UNSUPERVISED NOISE-AWARE ADAPTIVE FEEDBACK CANCELLATION FOR HEARING AID DEVICES UNDER NOISY SPEECH FRAMEWORK

Parth Mishra, Anshuman Ganguly, Abdullah Küçük, Issa M.S. Panahi

Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX, USA
{ pxm161830, axg124330, axk166230, imp015000 }@utd.edu

***Abstract*— This paper proposes a novel adaptive feedback cancellation (AFC) architecture to improve the performance of an existing robust AFC method in the presence of noisy speech conditions. By employing a computationally efficient Spectral flux (SF) feature-based unsupervised voice activity detector (VAD), we adaptively control the step sizes in the proposed AFC algorithm (SFPEM-AFC). The proposed AFC method achieves faster convergence and lower misalignment errors than earlier methods. Objective evaluation of the AFC algorithm is presented using Signal to Feedback Ratio (SFR) and Misalignment (MISA) values for several noisy conditions. The Proposed method shows a significant reduction in the MISA values while maintaining higher SFR and higher perceptual quality over the earlier methods. Experimental results are presented for realistic noisy conditions to demonstrate the superiority of the proposed noise-adaptive AFC method for hearing aid devices (HADs).**

I. INTRODUCTION

Acoustic feedback in hearing aid devices (HADs) is the unpleasant high-pitched whistling sound that arises when the loudspeaker output is repeatedly amplified in a closed-loop configuration. Thus, this recurring amplification process leads to instability, resulting in loud irritating 'ringing' (howling) sound. If this feedback signal is not controlled, it could lead to large distortion in speech or even permanently damage the HAD, in some cases. Acoustic feedback is a major concern in hearing aid devices and still one of the main factors causing hearing aid user dissatisfaction. [1] Figure 1 shows acoustic coupling between the loudspeaker and the microphone of the HAD. Reports [1], [2] indicate that as many as 10% to 15% of in-the-ear hearing aids are likely to be returned to the manufacturing plant within the first ninety days after manufacture for feedback-related problems. Obviously, this adds to the overall cost of the HADs as well as causes inconvenience to the user [2]. It is still an open and challenging research problem and requires additional research into improving convergence speed without hurting the system stability. Thus, this problem requires an efficient solution without any compromise in speech quality and intelligibility.

Adaptive feedback cancellation (AFC) is an efficient way to tackle the acoustic feedback problem in an adaptive framework. A typical AFC algorithm first estimates the unknown feedback path adaptively and then cancels feedback signal using the estimated acoustic feedback path. Different methods have been used to tackle this problem and increase the precision of the AFC system. Prediction Error Method AFC (PEM-AFC) [3] is a very successful method in this area in which whitening technique is performed to reduce the correlation between the desired signal and feedback signal [3], [4]. This approach utilizes two types of adaptive filters: First to pre-filter the signals and the second to estimate the feedback path transfer function. It typically uses Partitioned Block Frequency Domain Normalized Least Mean Square (PBFD-NLMS) algorithm for the feedback path estimation because frequency-domain adaptive filters have lower computational complexity and better control over frequency-domain step size (μ) [5-7].

