Applied Acoustics 181 (2021) 108156

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Maximum Versoria-criterion (MVC)-based adaptive filtering method for mitigating acoustic feedback in hearing-aid devices

Muhammad Tahir Akhtar [a,*], Felix Albu [b], Akinori Nishihara [c]

[a] Department of Electrical and Computer Engineering, School of Engineering and Digital Sciences, Nazarbayev University, Kabanbay Batyr Ave. 53, Nur-Sultan 010000, Kazakhstan
[b] Department of Electronics, Valahia University of Targoviste, 130082 Targoviste, Romania
[c] Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152-8552, Japan

ARTICLE INFO

Article history:
Received 25 August 2020
Received in revised form 17 March 2021
Accepted 2 May 2021
Available online 23 May 2021

Keywords:
Acoustic feedback cancellation
Hearing aids
Adaptive filtering
NLMS algorithm
Modeling signal
Maximum Versoria-criterion

ABSTRACT

In any digital hearing aid (HAid), the input microphone and the receiver loudspeaker are closely located. Furthermore, HAid cannot be tightly fitted for comfort reasons. Therefore, it is almost impossible to avoid the acoustic feedback present due to the leakage path between the receiver loudspeaker and the input microphone. This paper investigates acoustic feedback cancellation (AFC) by developing a hybrid adaptive filtering approach. It is proposed to use two simultaneously adapted adaptive filters as in the previous method. The adaptation algorithms are based on a delay-based normalized least mean square (NLMS) algorithm. Furthermore, gradient information from a maximum Versoria-criterion (MVC)-based algorithm is also incorporated in the proposed approach. The effectiveness of this approach is demonstrated through extensive computer simulations.

1. Introduction

A block diagram for a typical hearing aid (HAid) is shown in Fig. 1. As shown, a typical HAid consists of three essential units: the input microphone to pickup the source signal, a feedforward block $G(z)$ which is responsible for all signal processing for example for the user-ear-response dependent frequency domain amplification, and a receiver loudspeaker for generating the signal received by the user. In Fig. 1, the block represented as $F(z)$ denotes the leakage path or the so-called acoustic feedback path (AFP) between the receiver loudspeaker and the input microphone. The source signal (speech, telephone, music, etc.) is $s(n)$, while the processed (received) signal that reaches the user's ear, $x(n)$, also reaches the input microphone via $F(z)$. Therefore, the consequence of the presence of unavoidable $F(z)$ is the undesired feedback signal $x_f(n)$ that may corrupt the input signal. Therefore, the input signal $u(n)$ for the HAid feedforward processing unit $G(z)$, includes the source signal $s(n)$ and the feedback component $x_f(n)$. The block diagram in Fig. 1 can be analyzed in z-domain to get the transfer function between the source signal $s(n)$ and the received signal $x(n)$, as

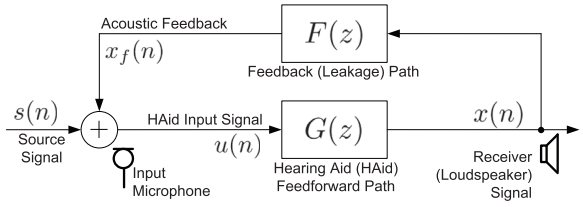


Fig. 1. Illustration of acoustic feedback in a typical hearing aid.

$$H(z) = \frac{G(z)}{1 - G(z)F(z)}. \tag{1}$$

Therefore, it is evident that if $G(z)F(z) = 1$ (e.g., when a large gain in $G(z)$ is desired) at some frequencies, the HAid will be unstable, and the speech quality will be reduced. In these cases, if a high gain value is used, then oscillations are generated which are perceived as annoying whistling and/or screeching sound. This is the so-called howling effect. This limits the maximum gain allowed in the feedforward block $G(z)$ without causing oscillations and/or instability. The acoustic feedback is, therefore, a major problem in HAids that has been studied for many decades [1–6]. The adaptive filtering approach is one of the most useful approach to reduce the acoustic feedback.

* Corresponding author.
E-mail addresses: muhammad.akhtar@nu.edu.kz, akhtar@ieee.org (M.T. Akhtar), felix.albu@valahia.ro (F. Albu), aki@wise-sss.titech.ac.jp (A. Nishihara).

https://doi.org/10.1016/j.apacoust.2021.108156
0003-682X/

Since its inception, adaptive filtering has resulted in successful implementation for a wide range of applications in signal processing, control, and communication [8,7]. In a typical system identification scenario, the objective is to identify the characteristics of the unknown (possibly time-varying) plant by an adaptive filter (AF) [7]. The most popular adaptive algorithm is the Widrow's least mean square (LMS) algorithm [8], where the key idea is to update the AF coefficients by a term comprising gradient information scaled by a learning parameter called the step-size. The main reasons behind popularity of the LMS algorithm are its easy implementation due mainly to a reduced computational complexity for many applications and a robust performance. The normalized LMS (NLMS) algorithm normalizes the step-size with respect to the power/energy of the input signal [9]. This improves convergence of the AF especially for signals with fluctuations (as speech).

In a standard system identification-based application of adaptive filtering for removing or at-least reducing the acoustic feedback in HAids, the desired response and the excitation signals for the AF are derived from the input microphone and the receiver loudspeaker, respectively. The error signal (computed by comparing the microphone signal with the output of the AF) is used to update the AF coefficients as well as excite the HAid feedforward processing unit $G(z)$. It is expected that the AF identifies the (unknown) AFP $F(z)$, and hence generates a replica of the acoustic feedback component $x_f(n)$ which is in fact subtracted from the microphone signal to generate the error signal. Therefore, the error signal of AF is a good candidate to be used as an input $u(n)$ to the HAid feedforward block $G(z)$. However, it is very important to notice that the received signal $x(n)$ and microphone signal (combination of source signal $s(n)$ and acoustic feedback $x_f(n)$) are strongly correlated, as all signals (in a typical setup shown in Fig. 1) are essentially from the same source signal. Due to such a high correlation present between the input and desired response of the AF, it suffers from a bias convergence issue [10].

The techniques developed for the acoustic feedback cancellation (AFC) in HAids can be broadly categorized into one of two approaches. In the first category, the approaches rely on the signals available in the HAid and envisage a mechanism to perform decorrelation of signals for an improved adaptive filtering by inserting a delay in the error signal path [1], or in the forward path [11]. Unfortunately, these approaches can lead to reduced speech quality. An alternative was the filtered-x approach where the error and/or input signals are filtered with decorrelation filters [12,13]. Another alternative were the frequency domain adaptive filtering algorithms [14], but they were not battery power efficient [2]. The dual microphone-based approach for AFC was investigated in [15,16], and its frequency domain implementation presented in [17]. The AFC method proposed in [18] employs affine combination of two adaptive filters with different step-sizes to realize a fast convergence and low steady-state misalignment of the combined filter. A hybrid AFC algorithm has been developed in [19,20] which selects between the classical NLMS update rule or a prediction error method (PEM)-based adaptation [21], depending upon the convergence condition of the AFC system. A soft-clipping-based stability detector helps in making this choice. The key idea here is to exploit fast convergence and low bias properties of conventional NLMS, and PEM-NLMS, respectively [19].

In another class of approaches, a modeling signal (additive random noise) is internally generated and injected into the HAid. Being uncorrelated with the rest of signals present in the HAid, it improves the adaptive filtering for AFC, but it reduces the signal-to-noise ratio (SNR) of the receiver loudspeaker signal. One proposed solution was to inject noise only when the howling occurs [22]. Unfortunately, the sudden ON/OFF switching of the modeling signal is negatively perceived by the HAid user. A continuous injection of a modeling signal reduces this annoyance, but, for a reasonable SNR, the level of modeling signal should be low [23]. In [24], the use of perceptually masking filters for the modeling signal was proposed. A very promising approach was the automatic tune of the modeling signal [25,26]. In our previously published papers, a large-level modeling signal is injected in order to achieve a fast convergence during the transient state and reduced to low level when the adaptive AFC filter converges. Also, two adaptive filters (AFs) are working in tandem, being adapted with a delay-based NLMS algorithm as in [27]. In this paper, we use a hybrid approach for updating the coefficients of AFs in [25,26] and include gradient information from maximum Versoria-criterion (MVC)-based cost function recently investigated in [28]. The MVC-based adaptive filtering provides good performance for non-Gaussian processes [28,29]. The included computer simulations show the effectiveness of the proposed approach for (speech) signals in HAids operating with a short processing delay.

Section 2 briefly describes the biasing problem in the conventional approach for adaptive filtering in HAids, and Section 3 briefs the previous approach [25,26]. The proposed approach, which essentially builds upon the previous one, is explained in detail in Section 4. The simulation results for a variety of signals are presented in Section 5, and the concluding remarks are provided in Section 6. A short version of this paper was presented at a conference [30].

2. Biased convergence in conventional method

The conventional adaptive filtering approach for AFC in digital HAids is shown in Fig. 2 [31]. The microphone signal $m(n)$ comprises the source signal $s(n)$ and a component $x_f(n)$ which is the acoustic feedback from the receiver loudspeaker signal $x(n)$, and can be written as

$$\begin{aligned} m(n) &= s(n) + x_f(n), \\ &= s(n) + \boldsymbol{x}^T(n)\boldsymbol{f}, \end{aligned} \tag{2}$$

where $\boldsymbol{f}$ denotes the impulse-response coefficient vector for the feedback path $F(z)$ and $\boldsymbol{x}(n)$ is the corresponding vector collecting samples of $x(n)$. For discussion presented in this Section, it is assumed without loss of generality that $F(z)$ is a linear time-invariant system with a coefficient vector $\boldsymbol{f}$. In Fig. 2, $\hat{F}(z)$ denotes the AF whose job is to identify the (unknown) AFP $F(z)$. The error signal $e(n)$ is the difference between the microphone signal $m(n)$ and the output $y(n)$ of the AF $\hat{F}(z)$, and is given as

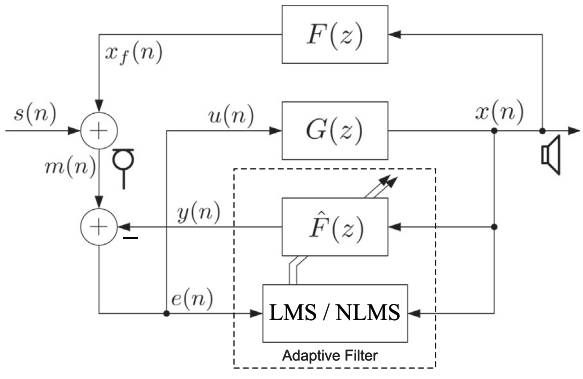


Fig. 2. The conventional AFC scheme using LMS/NLMS algorithms.

$$\begin{aligned} e(n) &= m(n) - y(n), \\ &= m(n) - \boldsymbol{x}^T(n)\hat{\boldsymbol{f}}(n), \\ &= s(n) - \boldsymbol{x}^T(n)\left[\boldsymbol{f} - \hat{\boldsymbol{f}}(n)\right]. \end{aligned} \tag{3}$$

