation VSS algorithms in AFC methods for HAs have often been combined with the PEM, e.g., [14, 15, 20–22]. However, since the PEM can not completely decorrelate the loudspeaker and incoming signals the estimation of the VSS in the time-domain remains challenging.

In this paper, we implement an improved practical VSS (IPVSS) algorithm in the PBFD-PEM for hearing aids, making use of the generally known properties of PBFD processing such as complexity reduction and convergence improvement [11]. In addition, this implementation makes use of the upper and lower step-size bounds to control the step-size such that when the system is (or close to) unstable, the large step-sizes are picked up and when the system converges, the small step-sizes are selected [21]. As a result, the proposed method, PBFD-PEM-IPVSS, achieves fast convergence and tracking rates while maintaining a low steady state-error. The proposed PBFD-PEM-IPVSS is evaluated for different types of speech incoming signals, including concatenated male-female speech, male speech and female speech as well as for a sudden change of the measured acoustic feedback channels. The simulation results show that the proposed PBFD-PEM-IPVSS and the existing PBFD-PEM-AffComb outperform the PBFD-PEM-NLMS utilizing only single step-size. Furthermore, the PBFD-PEM-IPVSS leads to an improved
performance compared to the PBFD-PEM-AffComb.

2. AFC METHODS

For simplicity, in this paper we assume that all AFC methods are discrete-time, and that the incoming signal is stationary. Fig. 1 depicts a conventional AFC method. In this method, an adaptive filter $\hat{F}(q)$ with $L_f$-dimensional impulse response (IR) $\hat{F} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{L_f-1}]^T$ is utilized to estimate the IR of the true feedback path, where $q^j$ is the discrete-time delay operator. The $L_f$-dimensional vector, $\mathbf{f} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_{L_f-1}]^T$, denotes the IR of the true feedback path $F(q)$. The microphone signal is generated by adding the feedback signal $v(k) = F(q) y(k)$ to the incoming signal $u(k)$, i.e.,

$$x(k) = u(k) + F(q) y(k),$$  

(1)

where $k$ is the discrete-time index and $y(k)$ denotes the loudspeaker signal. The error signal $e(k)$ is produced by subtracting the estimated feedback signal, $\hat{v}(k) = \hat{F}(q) y(k)$, from the microphone signal $x(k)$, i.e.,

$$e(k) = x(k) - \hat{F}(q) y(k),$$  

(2)

where $\hat{F}(q)$ is an estimate of $F(q)$. The loudspeaker signal can be computed as follows

$$y(k) = K(q) e(k),$$  

(3)

where $K(q)$ is the forward path gain with a delay $d_k \geq 1$ [3, 9]. Minimizing the mean-square error (MSE), $\min \{ E\{ e^2(k) \} \}$ with assumption of a sufficient order, i.e., $L_f = L_y$ and a linear time-invariant feedback path, results in the Wiener filter solution

$$\hat{f}_0 = f + R_{yu}^{-1} r_{yy},$$  

(4)

where $y(k) = [y(k), y(k-1), \ldots, y(k-L_f+1)]^T$ is a $L_f$-dimensional vector of the loudspeaker signal, $E\{ \cdot \}$ denotes the expectation operator, $R_{yx} = E\{ y(k) y^T(k) \}$ is the auto-correlation matrix of $y(k)$, and $r_{yu} = E\{ y(k) u(k) \}$ is the cross-correlation vector between $y(k)$ and $u(k)$. From the bias term in (4), we can see that the estimate of the feedback path suffers from the correlation between the loudspeaker signal and the incoming signal.

Moreover, the coefficients of the estimated feedback path can be recursively approximated using the NLMS algorithm as

$$\hat{f}(k) = \hat{f}(k-1) + \frac{\mu}{y^T(k) y(k) + \delta} y(k) e(k),$$  

(5)

where $\mu$ is a fixed step-size, and $\delta$ is a small regularization parameter.

3. PROPOSED PBFD-PEM-IPVSS

While the NLMS algorithm uses a fixed step-size, in this section we propose to implement a variable step-size as used in the IPVSS algorithm. To improve the performance, we integrate the IPVSS in the PBFD-PEM, cf. Fig. 2. In the PBFD-PEM [3, 11], the adaptive filter $\hat{F}(q)$ is partitioned into $L_f/P$ chunks, each chunk is a $P$-dimensional vector with IR, $\hat{F} = [\hat{F}_P, \hat{F}_{P+1}, \ldots, \hat{F}_{(L_f/P)-1}]^T$, $i = 0, \ldots, L_f/P - 1$. Then these partitioned impulse responses are transformed to the frequency domain as follows

$$\hat{F}_i = \mathcal{F} \left[ \hat{f}_i \right],$$  

(6)

where $\mathcal{F}$ denotes the $(N \times N)$-dimensional DFT matrix, and $N$ is the DFT length. The loudspeaker and microphone signals are pre-whitened by using the filter $\hat{G}(q)$ which is an estimate of the inverse AR model $G(q)$ of the incoming signal, i.e., $y^p(nL + j) = \hat{G}(q) y(nL + j)$ and $x^p(nL + j) = \hat{G}(q) x(nL + j)$, where $L$ is the block length, $j = 1, \ldots, L$, and $n$ is the block time index. The condition $N \geq P + L - 1$ needs to be fulfilled to ensure proper operation [11]. The pre-whitened input of adaptive filter for each chunk can be calculated as

$$Y_i^p(n) = \text{diag} \left\{ \mathcal{F} \left[ y^p \left[ (n+1) \cdot L - iP - N + 1 \right] \right], \ldots, y^p \left[ (n+1) \cdot L - iP \right] \right\}.$$  