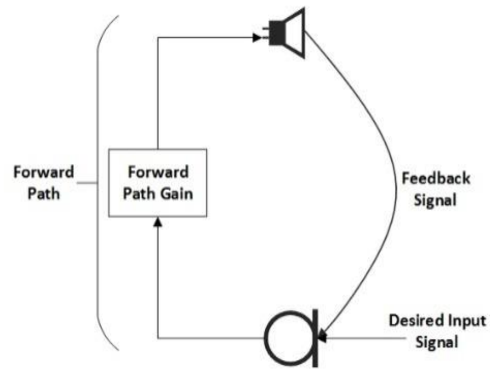

Figure 1. Acoustic coupling between Loudspeaker and Microphone

In several earlier works [8-11], the performance of AFC algorithms is studied only for clean speech input signals. However, the input signal is often corrupted with background noises, which can affect the performance of the AFC algorithms. It is well known [8-11] that the choice of μ reflects a trade-off between fast convergence (Large μ) and low misalignment (MISA) (Small μ) in the feedback path estimation. In [8], AFC uses a pitch based VAD to improve speech quality and continuously attenuate the step index value as speech frames are detected in the input data stream. In [9], the analysis of static and variable feedback path is carried out with varying μ for PEM AFC using PBFD-NLMS for clean speech. Optimum μ is obtained empirically based on the behavior of misalignment. A joint AFC and noise reduction scheme was proposed for PEM-AFC in [10], wherein the focus was on noise suppression rather than on the performance analysis in presence of noisy input signals.

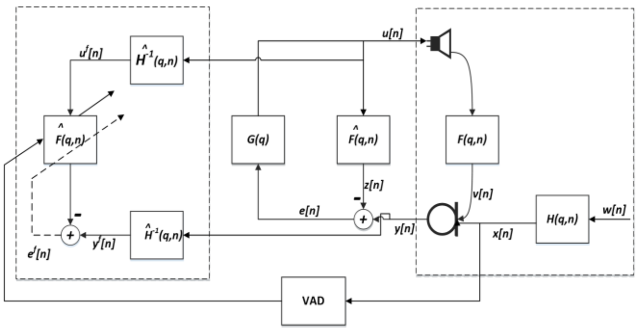

Figure 2. Block Diagram of Proposed AFC Method

In this paper, we present the performance analysis of similar method in [9] using PBFD-NLMS algorithm for noisy speech signals. We present our approach that makes use of an unsupervised VAD's decision to efficiently control two values of μ in the PBFD-NLMS weight update equation by achieving a balance between two critical objectives The values of μ are obtained by using a grid refinement approach by minimizing MISA over noise only and noisy speech separately. This also allows us to control μ disjointly over noisy speech (Smaller μ) and noise only (Larger μ) segments in an efficient manner. As shown in our results, this approach achieves faster convergence and lower MISA, while maintaining higher SFR and Perceptual Evaluation of Speech Quality (PESQ) [15] values over earlier methods.

Performance of this new proposed VAD based PEM-AFC algorithm is analyzed in presence of different types of additive background noises at different signal to noise ratios (SNRs) and presented in this paper. We quantify and compare the performance of the AFC algorithm using Misalignment (MISA) [9], Signal to Feedback Ratio (SFR) [11] and PESQ [15] for noisy speech framework.

The paper is organized as follows: Section II introduces the PEM-AFC algorithm. Adaptive filter using PBFD-NLMS algorithm for estimating feedback path coefficients is presented in Section III. Section IV presents the Voice Activity Detector used. Section V presents the proposed method and Section VI presents the experimental results and analysis. Conclusions are drawn in Section VII.

II. PEM-AFC ALGORITHM

The closed loop system leads to the correlation between the feedback signal and the desired signal. Hence, it provides a biased estimate of feedback transfer function. Thus, this correlation can be reduced by means of whitening filters as depicted in Figure 2. Here we assume that the desired signal $x(n)$ can be modeled well as white noise sequence filtered through a 'P' order time-varying AR (autoregressive) model [3] given by $x[n] = H(q,n)w[n]$ where $H(q,n)$ is an AR model and inversely stable, $x[n]$ is the input noisy speech signal, $w[n]$ is a zero-mean white noise sequence or an impulse train for voiced or unvoiced phonemes respectively. So, that the inverse filter of $H(q)$ (say $A(q,n) = H(q,n)^{-1}$) can be used to decorrelate the signals in the identification task.