Assuming $\hat{F}(z)$ is well estimating $F(z)$ (available from some offline measurements, for example), then $e(n) \approx s(n)$ and hence, $e(n)$ can be used as input to the HAid feedforward processing unit $G(z)$, i.e., $u(n) = e(n)$. This (ideal) assumption is seriously violated in actual practice, when $\hat{F}(z)$ is adapted during the operation of HAid. The cost function is defined as the mean-squared-error, $J(n) \triangleq E\left\{|e(n)|^2\right\}$, which is minimized to give the Wiener optimal solution for $\hat{F}(z)$ as [7]

$$\hat{\boldsymbol{f}}_o = \boldsymbol{R}_{\boldsymbol{xx}}^{-1}\boldsymbol{r}_{\boldsymbol{xm}}, \tag{4}$$

where $\boldsymbol{R}_{\boldsymbol{xx}} = E\{\boldsymbol{x}(n)\boldsymbol{x}^T(n)\}$ in the auto-correlation matrix for the received signal $x(n)$, and $\boldsymbol{r}_{\boldsymbol{xm}} = E\{\boldsymbol{x}(n)m(n)\}$ denotes the cross-correlation vector between the received signal $x(n)$ and the microphone signal $m(n)$. Substituting Eq. (2) in Eq. (4), it is straight forward to get [10,31]

$$\hat{\boldsymbol{f}}_o = \boldsymbol{f} + \boldsymbol{R}_{\boldsymbol{xx}}^{-1}\boldsymbol{r}_{\boldsymbol{xs}}, \tag{5}$$

where the second term on the right-hand-side (RHS) is a biasing present there due to the strong correlation between the source signal $s(n)$ and the received signal $x(n)$. The NLMS algorithm is used to update the coefficients of AF $\hat{F}(z)$ [9]

$$\hat{\boldsymbol{f}}(n+1) = \hat{\boldsymbol{f}}(n) + \frac{\mu}{\|\boldsymbol{x}(n)\|^2 + \epsilon} e(n)\boldsymbol{x}(n), \tag{6}$$

where $\|\cdot\|$ denotes the Euclidean norm, μ is a fixed step-size (FSS) parameter, and ϵ is the regularization parameter. The objective is $\hat{\boldsymbol{f}}(n) \to \hat{\boldsymbol{f}}_o$ as $n \to \infty$. Unfortunately, a biased convergence can be noticed due to correlation between $s(n)$ and $x(n)$ [10,31]. Therefore, the approach of Fig. 2 cannot be used in a continuous adaptation mode, and the adaptation must be stopped once a 'reasonable' convergence has been achieved

3. Previous method

Consider Fig. 3, which shows a block diagram for the approach previously developed in [25,26]. The input to the first AF $W(z)$ is $x(n)$ and, therefore, job of $W(z)$ is to counteract the acoustic feedback $x_f(n)$. The modeling signal $v(n)$ is used as an input for the second AF $H(z)$, and hence job of $H(z)$ is to take care of the feedback component $v_f(n)$. Here, the uncorrelated modeling signal $v(n)$ is generated from a white Gaussian noise $v_0(n)$ via adaptive gain control parameter $\rho(n)$. The convergence of $W(z)$ is very good, but it may converge to a biased solution as mentioned above. In contrast, $H(z)$ provide a good steady-state estimate of the AFP $F(z)$, but its convergence is slow due to its excitation by the low-level modeling signal $v(n)$. It is clear that both $W(z)$ and $H(z)$ would estimate well $F(z)$ if the biased convergence of $W(z)$ is avoided and the initial convergence of $H(z)$ improved. This goal is achieved by exchanging their coefficients using the coefficient-transfer strategy presented in [25,26], both filters being updated using the (delay-based) NLMS algorithms and reducing the $\rho(n)$ value to a small value when the system converges. This reduces the level of injected modeling signal, and hence improves the achieved SNR. The role of the delay block z^{-D} is to monitor the convergence status of the modeling filter (as explained later while describing the proposed method). It is worth to mention here that including such delay block would increase the overall delay of the feedforward path between the input microphone and the receiver loudspeaker. This eventually

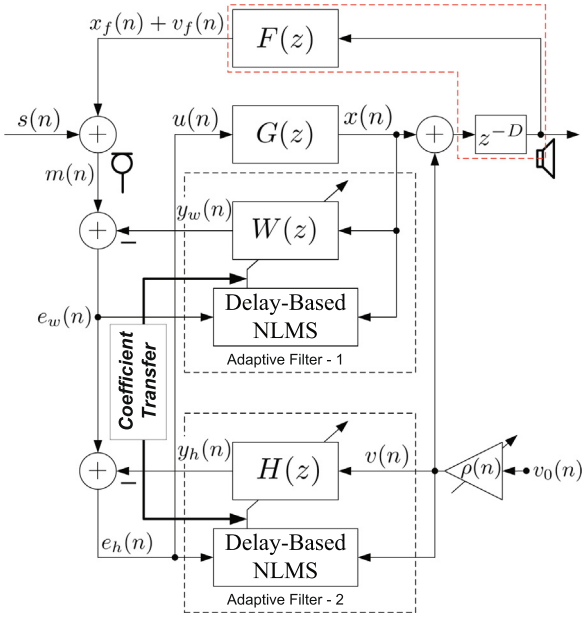


Fig. 3. Block diagram for previous method.

reduces the processing delay available for the HAid signal processing unit $G(z)$. Thus, it is very important to investigate adaptive algorithms able to work with a small processing delay.

Considering non-stationarity of speech and music signals (the main target for any HAid user), in this paper we propose to incorporate MVC-based adaptive filtering, as detailed below. The main idea is to improve upon the convergence speed of adaptive filtering process, so that AFC can be achieved with a small processing delay for the HAid processing unit $G(z)$.

4. Proposed approach

The proposed approach is drawn in Fig. 4. The improvement in convergence of both $W(z)$ and $H(z)$ is achieved by hybrid adaptive filtering approach comprising the (delay-based) NLMS algorithm and MVC-based processing.

4.1. Implementation structure

In signal processing community, it is customary to consider a gain-delay model for the HAid processing unit, i.e., $G(z) = Kz^{-\Delta}$, where K is the gain and Δ is the allowable delay in the feedforward block $G(z)$ of the HAid. Using mixed notation for the sake of convenience, the output of HAid feedforward processing unit $G(z)$ can be expressed as

$$x(n) = G(z)u(n), \tag{7}$$

where the HAid input $u(n)$ is derived from the error signal of second AF $H(z)$. At the receiver (loudspeaker) side, an uncorrelated modeling signal $v(n)$ is injected and a delay unit (see block z^{-D}) is inserted. Thus, the composite received signal (at the loudspeaker) is given by $x(n-D) + v(n-D)$. This composite signal is received by the HAid user as well as fed back via the AFP $F(z)$ towards the input microphone. Therefore, the input microphone signal $m(n)$ can now be expressed as

$$m(n) = s(n) + x_f(n) + v_f(n), \tag{8}$$

where $x_f(n) = f(n) * x(n-D)$ and $v_f(n) = f(n) * v(n-D)$ are the feedback components due to $x(n)$ and $v(n)$, respectively. and where $f(n)$ stands for the impulse response of the leakage AFP $F(z)$.

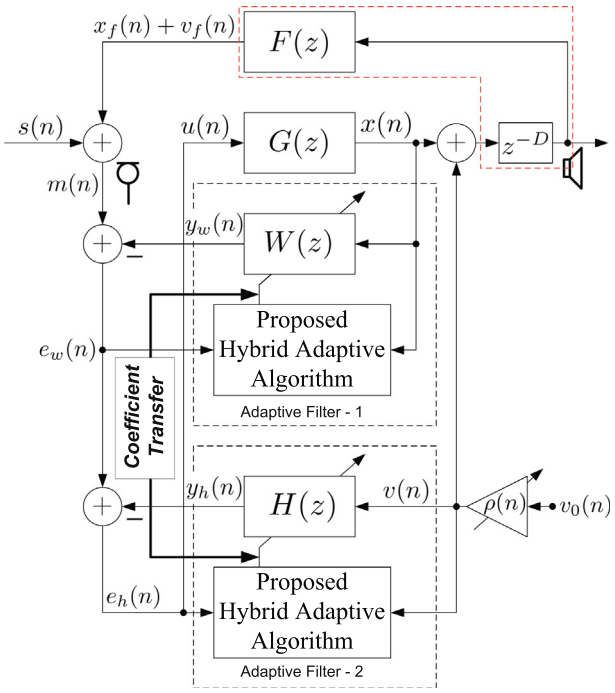


Fig. 4. Block diagram for proposed method.

Without loss of generality, the AFP $F(z)$ is assumed to be an FIR filter of length L.

The delay block z^{-D} inserted in the path for the receiver loudspeaker signal increases the effective length of the overall feedback path. For example, the signal $x(n)$ travels via the inserted delay unit and the AFP $F(z)$ before arriving at the location of the input microphone. Using $x(n)$ as input and $m(n)$ as a desired response, the task of AF $W(z)$ is to identify both the inserted delay as well as the AFP $F(z)$. The AF $W(z)$ is, therefore, an FIR filter of length $D+L$ and its coefficient vector can be written as

$$\boldsymbol{w}(n) = \begin{bmatrix} \boldsymbol{w}_{\mathrm{D}}(n) \\ \boldsymbol{w}_{\mathrm{F}}(n) \end{bmatrix}_{(D+L)\times 1}, \tag{9}$$

where the initial coefficients $\boldsymbol{w}_{\mathrm{D}}(n) = [w_0(n), w_1(n), \ldots, w_{D-1}(n)]^T$ attempt to model the inserted delay block, and the coefficients $\boldsymbol{w}_{\mathrm{F}}(n) = [w_D(n), w_{D+1}(n), \ldots, w_{D+L-1}(n)]^T$ attempt to estimate the coefficients of the AFP $F(z)$. The desired solution for the initial coefficients $\boldsymbol{w}_{\mathrm{D}}(n)$ of $W(z)$, is known a priori as these coefficients are expected to converge to zero to model z^{-D}.

The location of the delay z^{-D} dictates that the uncorrelated modeling signal $v(n)$ also travels via the overall path comprising inserted delay z^{-D} followed by the AFP $F(z)$. Therefore, the second AF $H(z)$ is also considered to be an extended-length FIR filter having coefficient vector

$$\boldsymbol{h}(n) = \begin{bmatrix} \boldsymbol{h}_{\mathrm{D}}(n) \\ \boldsymbol{h}_{\mathrm{F}}(n) \end{bmatrix}_{(D+L)\times 1}, \tag{10}$$

where the initial coefficients $\boldsymbol{h}_{\mathrm{D}}(n) = [h_0(n), h_1(n), \ldots, h_{D-1}(n)]^T$ attempt to model the delay block z^{-D}, and the latter coefficients $\boldsymbol{h}_{\mathrm{F}}(n) = [h_D(n), h_{D+1}(n), \ldots, h_{D+L-1}(n)]^T$ attempt to the identify the AFP $F(z)$.