(7)

The pre-whitened block error signal is denoted in frequency domain as

$$E_i^p(n) = \mathcal{F} \left[ x^p(n) - \psi^p(n) \right];$$  

(8)

where $x^p(n)$ is the $L$-dimensional block microphone signal, $x^p(n) = [x^p(nL + 1) \ldots \cdots x^p((n+1) \cdot L)]^T$ and $\psi^p(n)$ is the $L$-dimensional block estimated feedback signal after pre-whitening,

$$\psi^p(n) = \begin{bmatrix} 0 & I_L \end{bmatrix} \mathcal{F}^{-1} \sum_{i=0}^{L_f/P-1} Y_i^p(n) \hat{F}_i(n).$$  

(9)

The coefficients of the adaptive filter can be recursively updated as follows

$$\hat{F}_i(n + 1) = \hat{F}_i(n) + \Gamma(n) \Theta \Theta^{-1} Y_i^p H (n) E_i^p(n),$$  

(10)

where $\Gamma(n) = \text{diag} \{ \mu_0(n) \ldots \mu_{N-1}(n) \}$ is the step-size matrix and $\Theta = \begin{bmatrix} I_P & 0 \\ 0 & 0_{N-P} \end{bmatrix}$ is the gradient-constraint matrix. Each
frequency-dependent step-size \( \mu_l(n) \), \( l = 0, \ldots, N - 1 \) in the matrix \( \Gamma(n) \) is normalized, i.e., [11]

\[
\mu_l(n) = \frac{\tilde{\mu}_l}{\Xi_{\mu,l}(n)} + \Xi_{\mu,l}(n) + \epsilon,
\]

where \( \tilde{\mu}_l \) is a fixed step-size; \( \epsilon \) is a constant regularization parameter. Parameters \( \Xi_{\mu,l}(n) \), \( \Xi_{\mu,l}(n) \), and \( \Xi_{\mu,l}(n) \), represent the input power, the microphone power, the estimated feedback signal power, and the error power after pre-whitening, respectively.

\[
\Xi_{\mu,l}(n) = \sum_{i=0}^{L_f,\hat{P}^{-1}} |Y_{\mu,l}(n)|^2, \quad \Xi_{\mu,l}(n) = \frac{\hat{L}_f}{\hat{L}_f} |X_l(n)|^2, \\
\Xi_{\mu,l}(n) = \frac{\hat{L}_f}{\hat{L}_f} |\hat{Y}_{\mu,l}(n)|^2, \\
\Xi_{\epsilon,l}(n) = \frac{\hat{L}_f}{\hat{L}_f} |E_l(n)|^2.
\]

In the proposed PBFD-PEM-IVPSS, we implement the IPVSS algorithm in the PBFD-PEM in order to obtain a better compromise solution between fast convergence/tracking rates and low steady-state misalignment. In fact, the fixed step-size \( \tilde{\mu}_l \) in (11) is replaced by a variable step-size as used in the IPVSS algorithm, i.e.,

\[
\tilde{\mu}_l(n) = \begin{cases} 
\mu_1 & \text{if } \mu_{1,c}(n) > \mu_1 \\
\mu_2 & \text{if } \mu_{1,c}(n) < \mu_2 \\
\mu_{1,c}(n) & \text{otherwise}
\end{cases}
\]

where \( \mu_1 \) and \( \mu_2 \) are the constant upper and lower bounds of the step-size \( \mu_l > \mu_2 \), \( \mu_{1,c}(n) = \mu_1 \cdot \bar{\mu}_l(n) \), with \( \bar{\mu}_l(n) = 1 - \sqrt{\frac{\Xi_{\mu,l}(n) - \Xi_{\mu,l}(n) + \zeta}{\Xi_{\mu,l}(n) + \zeta}} \) [23], \( \zeta \) is a small positive value. As mentioned in [21], the constant step-sizes \( \mu_1 \) and \( \mu_2 \) are used to limit the fluctuation of the variable step-size in a smaller range, resulting in smoother system performance. Moreover, the benefits of using a variable step-size for convergence and tracking rates as well as for steady-state misalignment are still maintained.