We assume that

$$A(q,n) = H^{-1}(q,n) = 1 + q^{-1}\bar{A}(q,n) \tag{1}$$

Where $\bar{A}(q,n)$ is a FIR filter, q is the discrete-time delay and n is the time index. However, $H(q,n)$ is not only unknown but also time varying therefore, it needs to be estimated in an adaptive way along with $F(q,n)$. In Figure 2, $G(q)$ is a forward path gain and $v[n]$ is a feedback signal. $H(q,n)$ is a FIR filter with L_F number of coefficients and can be represented as

$$H(q,n) = \boldsymbol{h}^T[n]\boldsymbol{q} \tag{2}$$

Where vector $\boldsymbol{h}[n] = \left[h_0[n] \;\; h_1[n] \;\; \ldots \;\; h_{L_F-1}[n]\right]^T$ contains filter coefficients, $\boldsymbol{q} = [1 \;\; q^{-1} \;\; \ldots \;\; q^{-L_F+1}]^T$ the delay operator. Superscript T denotes transpose operation. Equation (1) is represented as:

$$x[n] = H(q,n)w[n] = \boldsymbol{h}^T\boldsymbol{w} \tag{3}$$

where $\boldsymbol{w} = \left[w[n] \;\; w[n-1] \;\; \ldots \;\; w[n-L_F+1]\right]^T$.

The AR model is estimated using the Levinson-Durbin recursion procedure (as in [8]) in order to obtain a whitened signal, for each frame of the error signal $e[n]$. The inverse of this AR model is used to whiten the signals $y[n]$ and $u[n]$ to obtain $y^f[n]$ and $u^f[n]$ respectively. These pre-filtered signals are used as the input signal and the desired signal of the adaptive filter. Optimum estimation of the feedback path is found by minimizing the energy of the prediction error defined as

$$e[n] = \hat{H}^{-1}(q,n)(y[n] - \hat{F}(q,n)u[n]) \tag{4}$$

$$\boldsymbol{J}(\hat{\boldsymbol{f}}[n]) = E\left\{ |\hat{H}^{-1}(q,n)(y[n] - \hat{F}(q,n)u[n])|^2 \right\} \tag{5}$$

$$= E\left\{ |y^f[n] - \hat{\mathbf{f}}^T[n]\,\boldsymbol{u}^f[n])|^2 \right\} \tag{6}$$

Where $\hat{\mathbf{f}}[n]$ is filter coefficients of adaptive feedback canceller $\hat{F}(q,n)$. $\boldsymbol{J}()$ represents the cost function and $E\{\}$ denotes the expectation operator over n.

$$\boldsymbol{u}^f[n] = \left[u^f[n] \;\; u^f[n-1] \;\; \ldots \;\; u^f[n-L_F+1]\right]^T \tag{7}$$

$$u^f[n] = \hat{H}^{-1}(q,n)u[n] \tag{8}$$

and $$y^f[n] = \hat{H}^{-1}(\mathrm{q,n})\ \mathrm{y[n]} \tag{9}$$

Minimization of the $\boldsymbol{J}(\hat{\boldsymbol{f}}[n])$ results in:

$$\hat{\mathbf{f}}[n] = \varepsilon\,\{\hat{\boldsymbol{u}}^f[n]\hat{\boldsymbol{u}}^{f\,T}[n]\}^{-1}\varepsilon\,\{\hat{\boldsymbol{u}}^f[n]y^f[n]\} \tag{10}$$

Unbiased feedback path estimation results if $\hat{H}(q,n)$ is assumed to be equal to the true desired signal model $H(q,n)$ and therefore

$$y^f[n] = w[n] + \hat{F}(q,n)u^f[n] \tag{11}$$

III. PBFD-NLMS ALGORITHM

PBFD-NLMS estimates coefficients of feedback path by implementing NLMS algorithm in a frequency domain in block mode of operation [6-8]. In this method, the adaptive filter $\hat{F}(q,n)$ is divided into several subfilters. The output of the $\hat{F}(q,n)$ is computed by the following convolution between the impulse response of the filter $\hat{f}_m[n]$ and the input signal $u[n]$:

$$z[n] = \sum_{m=0}^{N-1} \hat{f}_m[n]u[n-m] \tag{12}$$

Where $z[n]$ is the output and $\hat{f}_m[n]$ is the m^{th} coefficient of the impulse response. Let

$$z[n] = \sum_{p=0}^{P-1} z_p[n] \tag{13}$$

$$z[n] = \sum_{l=0}^{M-1} \hat{f}_{l+pM}[n]u[n-l-pM] \tag{14}$$

Where P and M are number of partitions (or sub-filters) and length of each sub-filter, respectively. Assuming $\hat{F}(q,n)$ has length N ($N = P$ x M), PBFD-NLMS partitions this convolution into smaller convolutions [6]. These convolutions are individually calculated in the frequency domain and stacked together to provide the output [7]. The main idea is to partition the adaptive filter into sub filters and implement each of them separately in the frequency domain using FFT. Generally, 50% overlap is considered between consecutive blocks of length L [7].

IV. VOICE ACTIVITY DETECTOR (VAD)

With our proposed method, we test the AFC algorithm in presence of non-stationary background noise, so in our approach, a VAD is used to detect the speech frames in the noisy speech data stream. The proposed AFC algorithm works frame by frame and each frame is processed by a simple 'Spectral Flux' (SF) based VAD [12] which provides speech or non-speech decision signal.

The SF measures how quickly the power spectrum of the signal changes, and for this it measures the difference in the spectrum between two adjacent frames. SF is calculated using the Short Time Fourier Spectrum [13]. Spectral Flux of nth frame and the kth frequency bin, represented by $SF(k,n)$ is calculated using (15) as described below.

$$SF(k,n) = \frac{\sum_{k=0}^{N/2-1}(|X(n,k)|-|X(n-1,k)|)^2}{N} \tag{15}$$

where $X(n,k)$ represents Short Time Fourier Spectrum of nth windowed frame and N is the frame length of the VAD Buffer frame.

During the initial training period, the average value of Spectral Flux - SF_{avg} and mean value of maximum Spectral Flux - SF_{max} are evaluated. In the simulation, we use first 2 frames for setting up the threshold in the training phase, after the training period, SF_{avg} and SF_{max} are used to compare the SF of the incoming data frames. If the SF of the incoming data frame remains higher than the SF_{avg} and SF_{max} for a certain number of data frames, $SpeechFlag$ is raised and this denotes that current frame under observation is a speech frame else noise.

V. PROPOSED ALGORITHM

In our proposed method, we control the step size μ of in the PBFD-NLMS algorithm according to the decisions made by the Voice Activity Detector based on the following weighted step index update equation:

$$\mu = \Upsilon \mu_{ns} + (1-\Upsilon)\mu_n \tag{16}$$

We introduce a new step size control factor Υ in (16), where Υ denotes speech activity for every frame of the input signal as follows:

$$\Upsilon = \begin{cases} 1, & \text{(Noisy speech)} \\ 0, & \text{(Noise only)} \end{cases}$$

In (16), μ_{ns} and μ_n denote the optimum step index values for the noisy speech and the noise respectively. For simulation, we use the values of μ_{ns} and μ_n obtained (using adaptive grid refinement technique) as 0.05 and 0.1 respectively. Thus, the step size control factor (Υ) controls the entire filter weights update process. Here we have used one feature VAD using SF to make decisions regarding the speech presence/absence in a specific frame and hence this proposed method is called Spectral Flux based VAD PEM-AFC algorithm (SFPEM-AFC).

VI. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we provide several simulation experiments and their results to support and justify the improvements in the proposed method.