4.2. Adaptation of adaptive filter $W(z)$

The output $y_w(n)$ of the first AF $W(z)$ is

$$y_w(n) = \boldsymbol{w}^T(n)\boldsymbol{x}(n), \tag{11}$$

where $\boldsymbol{w}(n) = \left[\boldsymbol{w}_{\mathrm{D}}^T(n)\boldsymbol{w}_{\mathrm{F}}^T(n)\right]^T = [w_0(n), w_1(n), \ldots, w_{D+L-1}(n)]^T$ is the coefficient vector (see Eq. (9)) and $\boldsymbol{x}(n) = [x(n), x(n-1), x(n-2), \ldots, x(n-(D+L-1))]^T$ is extended-length signal vector comprising $D+L$ recent samples of the received signal $x(n)$. Considering the extended-length nature of the AF $W(z)$, the corresponding signal vector $\boldsymbol{x}(n)$ can be thought of being comprised of two parts as $\boldsymbol{x}(n) = \left[\boldsymbol{x}_{\mathrm{D}}^T(n)\boldsymbol{x}_{\mathrm{F}}^T(n)\right]^T$ where $\boldsymbol{x}_{\mathrm{D}}(n) = [x(n), x(n-1), \ldots, x(n-(D-1))]^T$, and $\boldsymbol{x}_{\mathrm{F}}^T(n) = [x(n-D), \ldots, x(n-(D+L-1))]^T$.

The composite signal $m(n)$ at the input microphone (see Eq. (8)) is used as a desired response for the AF $W(z)$. Therefore, the error signal $e_w(n)$ needed for $W(z)$ adaptation is computed as

$$\begin{aligned} e_w(n) &= m(n) - y_w(n) \\ &= s(n) + v_f(n) + \left[f(n) * x(n-D) - \boldsymbol{x}^T(n)\boldsymbol{w}(n)\right], \end{aligned} \tag{12}$$

where $v_f(n)$ is an uncorrelated component; however, $s(n)$ is not uncorrelated with the other components. This may cause a biased convergence (as in the conventional approach) and must be addressed (as discussed later). The coefficients of the AF $W(z)$ are updated using the proposed hybrid algorithm, which combines the delay-based NLMS algorithm [27,25] and the gradient information computed from MVC-based processing [28], as

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \mu_w(n)\Delta\boldsymbol{W}_{\mathrm{NLMS}}(n) + \tilde{\mu}_w\Delta\boldsymbol{W}_{\mathrm{MVC}}(n), \tag{13}$$

where $\Delta\boldsymbol{W}_{\mathrm{NLMS}}(n)$ denotes the NLMS algorithm-based gradient information

$$\Delta\boldsymbol{W}_{\mathrm{NLMS}}(n) = \frac{e_w(n)\boldsymbol{x}(n)}{||\boldsymbol{x}(n)||^2 + \epsilon}, \tag{14}$$

and $\mu_w(n)$ is the normalized variable step-size (NVSS) obtained using the delay-based strategy [27]

$$\mu_w(n) = \begin{cases} \frac{\hat{N}_{\mathrm{D}_w}(n)}{P_{e_w}(n)+\epsilon}; & \frac{\hat{N}_{\mathrm{D}_w}(n)}{P_{e_w}(n)} > \mu_{w_{\min}} \\ \mu_{w_{\min}}; & \text{otherwise} \end{cases}, \tag{15}$$

where $\mu_{w_{\min}}$ is empirically selected lower bound on the step-size $\mu_w(n), P_{e_w}(n)$ estimates the power of error signal $e_w(n)$ obtained via a lowpass estimator (with a forgetting factor λ) as

$$P_{e_w}(n) = \lambda P_{e_w}(n-1) + (1-\lambda)e_w^2(n), \tag{16}$$

and $\hat{N}_{\mathrm{D}_w}(n)$ is a parameter indicating the convergence status of the AF. It is computed using initial coefficients of $W(z)$ (modeling the inserted delay), $\boldsymbol{w}_{\mathrm{D}}(n)$, and the corresponding signal vector $\boldsymbol{x}_{\mathrm{D}}(n)$ as [27]

$$\hat{N}_{\mathrm{D}_w}(n) = \lambda\hat{N}_{\mathrm{D}_w}(n-1) + (1-\lambda)\frac{\left(\|\boldsymbol{w}_{\mathrm{D}}(n)\|\|\boldsymbol{x}_{\mathrm{D}}(n)\|\right)^2}{D}, \tag{17}$$

where lowpass estimation (with a forgetting λ) is employed to smooth out any random fluctuations, as in Eq. (16).

In Eq. (13), $\tilde{\mu}_w$ is a FSS parameter, and $\Delta\boldsymbol{W}_{\mathrm{MVC}}(n)$ is calculated using the normalized version of MVC-based algorithm as [28]

$$\Delta\boldsymbol{W}_{\mathrm{MVC}}(n) = \frac{\mathrm{sgn}\{e_w(n)\}|e_w(n)|^{p-1}\boldsymbol{x}(n)}{\left[1+\tau_w\left(|e_w(n)|/\|\boldsymbol{x}(n)\|_p\right)^p\right]^2\left(\|\boldsymbol{x}(n)\|_p^p + \epsilon\right)}, \tag{18}$$

where $|\cdot|$ denotes the absolute value of quantity inside, $\|\cdot\|_p$ denotes the pth norm, $\tau_w = (2\theta_w)^{-p}$, and where p is a shape parameter which signifies the non-Gaussian nature of the signals involved. In [28], it has been recommended to compute θ_w as

$$\theta_w = \begin{cases} C_w; & C_w\overline{e_w}(n) > C_w \\ 0; & C_w\overline{e_w}(n) < 0 \\ C_w\overline{e_w}(n); & \text{otherwise} \end{cases}, \tag{19}$$

where C_w is an empirically chosen constant, and $\overline{e_w}(n)$ is estimated as

$$\overline{e_w}(n) = \lambda\overline{e_w}(n-1) + (1-\lambda) \times \min\{|e_w(n)|, \ldots, |e_w(n-N+1)|\}, \tag{20}$$

where N is length of the estimation window [28].

4.3. Adaptation of the adaptive filter $H(z)$

The input signal and the desired response for the AF $H(z)$ are the injected modeling signal $v(n)$ and the error signal $e_w(n)$ of $W(z)$, respectively. The output $y_h(n)$ of the AF $H(z)$ is given as

$$y_h(n) = \boldsymbol{h}^T(n)\boldsymbol{v}(n), \tag{21}$$

where $\boldsymbol{h}(n) = \left[\boldsymbol{h}_{\mathrm{D}}^T(n)\boldsymbol{h}_{\mathrm{F}}^T(n)\right]^T = [h_0(n), h_1(n), \ldots, h_{D+L-1}(n)]^T$ is the coefficient vector (see Eq. (10)) and $\boldsymbol{v}(n) = \left[\boldsymbol{v}_{\mathrm{D}}^T(n) \;\; \boldsymbol{v}_{\mathrm{F}}^T(n)\right]^T = [v(n), v(n-1), v(n-2), \ldots, v(n-(D+L-1))]^T$ is extended-length signal vector collecting $D+L$ recent samples of the modeling signal $v(n)$. Here $\boldsymbol{v}_{\mathrm{D}}(n) = [v(n), v(n-1), \ldots, \; x(n-(D-1))]^T$, and $\boldsymbol{v}_{\mathrm{F}}^T(n) = [v(n-D), \ldots, v(n-(D+L-1))]^T$ are component vectors corresponding to $\boldsymbol{h}_{\mathrm{D}}(n)$ and $\boldsymbol{h}_{\mathrm{F}}(n)$, respectively.

Using the microphone signal $e_w(n)$ Eq. (12) and output $y_h(n)$ Eq. (21), the error signal $e_h(n)$ is calculated as

$$\begin{aligned} e_h(n) \;\; &= e_w(n) - y_h(n) = s(n) + \left[v_f(n) - \boldsymbol{h}^T(n)\boldsymbol{v}(n)\right] \\ &+ \left[f(n) * x(n-D) - \boldsymbol{w}^T(n)\boldsymbol{x}(n)\right]. \end{aligned} \tag{22}$$

The AF $H(z)$ is adapted using the proposed hybrid algorithm as

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu_h(n)\Delta\boldsymbol{H}_{\mathrm{NLMS}}(n) + \tilde{\mu}_h\Delta\boldsymbol{H}_{\mathrm{MVC}}(n), \tag{23}$$

where $\Delta\boldsymbol{H}_{\mathrm{NLMS}}(n)$ denotes NLMS algorithm-based gradient information [27,25]

$$\Delta\boldsymbol{H}_{\mathrm{NLMS}}(n) = \frac{e_h(n)\boldsymbol{v}(n)}{||\boldsymbol{v}(n)||^2 + \epsilon}, \tag{24}$$

and $\mu_h(n)$ is the corresponding NVSS which is computed using a similar approach as for the NVSS $\mu_w(n)$ for $W(z)$. The necessary steps in computing $\mu_h(n)$ are summarized below [27]

$$P_{e_h}(n) = \lambda P_{e_h}(n-1) + (1-\lambda)e_h^2(n), \tag{25a}$$

$$\hat{N}_{\mathrm{D}_h}(n) = \lambda\hat{N}_{\mathrm{D}_h}(n-1) + (1-\lambda)\frac{(||\boldsymbol{h}_{\mathrm{D}}(n)||||\boldsymbol{v}_{\mathrm{D}}(n)||)^2}{D}, \tag{25b}$$

$$\mu_h(n) = \begin{cases} \frac{\hat{N}_{\mathrm{D}_h}(n)}{P_{e_h}(n)+\epsilon}; & \frac{\hat{N}_{\mathrm{D}_h}(n)}{P_{e_h}(n)} > \mu_{h_{\min}} \\ \mu_{h_{\min}}; & \text{otherwise} \end{cases}, \tag{25c}$$

where $\mu_{h_{\min}}$ is empirically selected lower bound on the step-size $\mu_h(n)$.

In (23), $\tilde{\mu}_h$ is the FSS parameter and $\Delta\boldsymbol{H}_{\mathrm{MVC}}(n)$ is calculated using the standard MVC-based algorithm as [28]

$$\Delta\boldsymbol{H}_{\mathrm{MVC}}(n) = \frac{\mathrm{sgn}\{e_h(n)\}|e_h(n)|^{p-1}\boldsymbol{v}(n)}{\left[1+\tau_h|e_h(n)|^p\right]^2}, \tag{26}$$

where the parameter τ_h is computed in a similar fashion as τ_w computed for $W(z)$. These computations are summarized as follows:

$$\overline{e_h}(n) = \lambda\overline{e_h}(n-1) + (1-\lambda)\min\{|e_h(n)|, \ldots, |e_h(n-N+1)|\}, \tag{27a}$$

$$\theta_h = \begin{cases} C_h; & C_h\overline{e_h}(n) > C_h \\ 0; & C_h\overline{e_h}(n) < 0 \\ C_h\overline{e_h}(n); & \text{otherwise} \end{cases}, \tag{27b}$$

$$\tau_h = (2\theta_h)^{-p}, \tag{27c}$$

where C_h is an empirical constant.