4. SIMULATION RESULTS

Fig. 3 shows the amplitude responses of the measured acoustic feedback paths with a two-microphone behind-the-ear hearing aid [24] used for the simulations. It can be seen that the feedback path with a telephone placed close to the ear has larger amplitude than the free-field feedback path. In all simulations, three types of incoming signals generated by the concatenated male and female speech patterns, only the male speech patterns and only the female speech patterns extracted from NOIZEUS database [25] were used. All incoming signals were of length 80 s and recorded by using a microphone of a mock-up hearing aid. The misalignment and added stable gain are

\[
\text{MIS} = 10 \log_{10} \left( \frac{\int_{\Omega} \int_{\Omega} |F(e^{j\Omega}) - \hat{F}(e^{j\Omega})|^2 d\Omega}{\int_{\Omega} \int_{\Omega} |F(e^{j\Omega})|^2 d\Omega} \right),
\]

\[
\text{ASG} = 20 \log_{10} \left( \min_{\Omega} \left| \frac{1}{F(e^{j\Omega})} \right| \right) - 20 \log_{10} \left( \min_{\Omega} \left| \frac{1}{\hat{F}(e^{j\Omega})} \right| \right),
\]

where \( F(e^{j\Omega}) \) and \( \hat{F}(e^{j\Omega}) \) are frequency responses of the true and the estimated feedback paths at the normalized frequency \( \Omega \), respectively. We also chose the following parameters for all simulations: the delay in the forward path \( d_f = 96 \) samples, the gain in the forward path \( |K| = 20 \) dB, the sampling frequency \( f_s = 16 \) kHz, \( P = 32, N = 128, L = 32, \epsilon = 10^{-10} \) and the length of the true and estimated feedback paths \( L_f = 100 \) and \( L_j = 64 \), respectively.

We used the following parameters for the proposed PBFD-PEM-IVPSS: the upper and lower bounds of step sizes \( \mu_1 = 0.015 \) and \( \mu_2 = 0.001 \); a small positive constant \( \zeta = 10^{-6} \). Furthermore, we compare with the PBFD-PEM-AffComb and use the parameters as proposed in [13]. In all AFC methods using the PEM, a 20-order prediction-error filter \( \hat{G}(q) \) was computed for every frame of 160 samples by using the Levinson-Durbin algorithm [27]. All mentioned AFC methods were evaluated with a sudden change of the feedback path after 40 s.
Fig. 4 compares the misalignment and ASG of the proposed method with the PBFD-PEM using either $\mu_1$ or $\mu_2$ and the PBFD-PEM-AffComb for concatenated (male and female) speech incoming signal. The PBFD-PEM-AffComb has as high initial convergence and tracking rates as the PBFD-PEM using $\mu_1$ and as low steady-state error as the PBFD-PEM using $\mu_2$ when the system has converged. We can see that the proposed method outperforms the PBFD-PEM using either $\mu_1$ or $\mu_2$. Although it initially converges slower than the PBFD-PEM using $\mu_1$ and the PBFD-PEM-AffComb for the free-field feedback path, it provides a much lower and smoother steady-state misalignment as well as a higher ASG. Furthermore, when the feedback path changes to the telephone-near feedback path after 40 s, the proposed method can track the changes almost as fast as the PBFD-PEM using $\mu_1$ and the PBFD-PEM-AffComb but it obtains approximately 1.5-2 dB and 2-4 dB misalignment improvements compared to the PBFD-PEM-AffComb and the PBFD-PEM using $\mu_1$, respectively.

Fig. 5 and Fig. 6 illustrate the performance of all mentioned AFC methods for the male and female speech incoming signals, respectively. It can be seen that the behaviour of the proposed PBFD-PEM-IPVSS is similar to the case with concatenated speech incoming signal. It still outperforms the PBFD-PEM using $\mu_2$ and yields a similar tracking rate but a lower steady-state misalignment and a higher ASG compared to the PBFD-PEM-AffComb as well as the PBFD-PEM using $\mu_1$. In particular, when the incoming signal is male speech, the proposed method achieves an improvement of 1.5-2 dB in term of misalignment before the change of feedback path and an improvement of 1-1.5 dB after the change compared to the PBFD-PEM-AffComb. When the incoming signal is female speech, the proposed method provides approximately 2-2.5 dB and 3-4 dB misalignment improvements compared to the PBFD-PEM-AffComb and the PBFD-PEM using $\mu_1$, respectively. It also provides much smoother performance and higher added ASG than those two methods. Its initial convergence seems to be slower than the PBFD-PEM-AffComb and the PBFD-PEM using $\mu_1$ but much quicker than the PBFD-PEM using $\mu_2$.

The initial slow convergence of the proposed method is from the high correlation between the loudspeaker and the incoming signal as well as the large mismatch between the estimated and the true feedback paths in this situation [28].

5. CONCLUSION

In the paper, we propose to implement the IPVSS algorithm in the PBFD-PEM in order to make use of the benefit of both. As a result, the proposed method achieves a better compromise solution between fast initial convergence and tracking rates and a low steady-state misalignment compared to other mentioned methods. Simulation results utilizing three types of speech incoming signals as well as two measured acoustic feedback paths show that the proposed method outperforms the PBFD-PEM employing either the upper or lower step-size limits used in the IPVSS algorithm. Moreover, it also provides smoother performance, a lower steady-state misalignment, and a higher ASG while maintaining a similar tracking rate compared to the PBFD-PEM-AffComb.
6. REFERENCES