A. Setup and Performance measure

For our experimentation, we have implemented SFPEM-AFC frame by frame, with a frame size of 20 msec. and the sampling frequency was fixed at 16 kHz. For the analysis of the SFPEM-AFC algorithm using PBFD-NLMS we consider the order of the AR model to be 20 samples and the forward path transfer function is considered [3] as $G(q) = Ge^{d_G}$ where $G$and d_G is set to 15 msec and 5 msec, respectively. We use a static feedback path for simulation, which is a FIR Filter of order 88 samples. The length of the adaptive filter is considered 64 samples and the input data block size is 32 samples. For our analysis, we use 10 different noisy speech files of length 15 secs each and three different non-stationary background noises- Babble, Machinery, and Traffic noise of length 15 secs each. These background noises are added to the speech files at different SNR values, ranging from -5 to +5dB. These data files are available at [14] upon request. To evaluate the performance of the algorithm, we use two quantitative performance metrics: SFR is used to quantify the reduction in feedback, MISA quantifies the estimation error in $\hat{F}(q,n)$.

1. Misalignment (MISA) [9]

It is the normalized energy of the error between the transfer functions of actual feedback path $F(e^{j\omega})$ and its estimate $\hat{F}(e^{j\omega})$:

$$\text{MISA (dB)} = 10\log_{10}\left(\frac{\int_o^{\pi}\left|F(e^{j\omega})-\hat{F}(e^{j\omega})\right|^2 d\omega}{\int_0^{\pi}\left|F(e^{j\omega})\right|^2 d\omega}\right) \tag{17}$$

Lower MISA value indicates better AFC performance.

2. Signal to Feedback Ratio (SFR) [11]

SFR over each data frame is defined as:

$$\text{SFR (dB)} = 10\log_{10}\left(\frac{||x[n]||^2}{||e[n]-x[n]||^2}\right) \tag{18}$$

Where $x[n]$ and $e[n]$ denote the input signal and the feedback signal, respectively. Higher SFR value demonstrates better AFC performance.

B. Simulation Results

Figure 3 depicts the comparative plot of average MISA versus SNR in presence of different background noises for the PEM- AFC utilizing μ suggested in [9] and the proposed method. From the plot, we observe that the average MISA is always least for the proposed method as compared to PEM-AFC for considered noise types under the specified SNRs. We also observe that the error between energies of the estimated and the actual feedback path decreases with decreasing SNR. Figure 4 shows the plot of SFR versus SNR for PEM-AFC utilizing μ suggested in [9] and the proposed method. From the figure, we observe that SFR values are higher for the proposed method in almost every test case, thus our proposed method offers higher feedback cancellation. Figure 5 depicts the plot of PESQ [15] values for the for PEM-AFC utilizing μ suggested in [9] and the proposed method. The proposed method shows improvement in PESQ values for all test cases, and thus with the proposed method the intelligibility of the output signal increases for the listener, i.e. for the Hearing Aid user. Figure 6 shows the convergence of MISA obtained by averaging the MISA values over 60 different noisy speech files. We observe that though PEM-AFC algorithm ('original') converges faster than the SFPEM-AFC algorithm ('proposed') initially due to higher μ. Finally MISA for the proposed method is lower than for the PEM-AFC method.

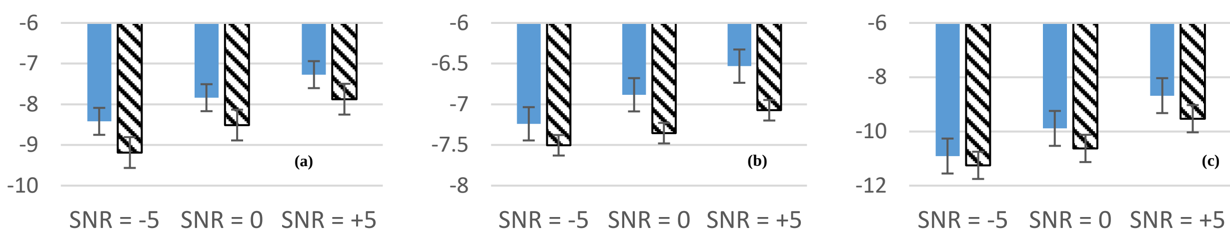

Figure 3. (Left to Right) Misalignment (MISA) (in dB) plot for PEM-AFC (solid) and Proposed Method (dashed) for (a) Babble (b) Machinery, and (c) Traffic noise at different SNR (dB) values. Lower MISA is better