4.3.1. A few remarks

1. Assuming that everything works as per plan and both AFs converge, i.e., $W(z)$ converges to take care of acoustic feedback due to $x(n)$ and $H(z)$ takes care of acoustic feedback due to the modeling signal $v(n)$. It is, therefore, expected that $y_w(n) \to x_f(n), y_h(n) \to v_f(n), \Rightarrow \;\; e_h(n) \approx \hat{s}(n)$, and hence, $u(n) = e_h(n)$ is used as an input to the HAid feedforward signal processing block $G(z)$. It is, therefore, very important that both $W(z)$ and $H(z)$ give good modeling accuracy for the composite path consisting of the inserted delay and the AFP $F(z)$.
2. While developing hybrid algorithm for the first AF $W(z)$, normalized version of MVC algorithm [28] is employed to compute $\Delta\boldsymbol{W}_{\mathrm{MVC}}(n)$ in Eq. (18). As noted in Eq. (18), here the gradient information employs normalization with respect to both the signal $x(n)$ (input for $W(z)$) and the error signal $e_w(n)$ which is computed by using microphone signal $m(n)$ as a desired response. The main component of the $m(n)$ is the input source signal $s(n)$. Furthermore, $x(n)$ is (expected to be) an amplified version of the input signal $s(n)$. Therefore, it is seen that $W(z)$ is having strong (non-white) speech signals in both in its input as well as error, and hence normalization with respect to both in Eq. (18) is justified.
3. The AF $H(z)$ is excited by an uncorrelated modeling signal $v(n)$ computed from injected random signal $v_0(n)$, and its adaptation error signal $e_h(n)$ is computed by using $e_w(n)$ (from AF $W(z)$) as a desired response. Therefore, only the error signal $e_h(n)$ carries (non-white) speech signal information,[1] and hence, gradient normalization is needed to be performed only with respect to $e_h(n)$. Therefore, the hybrid algorithm for $H(z)$ employs the standard MVC algorithm [28] to compute $\Delta\boldsymbol{H}_{\mathrm{MVC}}(n)$ where normalization is performed with respect to only $e_h(n)$ (see Eq.(26)).
4. It is mentioned earlier that $W(z)$ has fast convergence because its input is derived from the receiver loudspeaker signal $x(n)$ which is expected ideally to be an amplified version of the source signal $s(n)$. However, $W(z)$ may suffer from the biased convergence and therefore, its adaptation must be stopped as soon as it achieves a reasonable convergence.
5. An approach to perform convergence monitoring is discussed below in order to avoid a biased convergence of $W(z)$ and ensure good performance from the two AFs in identifying the AFP $F(z)$.

4.4. Convergence monitoring and controlling the modeling signal

As stated earlier, the initial coefficients $\boldsymbol{w}_{\mathrm{D}}(n)$ and $\boldsymbol{h}_{\mathrm{D}}(n)$ in $W(z)$ and $H(z)$, respectively, attempt to model the inserted delay z^{-D}. Since the adaptive algorithm would spread the adaptation error over the entire set of coefficients [27], by monitoring convergence of these coefficients in fact gives us information about the convergence of the overall extended-length AFs $W(z)$ and $H(z)$. Essentially, we compute the following parameters using the initial coefficients $\boldsymbol{w}_{\mathrm{D}}(n)$ and $\boldsymbol{h}_{\mathrm{D}}(n)$,

[1] In fact, it is expected and desired that $e_h(n) \approx \hat{s}(n)$, as explained in Remark 1.

$$\alpha_w(n) = \frac{\sum_{i=0}^{D-1} |w_i(n)|^2}{D}, \tag{28a}$$

$$\alpha_h(n) = \frac{\sum_{i=0}^{D-1} |h_i(n)|^2}{D}, \tag{28b}$$

respectively. Let us assume that the initial coefficients in the two AFs are all initialized by 1's, i.e., $\boldsymbol{w}_{\mathrm{D}}(0) = \boldsymbol{1}$ and $\boldsymbol{h}_{\mathrm{D}}(0) = \boldsymbol{1}$, where $\boldsymbol{1}$ indicates vector of all entries being equal to 1. Since coefficients corresponding to the appended delay are expected to converge to zero, it is straightforward to conclude that initially $\alpha_w(0) = 1$ and $\alpha_h(0) = 1$, and $\alpha_w(n) \to 0$ and $\alpha_h(n) \to 0$, as the two AFs converge. Therefore, it is easy to decide a threshold $0 < T_1 \ll 1$ on $\alpha_w(n)$ and $\alpha_h(n)$, indicating the convergence status of the two AFs.

While briefing the previous method, it has been stated that the convergence of $W(z)$ is faster than that of the $H(z)$ at the start-up. This can be checked by $\alpha_w(n) < \alpha_h(n)$, and if so, the coefficient attempting to model $F(z)$ are copied from $\boldsymbol{w}_{\mathrm{F}}(n)$ to $\boldsymbol{h}_{\mathrm{F}}(n)$. This speeds up the convergence speed of the AF $H(z)$ which is otherwise excited by a modeling signal $v(n)$ which is of low-level as compared with that of the input $x(n)$ to the AF $W(z)$) - note that ideally $x(n)$ is an amplified version of $u(n)$.

Table 1
List of speech signals used in computer simulations.

S01	Male speech (Japanese)
S02	Female speech (Japanese)
S03	Male–Female conversation (English) with background music
S04	Male news (English)
S05	Female speech (Japanese)
S06	Female news (English) with background noise
S07	Female news (English) with background noise
S08	Male speech (English)

As $\alpha_w(n)$ approaches a certain threshold T_1 indicating reasonable convergence of $W(z)$, and if $H(z)$ converges too (if $\alpha_h(n) < 1$), the adaptation of $W(z)$ is stopped. At this point an estimate of AFP, $\hat{\boldsymbol{f}}(n) \overset{z}{\leftrightarrow} \hat{F}(z)$, is obtained from $W(z)$ as $\hat{\boldsymbol{f}}(n) = \boldsymbol{w}_{\mathrm{F}}(n)$. The AF $H(z)$ is on the other hand continuously adapted as it is using an input signal $v(n)$ which is uncorrelated with the source signal. The estimation improvements of $F(z)$ by the AF $H(z)$ can now be directly monitored by computing normalized squared deviation (NSD) being defined for $H(z)$ as

$$\mathrm{NSD}_h(n) = 10 \log \left\{ \frac{\|\hat{\boldsymbol{f}}(n) - \boldsymbol{h}_{\mathrm{F}}(n)\|^2}{\|\hat{\boldsymbol{f}}(n)\|^2} \right\}, \tag{29}$$

where the subscript h signifies the computation using latter coefficients $\boldsymbol{h}_{\mathrm{F}}(n)$ of AF $H(z)$. After the initial estimate $\hat{\boldsymbol{f}}(n)$ is available, the adaptation of $W(z)$ freezes and biased convergence is avoided. The parameters $\alpha_w(n)$ and $\alpha_h(n)$ are computed as in (28). If $\alpha_h(n) < \alpha_w(n)$ then coefficients of $\boldsymbol{h}_{\mathrm{F}}(n)$ are transferred to $\boldsymbol{w}_{\mathrm{F}}(n)$, and vice versa.

Based on our simulations, the stability is ensured for a 10∼30 dB modeling accuracy of $F(z)$, and hence a threshold T_2 (for $\mathrm{NSD}_h(n)$) can be selected accordingly. The AFP has changed when $\mathrm{NSD}_h(n) > T_2$ and $P_{e_h}(n) > T_3$ (T_3 is a threshold for power of the error signal $e_h(n)$, and estimated from the past values of $P_{e_h}(n)$ (25a)). In this case, $W(z)$ and $H(z)$ are re-initialized such that a new estimate for $\hat{\boldsymbol{f}}(n)$ can be obtained.

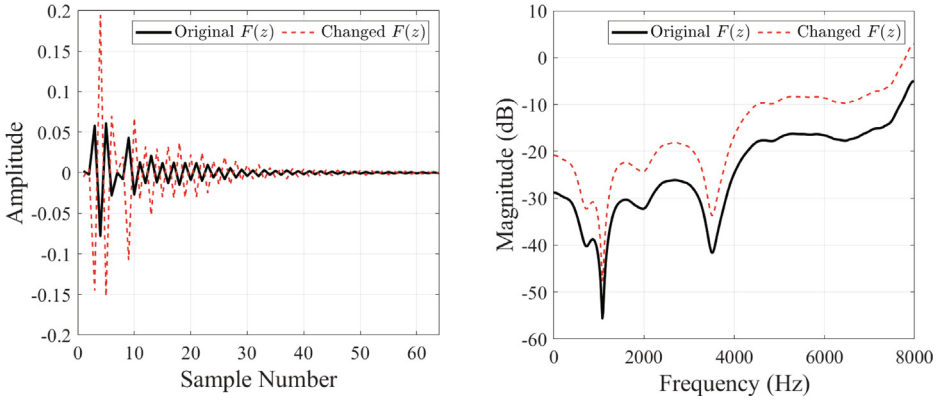


Fig. 5. The impulse response (left), and the magnitude response (right) characteristic of AFP $F(z)$ adopted from [33].

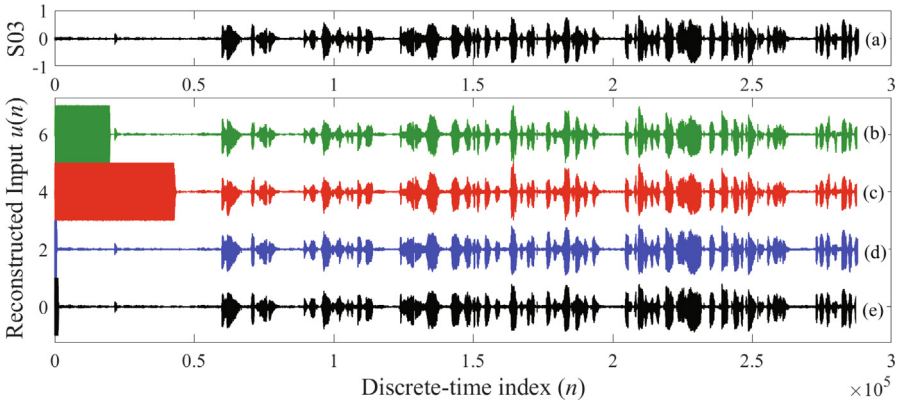


Fig. 6. The reconstructed signal $u(n)$ at the input of HAid feedforward signal processing block $G(z)$ in (b) Conventional, (c) Basic, (d) Previous, and (e) Proposed methods, in comparison with the (a) speech signal S03.

Finally, the modeling signal $v(n)$ is obtained from a white Gaussian noise $v_0(n)$ [25,26]

$$v(n) = \rho(n)v_0(n), \tag{30}$$

where $\rho(n)$ is the time-varying gain parameter calculated as follows

$$\rho(n) = \frac{\alpha_h(n)}{\alpha_h(n) + C}, \tag{31}$$

where C is an positive constant. When $\alpha_h(n)$ is large (indicating $H(z)$ is far from convergence) $v(n)$ tends to $v_0(n)$. On the other hand, when $\alpha_h(n)$ is small, the gain $\rho(n)$ and hence the modeling signal $v(n)$ is small. In summary, the modeling signal $v(n)$ is regulated according to the convergence status of the AF $H(z)$. As soon as system converges, the modeling signal is reduced.

5. Simulation results

This section presents simulation for various methods as follows:

- The conventional method using classical NLMS algorithm (see Fig. 2).
- The basic method using two AFs and employing an uncorrelated modeling signal for an unbiased convergence as reported in [32]. This method employs a delay in the path for $v(n)$ only, and no gain control is used for the modeling signal. Therefore, the modeling signal must be a low-level random signal.
- The previous method [25,26] employing two AFs along with a gain control strategy for the modeling signal. A delay is appended in the path for the overall received signal (combination of the HAid output and the uncorrelated modeling signal) and hence both AFs are extended length FIR filters. In fact, the previous methods provides the basis for the method developed in this paper.

Table 2
Scores for PESQ, NMSE, and SNR_{out} averaged over all speech signals for the HAid gain $K = 10$.

		PESQ	NMSE	SNR_{out}
Conventional	Mean	3.8023	−3.0651	–
	SD	0.5363	2.4771	–
	Median	3.9517	−2.4227	–
Basic Method	Mean	3.7618	−4.2435	34.6625
	SD	0.5594	2.2481	0.7476
	Median	3.9407	−4.4268	34.6000
Previous Method	Mean	4.2974	-14.5223	**64.2364**
	SD	0.2497	7.6302	1.5510
	Median	4.4029	-17.5765	63.6740
Proposed Method	Mean	**4.3867**	**−17.9959**	**63.9828**
	SD	0.1984	6.1221	1.5162
	Median	4.4500	−17.6917	63.7563

- The proposed method as detailed in this contribution.

The same set of speech signals (see Table 1) is used as in [26]. The impulse and magnitude response characteristics of the AFP $F(z)$ are shown in Fig. 5. The length of acoustic path is $L = 64$. This is the same acoustic path as used in many seminal papers including [10]. The interested reader may get the corresponding data from files available with [33].

The experiments have been performed for two values for gain as $K = 10, 20$. Since the modern digital HAids are equipped with very fast processing capabilities, the value of delay is selected as $\Delta = 10$ samples. Since each method considered in this paper employs a different strategy to address AFC their parameters are adjusted experimentally as follows.

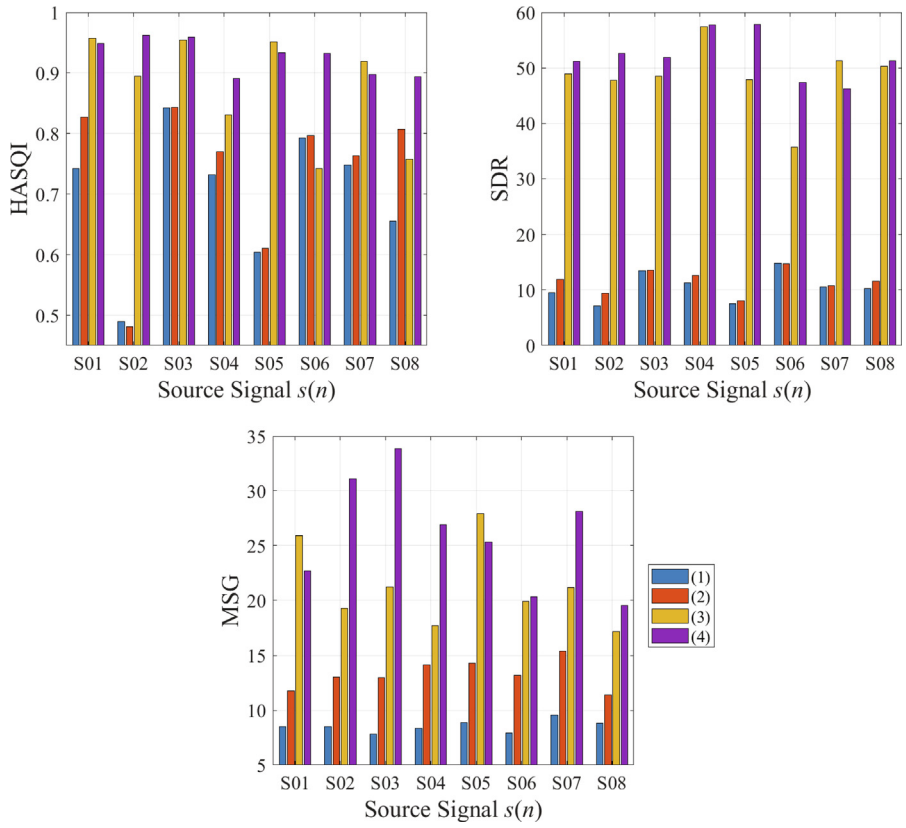


Fig. 7. Scores for HASQI (top-left), SDR (top-right), and MSG (bottom) for experiments with gain $K = 10$ for all speech signals considered in this paper. [Eq. (1) Conventional, Eq. (2) Basic, Eq. (3) Previous, and Eq. (4) Proposed methods].

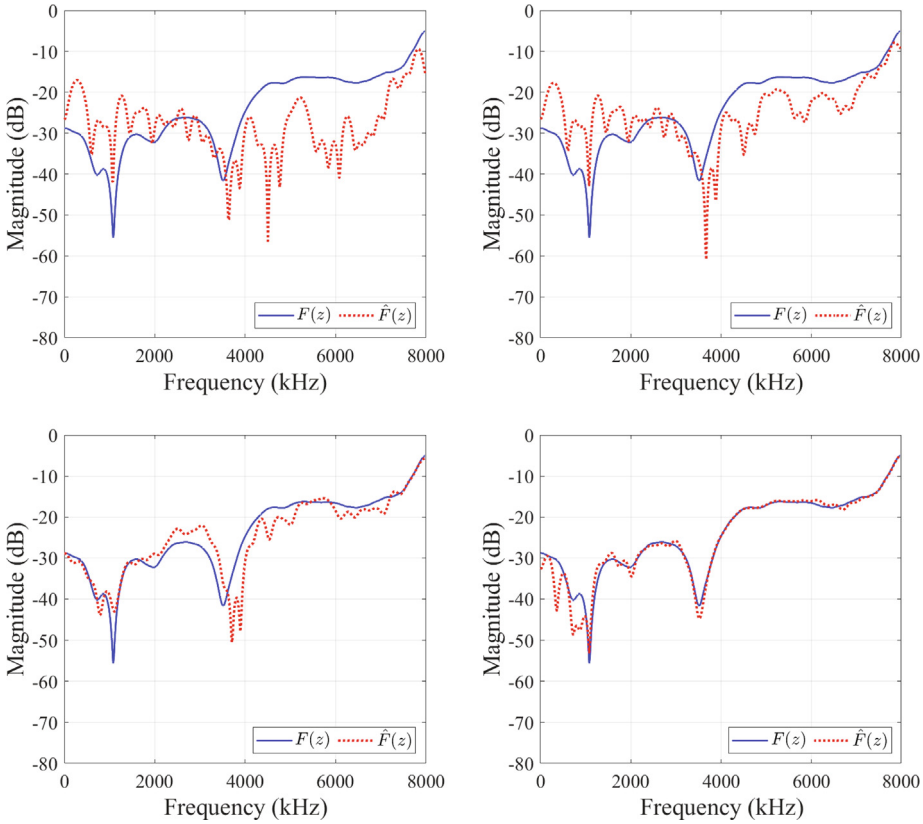


Fig. 8. The magnitude response characteristics of the estimated feedback path $\hat{F}(z)$ in comparison with that of the true AFP $F(z)$ for speech signal S03 in Conventional (top-left), Basic (top-right), Previous (bottom-left), and Proposed (bottom-right) methods.

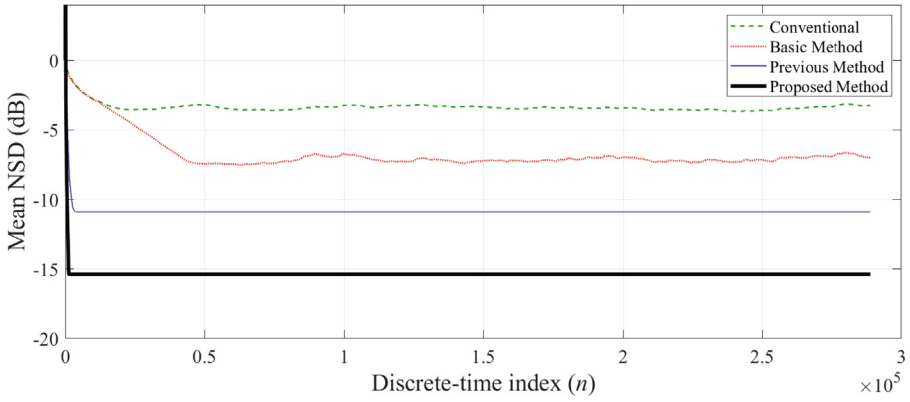


Fig. 9. The curves for NSD (in dB) (averaged over speech signals S01–S08) for experiments with HAid with gain $K = 10$.

1. NLMS-algorithm based conventional method: $\mu = 1 \times 10^{-3}$.
2. Basic method [32]: $\mu_1 = 1 \times 10^{-3}, \text{SNR}_{\text{probe}} = \sigma_v^2/\sigma_x^2 = -15$ dB, $D = 64, \mu_{2_{\min}} = 1 \times 10^{-6}, T_1 = 1 \times 10^{-3}$.
3. Previous method [25,26]: $T_1 = 1 \times 10^{-3}, T_2 = -20$ dB, $T_3 = 10, D = 8, \mu_{\min} = 1 \times 10^{-6}, \text{SNR}_{\text{probe}} = \sigma_v^2/\sigma_s^2 = 0$ dB, $C = 1.5$.
4. Proposed method: $\tilde{\mu}_w = 1 \times 10^{-1}, \tilde{\mu}_h = 1 \times 10^{-3}, C_w = C_h = 1, p = 1.75, N = 8$, and the rest of the parameters are set to the same value as in the Previous method.

It is worth to mention about different values selected for the appended delay D in various methods. This is mainly due to different location of the appended delay block z^{-D} in various methods. As stated earlier, the basic method employs a delay in the path for the modeling signal $v(n)$ (the reader is referred to [32] for details). Hence, D does not affect the overall delay for forward path the input microphone and the received loudspeaker. On the other hand, the delay is appended before the received loudspeaker in the previous and the proposed methods (see Figs. 3 and 4), which increases the overall delay available in the feedforward path. The regularization parameter in all methods is selected as $\epsilon = 1 \times 10^{-4}$. The forgetting factor for various lowpass estimators is also selected as a same value of $\lambda = 0.97$ for various methods. The sampling frequency is $F_s = 16$ kHz.

The following quantitative measures have been used to assess the performance of various methods.

- The HAid speech quality index (HASQI) that has been developed for HAids, with a maximum score of 1.0 indicating a clean signal with no degradation [34,35].

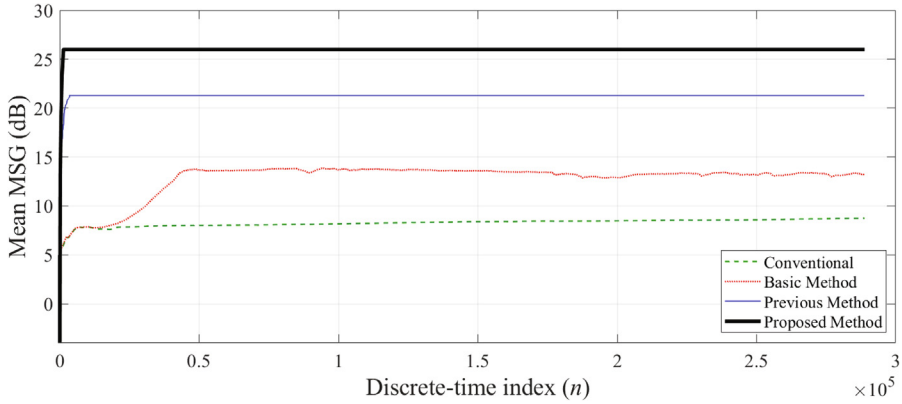


Fig. 10. The curves for MSG (in dB) (averaged over speech signals S01–S08) for experiments with HAid with gain $K = 10$.