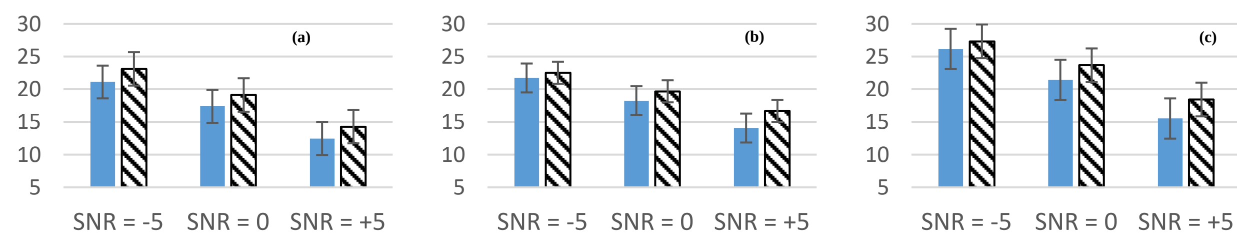

Figure 4. (Left to Right) Signal to Feedback Ratio (SFR) (in dB) plot for PEM-AFC (solid) and Proposed Method (dashed) for (a) Babble, (b) Machinery, and (c) Traffic noise at different SNR (dB) values. Higher SFR is better.

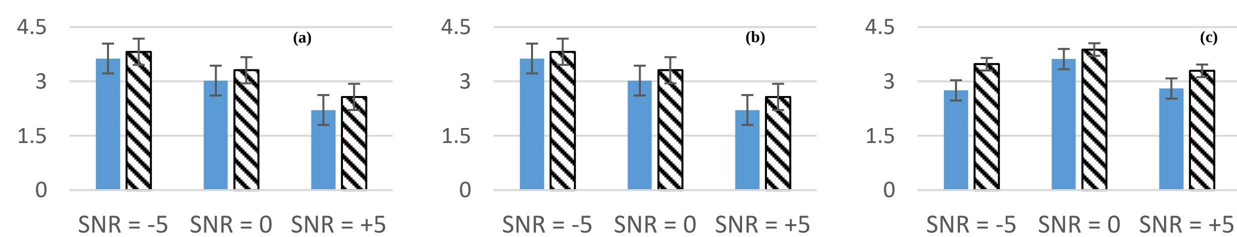

Figure 5. (Left to Right) Perceptual Evaluation of Speech Quality (PESQ) plot for PEM-AFC (solid) and Proposed Method (dashed) for (a) Babble, (b) Machinery, and (c) Traffic noise at different SNR (dB) values. Higher PESQ is better.

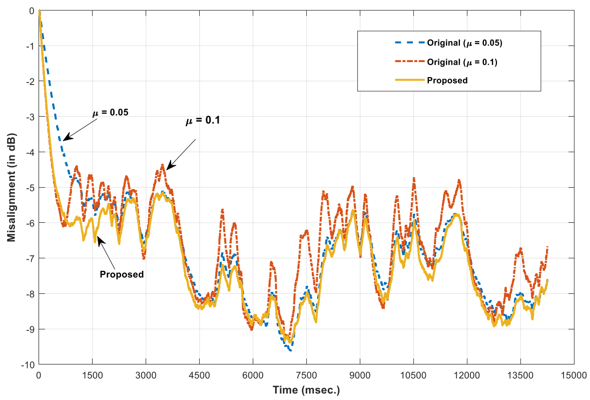

Figure 6. Average MISA for Proposed method. Sixty noisy speech files were averaged at +5dB SNR for Machinery noise. Original method is PEM-AFC in [9]. Proposed method maintains optimal (or lower) MISA value over time.