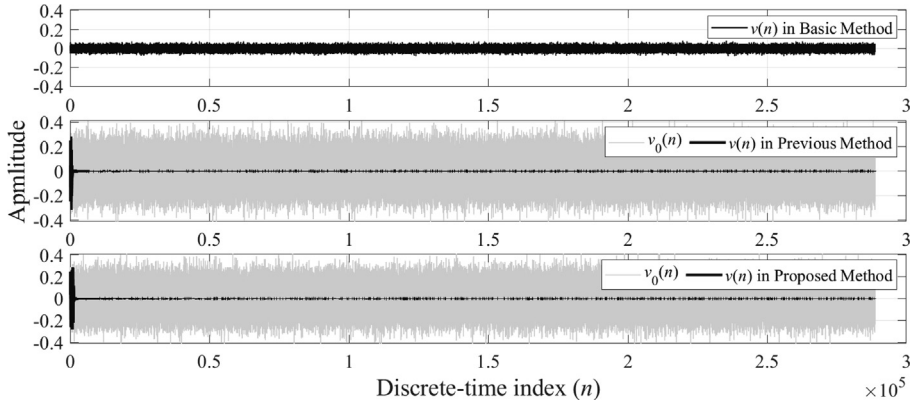


Fig. 11. Variation of modeling signal $v(n)$ for experiments with speech signal S03 in various methods.

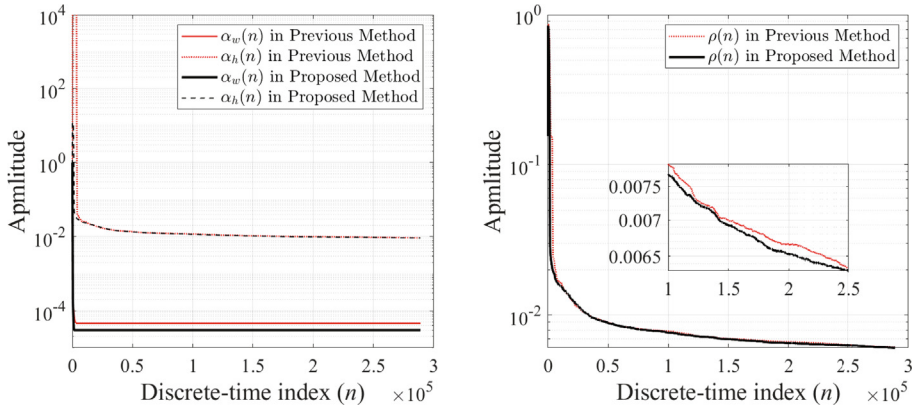


Fig. 12. Plots for variation of (left) parameters $\alpha_w(n)$ and $\alpha_h(n)$ and (right) gain controlling parameter $\rho(n)$ (for $v(n)$) for previous and proposed methods, averaged over all speech signals.

- The signal to distortion ratio (SDR) for nonlinear distortions [36–38].
- The maximum stable gain (MSG) is defined (in dBs) as

$$\text{MSG} = 20\log\left\{\max_{\omega}\|F(\omega) - \hat{F}(\omega)\|^2\right\}. \tag{32}$$

where $F(\omega)$ and $\hat{F}(\omega)$ denote Fourier transform of $\boldsymbol{f}(n)$ of acoustic path, and the estimate $\hat{\boldsymbol{f}}(n)$ obtained from the AFC filter, respectively. The system will only be unstable when the phase at that frequency equals a multiple of 2π [16].

- The perceptual evaluation of speech quality (PESQ) measures the quality of the processed speech in comparison with the true reference. The maximum score is 4.5 for a clean speech signal with no degradation [39].
- The normalized mean squared error (NMSE) is computed between the (true) source signal $s(n)$ and the HAid input signal $u(n)$ to have an overall idea about the quality of reconstructed signal $u(n)$ (which should ideally be same as that of the $s(n)$).
- The Output SNR is calculated as follows

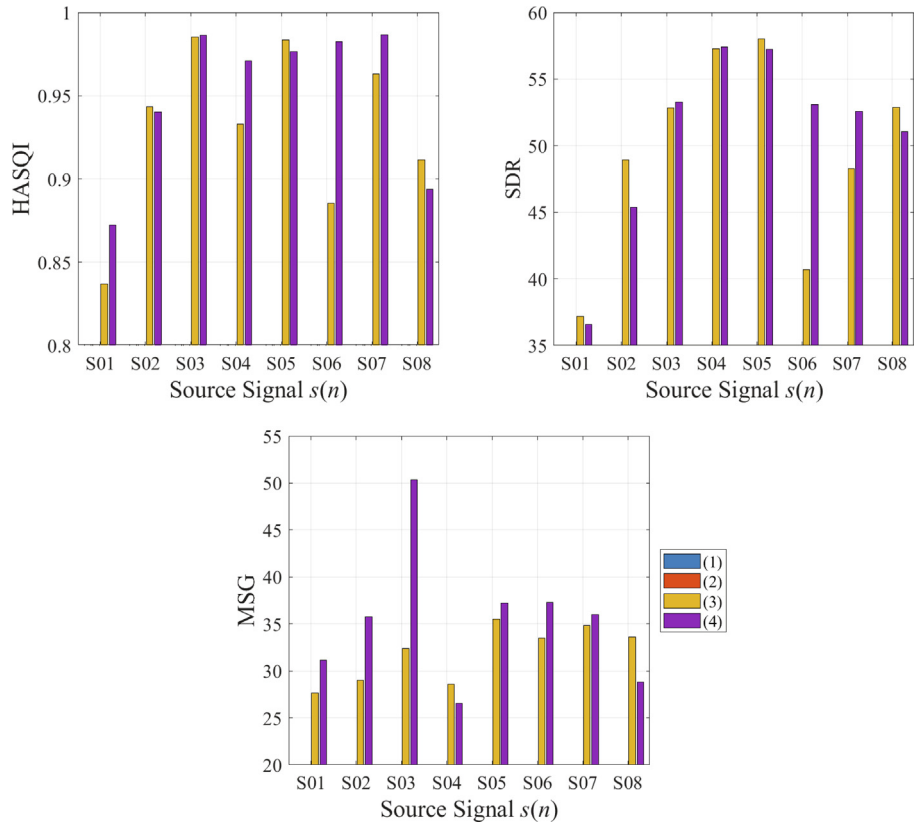


Fig. 13. Scores for HASQI (top-left), SDR (top-right), and MSG (bottom) for experiments with gain $K = 20$ for all speech signals considered in this paper. The description of legend is same as in Fig. 12.

Table 3
Scores for PESQ, NMSE, and SNR_{out} averaged over all speech signals for the HAid gain $K = 20$.

		PESQ	NMSE	SNR_out
Previous Method	Mean	4.4070	−16.5525	**70.2109**
	SD	0.0951	7.0549	1.4726
	Median	4.4380	−13.6482	69.7661
Proposed Method	Mean	**4.4511**	**−19.8444**	**70.1444**
	SD	0.0577	8.8712	1.3463
	Median	4.4848	−22.5150	69.8519

$$\text{SNR}_{\text{out}} = 10 \log \{\sigma_x^2 / \sigma_v^2\}, \tag{33}$$

where σ_x^2 and σ_v^2 denote variances of signals output signal $x(n)$ and modeling signal $v(n)$.

- The misalignment performance for the AFC filter is computed as $\text{NSD}_f(n)$ as follows:

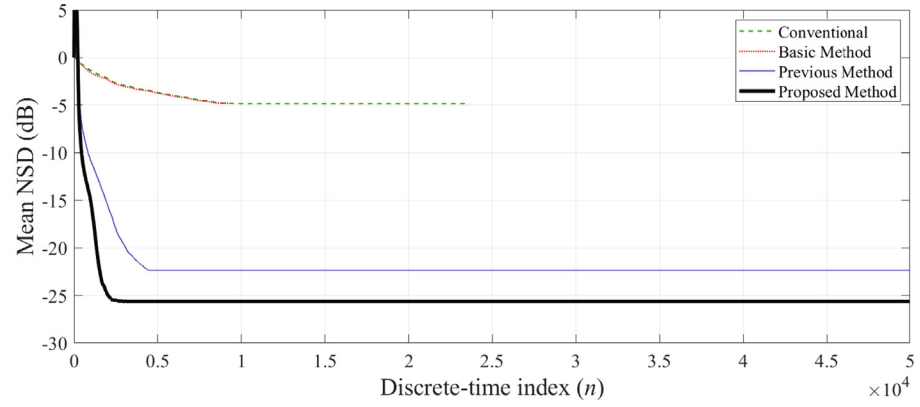
$$\text{NSD}_f(n) = 10 \log \left\{ \frac{\|\boldsymbol{f}(n) - \hat{\boldsymbol{f}}(n)\|^2}{\|\boldsymbol{f}(n)\|^2} \right\}, \tag{34}$$

where $\boldsymbol{f}(n)$ denotes coefficient vector for true AFP $F(z)$ and the $\hat{\boldsymbol{f}}(n)$ is the estimate obtained from the adaptive AFC filter. In the previous and proposed method, $\hat{\boldsymbol{f}}(n)$ is obtained from the (latter) coefficients $\boldsymbol{w}_{\text{F}}(n)$ (see Eq. (9)) of the first adaptive AF $W(z)$.

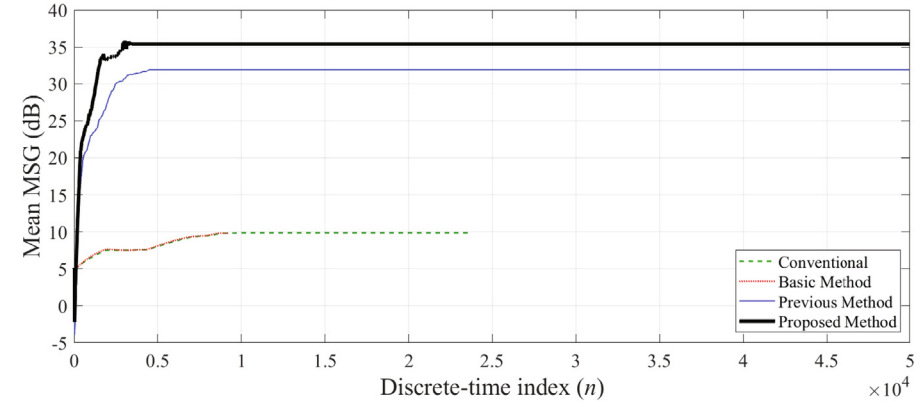
5.1. Case 1: Stationary environment

The objective of the first case study is to understand the behavior of various methods from the view points of initial convergence speed and the steady-state performance. For this purpose, the AFP $F(z)$ is considered as linear time-invariant system with characteristic as shown in Fig. 5 (see thick-black curves marked as 'Original $F(z)$'). First we discuss results for HAids with gain $K = 10$.

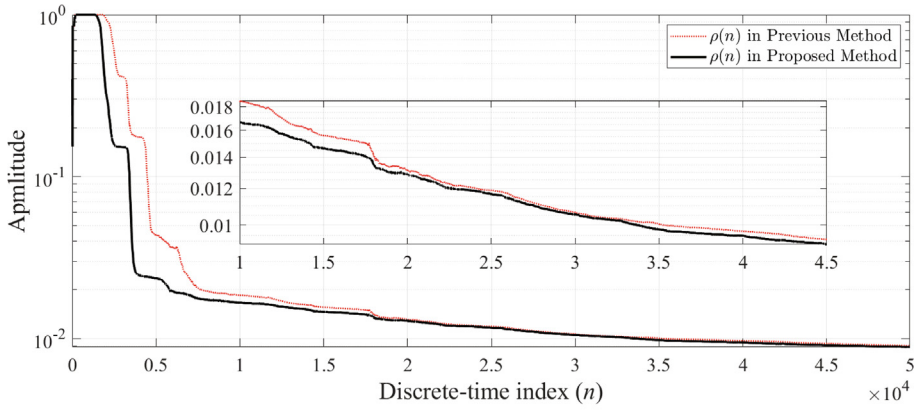
As shown in block diagrams (in Figs. 1–4), the key signal processing objective is to remove (or at least reduce) the acoustic feedback at the input microphone, and hence the reconstructed signal $u(n)$ at the input of HAid processing unit $G(z)$ resembles the input $s(n)$ as close as possible. One naive way to assess this performance is to visually look at the two speech signals. Fig. 6, in fact shows reconstructed signal $u(n)$ in comparison with input speech signal $s(n)$ (denoted as S03 in Table 1). It is observed that the conventional and basic methods suffer from a very long howling period when the AFC system is started. The initial howling periods for the previous and proposed methods are very short and AFC system converges very quickly in these methods. For the typical results shown in Fig. 6, the howling periods are about 1.25 s in the conventional method, 2.72 s in the basic method, 50 ms in the previous method, and 87.5 ms in the proposed method. Furthermore, the conventional method suffers from musical noise even at the steady-state (not obvious from the wave-plots shown in Fig. 6). A detailed comparative analysis in fact shows that proposed method is very efficient from the view point of various performance measures considered in this paper. Fig. 7 shows results for HASQI, SDR, and MSG for all speech signals considered in this paper. It is observed that for these performance measures, the proposed method outperforms all the other considered methods for (almost)



(a)



(b)



(c)

Fig. 14. Simulation results averaged over speech signals S01–S08 for HAid gain $K = 20$. (a) mean NSD (in dB), (b) mean MSG (in dB), and (c) variation of gain $\rho(n)$ for modeling signal $v(n)$.

all input signals. The corresponding results for PESQ, NMSE, and SNR_{out} (averaged over all input signals S01-S08) are tabulated in Table 2, where bold signifies the best performing method(s). It is observed that:

- The conventional method gives lowest scores for PESQ, and shows poor NMSE performance computed between $s(n)$ and $u(n)$.
- The (fixed) modeling signal is a low-level signal in the case of basic method, so that an appreciable SNR_{out} is archived. However, this degrades convergence performance of the AFC filter (discussed later), which in turn results in low scores for PESQ and NMSE as compared with those obtained with the previous and proposed methods.
- The proposed method gives improved performance for PESQ and NMSE, and a comparable performance from the view point of SNR_{out} if compared with the previous method.

As discussed earlier in Section 1, in order to perform AFC the key task for the adaptive AFC filter is to identify the characteristic of 'unknown' AFP. One way to understand this performance is to compare the magnitude response of coefficients of the AFC filter

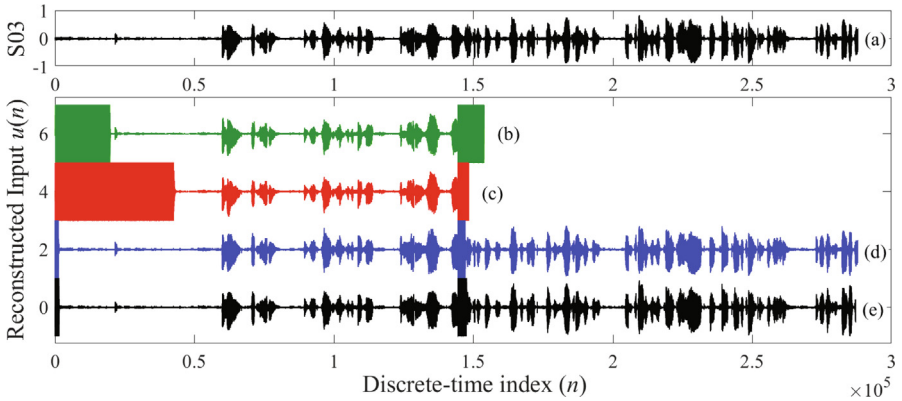


Fig. 15. Simulation results for HAid gain $K = 10$ and a sudden change in the acoustic path during the middle of simulation. (a) Input speech signal S03 and the corresponding reconstructed signal $u(n)$ at the input of HAid processing unit $G(z)$ in (b) Conventional, (c) Basic, (d) Previous, and (e) Proposed methods.

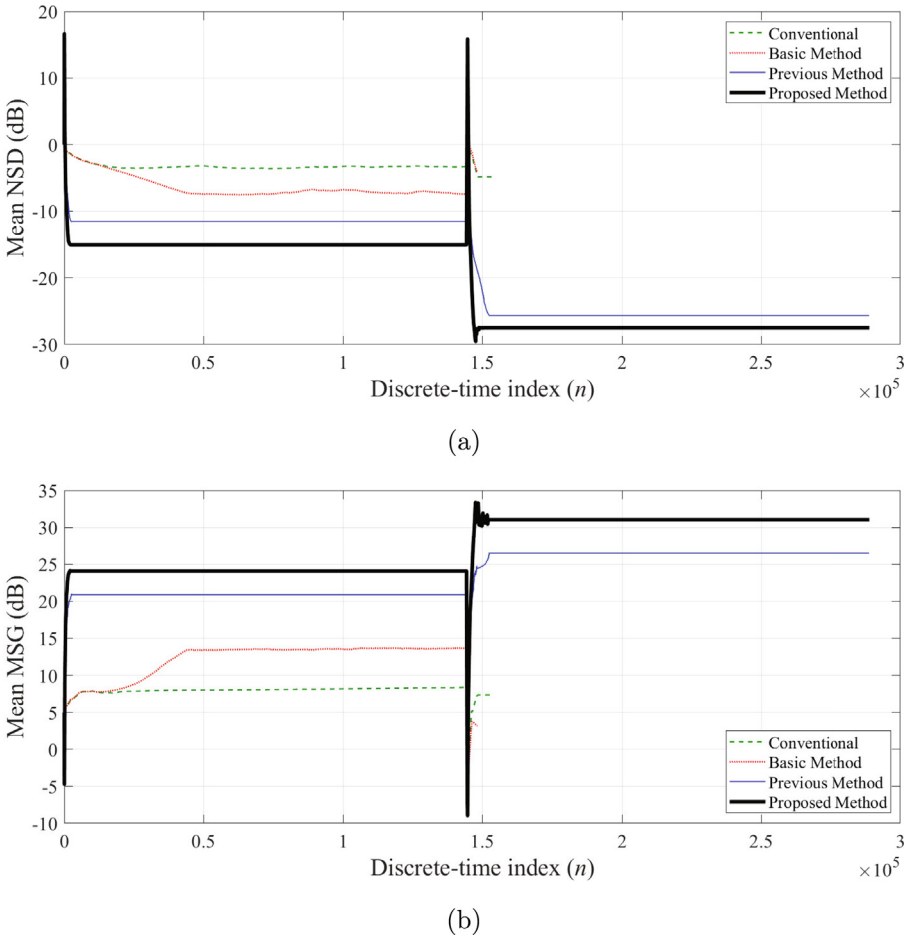


Fig. 16. Simulation results averaged over speech signals S01–S08 for HAid gain $K = 10$ and a sudden change in the acoustic path during the middle of simulation. (a) mean NSD (in dB), and (b) mean MSG (in dB).

with the true AFP. One such comparison for (speech signal S03) is shown in Fig. 8. It is observed that the estimation filter $\hat{F}(z)$ obtained via AFC system in the proposed method well identifies the magnitude response of the true AFP $F(z)$. In order to understand the overall performance from the system identification point of view, the curves for NSD (as defined in Eq. (34)) averaged over all speech signals are shown in Fig. 9. It can be seen that the proposed method gives the best NSD performance among the investigated methods. The corresponding curves for MSG Eq. (32), averaged over all speech signals S01-08 are shown in Fig. 10. It can be seen that the proposed method gives the largest MSG among the investigated methods.

One important advantage in previous and proposed methods is to have improved performance from the view point of SNR_{out} (see Table 2). This implies that the injected modeling signal has in fact been reduced to a very low level. This is depicted in Fig. 11 which

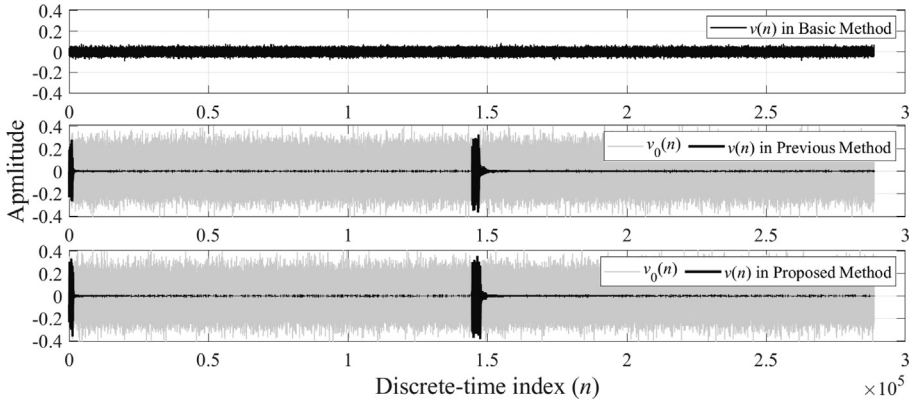


Fig. 17. Variation of modeling signal $v(n)$ for experiments with speech signal S03 in experiments for HAid gain $K = 10$ and a sudden change in the acoustic path during the middle of simulation.

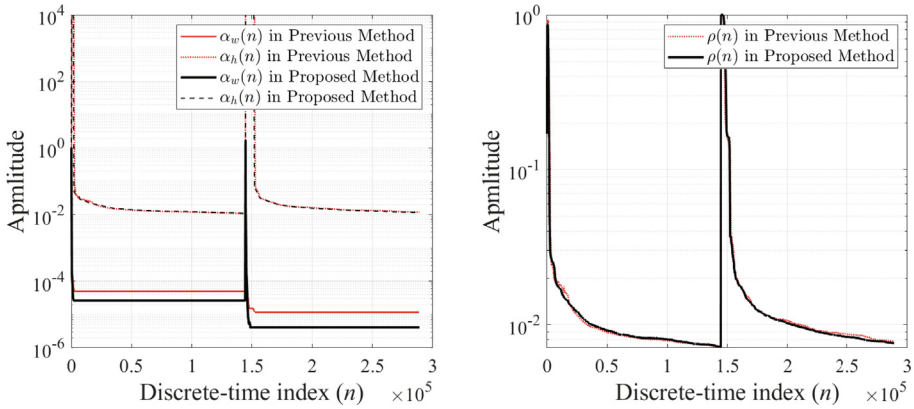


Fig. 18. (left) Variation of parameters $\alpha_w(n)$ and $\alpha_h(n)$ and (right) variation of gain parameter $\rho(n)$ in modeling signal $v(n)$ for previous and proposed methods, averaged over speech signals S01–S08 in experiments for HAid gain $K = 10$ and a sudden change in the acoustic path during the middle of simulation.

plots the variation of the modeling signal $v(n)$. As discussed earlier, a constant low-level modeling signal is used in the basic method to have a good SNR_{out}. The low-level signal, however, affects the convergence speed (see Figs. 9 and 10) and the accuracy (see Table 2 and Fig. 7) achieved by the basic method. A large-level random noise is injected for AFC in both the previous and proposed methods. This gives fast convergence speed for the AFC filter (see Figs. 9 and 10). The level of the modeling signal reduces to a very low level, as the AFC system converges in the previous and the proposed methods. This is achieved by a gain control parameter $\rho(n)$ computed on the basis of parameters $\alpha_w(n)$ and $\alpha_h(n)$, which are plotted in Fig. 12 (averaged over all speech signals). As expected, gain control parameter $\rho(n)$ is adjusted to 1 during the transient state of the AFC system. It gradually reduces (toward zero) as the AFC system converges, and hence, the level of the modeling signal is reduced at the steady-state which in turn improves the SNR_{out}.