So overall, the proposed method performs better than PEM-AFC method in terms of feedback path estimation errors (Lower MISA), feedback suppression (Higher SFR) and perceptual quality of speech quality (Higher PESQ) using VAD decision

VII. CONCLUSION

Study of new method 'SFPEM-AFC' in the presence of spectrally colored input signals in the different noisy environment is presented. The use of variable step-sizes allows us to automatically control/ customize the performance in the presence of realistic noisy conditions for HADs. Our experimental results show that by using a simple VAD, the proposed method shows a significant reduction in the MISA and satisfactory improvement in SFR across different noise types. Performance can be further improved using more powerful VAD and pre-filtering techniques. This approach can be deployed in real-time using systematic code optimization techniques for HADs in the future works.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of the Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health (NIH) under the grant number 5R01DC015430-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

REFERENCES

[1] Meng Guo, Ph.D. Thesis, Analysis, Design, and Evaluation of Acoustic Feedback Cancellation Systems for Hearing Aids

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4172233/

[3] A. Spriet, G. Rombouts, M. Moonen, J. Wouters, “Adaptive Feedback Cancellation in Hearing Aids,” Journal of Franklin Institute, Vol. 343, 2006, pp. 545-573.

[4] A. Spriet, I. Proudler, M. Moonen and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," in *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3749-3763, Oct. 2005

[5] J. J. Shynk, "Frequency-domain and multi rate adaptive filtering," in IEEE Signal Processing Magazine, vol. 9, no. 1, pp. 14-37

[6] Kh. S. Chan and B. Farhang-Boroujeny, “Analysis of the Partitioned Frequency-Domain Block LMS (PFBLMS) Algorithm,” IEEE Transactions on Signal Processing, Vol. 49, pp. 1860-1874, 2001.

[7] Egelmeers, G. P. M. (1995). Real time realization concepts of large adaptive filters, Eindhoven: Technische Universiteit

[8] Y. FanChiang, C. W. Wei, Y. L. Meng, Y. W. Lin, S. J. Jou and T. S. Chang, "Low Complexity Formant Estimation Adaptive Feedback Cancellation for Hearing Aids Using Pitch Based Processing," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 8, pp. 1248-1259, Aug. 2014.

[9] S. A. Khoubrouy and I. M. S. Panahi, "Adaptive cancellation of variable feedback path for hearing aid using misalignment-dependent step size values," *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, 2011, pp. 7912-7915.

[10] A. Bastari, S. Squartini and F. Piazza, "Joint Acoustic Feedback Cancellation and Noise Reduction Within the Prediction Error Method Framework," 2008 Hands-Free Speech Communication and Microphone Arrays, Trento, 2008, pp. 228-231

[11] A. Pandey, V. J. Mathews and M. Nilsson, "Adaptive gain processing to improve feedback cancellation in digital hearing aids," *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, 2008, pp. 357-360.

[12] Anshuman Ganguly, Abdullah Küçük, Issa Panahi, " Noise-robust Instantaneous Direction of Arrival estimation using improved GCC for Smartphones " *The Journal of the Acoustical Society of America* (2017) (Under Review).

[13] Robyn M. Cox., Joseph S. Matesich and Jeffrey N. Moore “Distribution of Short-Term RMS Levels in Conversational Speech,” in *The Journal of the Acoustical Society of America*, 1988.

[14] ‘Hearing Aid Project’ http://www.utdallas.edu/ssprl/hearing-aid-project/video-demonstration/

[15] A.W.Rix, J.G. Beerends, M. P Hollier, A. P. Hekstra, “Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs,” *IEEE Int. Conf. Acoust.,Speech, Signal Processing (ICASSP)*, 2, pp. 749-752., May 2001