The above experiment is repeated for HAid with gain $K = 20$. Fig. 13 shows results for HASQI, SDR, and MSG for all speech signals, and Table 3 presents corresponding averaged results for PESQ, NMSE, and SNR_{out}. It is observed that both NLMS algorithm-based conventional method and the basic method become unstable in this case. One reason for such a behavior is due to a very small processing delay used for the HAid processing unit $G(z)$ (recall $\Delta = 10$ samples). Due to such a small processing delay, the AF signals become strongly correlated when the HAid is operated at a high gain. The performance comparison between the previous and the proposed methods remain similar to that observed in experiments for HAid with gain $K = 10$, i.e., the proposed method gives a better or a similar performance for the various performance measures. The corresponding curves (averaged over all speech signals) for NSD, MSG, and for variation of the gain control parameter $\rho(n)$ are shown in Fig. 14. It is observed that in all these performance measures, the proposed method shows faster convergence speed and a better steady-state performance in comparison with the previous method.

5.2. Case 2: Non-stationary environment

The performance of the considered methods for time-varying acoustic environment is investigated by a sudden change in the AFP $F(z)$. In practice, this might happen when the HAid user takes his/her mobile phone near the ear. The experimental studies have shown that such situation results in increase of the amplitude of the impulse response of the AFP $F(z)$ [26,40]. Let us assume that at the start-up the AFP $F(z)$ is same considered in the previous case study. At the middle of the simulation, the AFP $F(z)$ is changed to having increased-amplitude characteristics as shown in Fig. 5 (see dashed-red curves marked as 'Changed $F(z)$'). For the results presented in this case, the HAid gain is adjusted to $K = 10$.

Fig. 15, shows reconstructed signal $u(n)$ in comparison with input speech signal $s(n)$ (S03 in Table 1). It is observed that the conventional and basic methods are not able to recover from the howling occurring when there is a change in the acoustic path.

The previous and proposed methods, on the other hand, show very robust performance, and can re-converge very quickly after a small period of howling at middle of simulation. In order to understand the overall/averaged performance of various methods, the averaged NSD and MSG curves are plotted in Fig. 16 and the tracking superiority of the proposed method is obvious. The variation of the modeling signal in comparison with the injected random signal is plotted in Fig. 17 for a single speech signal S03. As observed earlier, the modeling signal reduces to a very low level as the AFC converges at the steady-state because the gain control parameter $\rho(n)$ increases its value when the acoustic path changes, and reduces it value when the AFC system converges. Fig. 18 shows the variation of convergence monitoring parameters $\alpha_w(n)$ and $\alpha_h(n)$, and the gain control parameter $\rho(n)$ in the previous and proposed methods. These curves are obtained by performing averaging over all speech signals considered in this paper.

6. Conclusions

Inspired by the results of MVC-based adaptive filtering for non-Gaussian signal processing [28], we have proposed hybrid adaptive filtering to adapt the two AFs working in tandem for an effective AFC system. The simulation results convincingly demonstrate that the proposed approach provided not only a fast convergence speed and tracking behavior but also improved scores for various quantitative measures. Further study is needed to perform convergence analysis of the proposed algorithm.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research has been funded from the Faculty Development Competitive Research Grants Program of Nazarbayev University under the Grant No. 110119FD4525.

References

[1] Bustamante DK, Worrall TL, Williamson MJ. Measurement and adaptive suppression of acoustic feedback in hearing aids. In: Proc. IEEE ICASSP 1989, p. 2017–20. .
[2] Kates JM. Digital hearing aids. Plural Publishing 2008.
[3] Maxwell J, Zurek P. Reducing acoustic feedback in hearing aids. IEEE Trans Speech Audio Process 1995;4:304–13.
[4] Edwards BW. Signal processing techniques for a DSP hearing aid. In: Proc. IEEE ISCAS 1998, vol. VI, p. 586–9 .
[5] Kaelin A, Lindgren A, Wyrsch S. A digital frequency domain implementation of a very high gain hearing aid with compensation for recruitment of loudness and acoustic echo cancellation. Signal Process 1998;64:71–85.
[6] Kates JM. Constrained adaptation for feedback cancellation in hearing aids. J Acoust Soc Am 1999;106:1010–9.
[7] Haykin S. Adative filter theory. 4th ed. New Jersey: Prentice Hall; 2002.
[8] Widrow B, Stearns SD. Adaptive signal processing. New Jersey: Prentice Hall; 1985.
[9] Douglas SC. A family of normalized LMS algorithms. IEEE Signal Process. Lett. 1994;1(3):49–51.
[10] Siqueira MG, Alwan A. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids. IEEE Trans Speech Audio Process 2000;8(4):443–53.
[11] Estermann P, Kaelin A. Feedback cancellation in hearing aids: Results from using frequency-domain adaptive filters. In: Proc. IEEE ISCAS; 1994.pp. 257–60. .
[12] Hellgren J. Analysis of feedback cancellation in hearing aids with filtered-x LMS and the direct method of closed loop identification. IEEE Trans Speech Audio Process 2002;10(2):119–31.
[13] Sakai H, Fukuzono H. Analysis of adaptive filters in feedback cancellation for sinusoidal signals. In Proc. APSIPA-ASC 2009. p. 430–3. .
[14] Vicen-Bueno R, Martínez-Leira A, Gil-Pita R, Rosa-Zurera M. Modified LMS-based feedback-reduction subsystems in digital hearing aids based on WOLA filter bank. IEEE Trans Instrum Meas 2009;58(9):3177–90.
[15] Nakagawa CRC, Nordholm S, Yan WY. Dual microphone solution for acoustic feedback cancellation for assisitve learning. In: Proc. IEEE ICASSP 2012. p. 149–52. .
[16] Nakagawa CRC, Nordholm S, Yan WY. Analysis of two microphone method for feedback cancellation. IEEE Signal Process Lett 2015;22(1):35–9.
[17] Pradham S, George NV, Albu F, Nordholm S. Two microphone acoustic feedback cancellation in digital hearing aids: A step size controlled frequency domain approach. Appl Acoust Mar. 2018;132:142–51.
[18] Schepker H, Tran LTT, Nordholm S, Doclo S. Improving adaptive feedback cancellation in hearing aids using an affine combination of filters. In: Proc. IEEE ICASSP 2016. p. 231–5. .
[19] Nordholm S, Schepker H, Tran LTT, Doclo S. Stability-controlled hybrid adaptive feedback cancellation scheme for hearing aids. J Acoust Soc Am Jan. 2018;143(1):150–66.
[20] Albu F, Nordholm S, Tran LTT. The hybrid simplified Kalman filter for adaptive feedback cancellation. In: Proc. COMM 2018, Bucharest, Romania. p. 45–50. .
[21] Spriet A, Proudler I, Moonen M, Wouters J. Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal. IEEE Trans Signal Process 2005;53(10):3749–63.
[22] Greenberg JE, Zurek PM, Brantley M. Evaluation of feedback-reduction algorithms for hearing aids. J Acoust Soc Am 2000;108(5):2366–76.
[23] Guo M, Jensen SH, Jensen J. Novel acoustic feedback cancellation approaches in hearing aid applications using probe noise and probe noise enhancement. IEEE Trans Audio Speech Lang Process 2012;20(9):2549–63.
[24] Nakagawa CRC, Nordholm S, Yan WY. Feedback cancellation with probe shaping compensation. IEEE Signal Process Lett 2014;21(3):365–9.
[25] Akhtar MT, Nishihara A. Automatic tuning of probe noise for continuous acoustic feedback cancelation in hearing aids. In: Proc. EUSIPCO 2016, August 29 — September 2, 2016, Budapest Hungary. p. 888–92. .
[26] Akhtar MT, Albu F, Nishihara A. Acoustic feedback cancellation in hearing aids using dual adaptive filtering and gain-controlled probe signal. Biomedical Sig Process Control Jul. 2019;52:1–13.
[27] Mader A, Puder H, Schmidt GU. Step-size control for acoustic echo cancellation filters–An overview. Signal Process 2000;4:1697–719.
[28] Huang F, Zhang J, Zhang S. Maximum versoria criterion-based robust adaptive filtering algorithm. IEEE Trans Circuits Sys II: Express Briefs Oct. 2017;64 (10):1252–6.
[29] Radhika S, Albu F, Chandrasekar A. Steady state mean square analysis of standard maximum versoria criterion based adaptive algorithm. IEEE Transactions Circuits Systems II: Express Briefs; 2020.
[30] Akhtar MT. Nishihara A. Two-adaptive filter-based method using gain controlled probe noise for acoustic feedback neutralization in digital hearing aids. In: Proc. Inter. Workshop Acoustic Signal Enhancement (IWAENC2018), Sept. 2018, Tokyo, Japan. p. 176–80. .
[31] Tran LTT, Nordholm SE, Schepker H, Dam HH, Doclo S. Two-microphone hearing aids using prediction error method for adaptive feedback control. IEEE/ACM Trans Audio Speech Lang Process May 2018;26(5):909–23.
[32] Akhtar MT, Nishihara A. Acoustic feedback neutralization in digital hearing aids – A two adaptive filters-based solution. In: Proc. IEEE ISCAS 2013, May 19–23, 2013, Beijing, China. p. 529–32. .
[33] Lee KA, Gan WS, Kuo SM. Subband adaptive filtering: Theory and implementation. Wiley; 2009.
[34] Kates JM, Arehart KH. The hearing-aid speech quality index (HASQI). J Audio Eng Soc 2010;58(5):363–81.
[35] Kates JM, Arehart KH. The hearing-aid speech quality index (HASQI) version 2. J Audio Eng Soc 2014;62(3):99–117.
[36] Manders AJ, Simpson DM, Bell SL. Objective prediction of the sound quality of music processed by an adaptive feedback canceller. IEEE Trans Audio Speech Lang Process 2012;20(6):1734–45.
[37] Olofsson A, Hansen M. Objectively measured and subjectively perceived distortion in nonlinear systems. J Acoust Soc Amer 2006;120(6):3759–69.
[38] Vincent E, Gribonval R, Févotte C. Performance measurement in blind audio source separation. IEEE Trans Audio, Speech Lang Process 2006;14(4):1462–9.
[39] Loizou PC. Speech enhancement theory and practice. CRC Press; 2007.
[40] Sankowsky-Rothe T, Blau M. Static and dynamic measurements of the acoustic feedback path of hearing aids on human subjects. Proc Mtgs Acoust 2017;30 (050008). (doi: 10.1121/2.0000618). .