Received December 26, 2017, accepted February 2, 2018, date of publication February 8, 2018, date of current version April 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2804298

Convex Combined Adaptive Filtering Algorithm for Acoustic Echo Cancellation in Hostile Environments

YING-REN CHIEN, (Senior Member, IEEE), AND JIN LI-YOU
Adaptive and Autonomous Communication Laboratory, Department of Electrical Engineering, National Ilan University, Yilan 26041, Taiwan
Corresponding author: Ying-Ren Chien (yrchien@niu.edu.tw)

This work was supported by the Ministry of Science and Technology of Taiwan under Grant 106-2221-E-197-009.

ABSTRACT Research into combined adaptive filters has been attracting considerable attention; the performance of which is generally superior to that of the individual component filters. Formulating an adaptive filtering algorithm to handle a range of cases can be very difficult, particularly in complicated scenarios. This paper applies a robust convex combination method to the problem of acoustic echo cancellation. According to the minimum mean square error criterion, we first derive the optimal combination factor for a combination adaptive filter with two component filters. Next, we provide an approximation method to the inaccessible optimal combination factor. Finally, we extend this approximated factor to a robust one so that it is applicable even in situations with strong a ambient impulsive noise (IN) and abrupt changes in the echo channel. To do this, we non-linearly transfer the approximated combination factor so that this combination factor exhibits less transient time during the adaptation process of the combination factor and a fewer fluctuations in its steady state. Moreover, we detect the presence of the strong ambient IN and force the combined system output mainly contributed by one of the two component adaptive filters which is less sensitive to the the strong IN. We adopted computer simulations combining two affine projection sign algorithms with different step sizes as an example. Simulation results show that the proposed method is superior to comparable algorithms in terms of the normalized mean squared deviation and echo return loss enhancement.

INDEX TERMS Convex combination, adaptive filter, acoustic echo cancellation (AEC), impulsive noise (IN).

I. INTRODUCTION

Acoustic echo cancellation (AEC) is crucial to the removal of undesired acoustic echoes generated by hands-free audio devices [1]–[4]. When dealing with AEC, adaptive filters are commonly used to identify the path of the acoustic echo between the loudspeaker and microphone in order to generate an electronic replica of the acoustic echo. The echo is then canceled by subtracting the replica from the microphone signal. However, a number of challenging problems associated with AEC impair the performance of existing adaptive filtering techniques. First, the echo path can be quite long and may vary with the surrounding environment [5]. This means that adaptive filters must be very long to deal with this situation. In under-modeling situations, the AEC is unable to produce a replica to cancel the echo signal. Furthermore, long adaptive filters suffer from slow convergence [6], which can be slowed even more in cases involving variations in the acoustic echo due to the movement of objects or human bodies [7]. Second, ambient impulsive noise (IN), which can corrupt the microphone signal, is often strong and highly non-stationary, such that the adaptive filter is unable to converge or causes an increase in the weight misadjustment [8]. Third, when the input of an AEC is a speech signal (i.e., highly correlated), it tends to slow down the convergence of the adaptive filter. Fourth, full-duplex voice communication can result in a double-talk, resulting in the divergence of adaptive filters during the tap weight updating process [9]–[11]. In cases where the echo channel is time-varying, it can be difficult to detect double-talk events [12]. Unfortunately, there is a trade-off between convergence rate and steady-state misadjustment when using conventional adaptive algorithms. Several approaches have been proposed to address this issue. Variable step-size (VSS) algorithms can be used to enable rapid convergence as well as low steady-state misadjustment. VSS algorithms have been applied to various adaptive filtering algorithms, such as the normalized least-mean-square (NLMS) algorithm [13] or the affine projection algorithm (APA) [14]. Instead of using only one step-size to adjust all tap weights, the granularity of step-size control may further down to each tap of adaptive filters to improve performance even more [15], [16]. A generalized proportionate VSS algorithm was proposed in [17] for fast time-varying acoustic environments. Subband adaptive filtering (SAF) algorithms provide another framework by which to accelerate the convergence of adaptive filtering algorithms [18]. Sign algorithms can be used to alleviate the impact of IN on the rate of convergence of adaptive algorithms, including the signed regressor, signed error, and signed-signed algorithms [6]. Unfortunately, sign algorithms tend to reduce the rate of convergence and increase the degree of misadjustment in adaptive algorithms.

The idea of combining adaptive filters was proposed in [19]. This involves combining (either in an affine or convex fashion) two adaptive filters with different features to a single output. The different features could possibly be different adaptation algorithms, different cost functions during the adaptation process, different kernel functions for non-linear filtering problems [20], adaptive filters with different tap length, or the same adaptation algorithm with different step-sizes [21]–[23]. The principle is to combine the advantages of the two adaptive filters, such that the resulting adaptive filter combination has fast convergence and low steady-state misadjustment. For example, [24] combined the least-mean-square (LMS) and zero-attrator LMS (ZA-LMS) algorithms for sparse adaptive filtering applications. The ZA-LMS algorithm has lower steady-state excess mean square error than does the LMS algorithm due to the sparsity of the system; however, when the system is non-sparse, the ZA-LMS algorithm is unable to match the performance of the LMS algorithm. No single adaptive algorithm is ever good enough for the systems with variable sparsity. Das and Chakraborty [24] proposed a combined filter for the semi-sparse case, which converges to the better of the two filters or produces a solution that outperforms both of the constituent filters. Ferrer *et al.* [25] combined two APA algorithms with different projection orders. This approach can be extended to include more APA filters with different projection orders [26]. The combination of adaptive filters can also be extended from fullband to subband adaptive filters [27]–[29].

In this paper, we first mathematically derive the optimal convex combination factor of two adaptive filters. We then propose a practical method to approximate the optimal combination factor, which is unavailable due to the unknown impulse response of the system, such as the acoustic echo channel. By numerical analysis, we can show that our proposed approximation method outperform the comparable related works. We then extend the approximated factor to a robust combination factor to mitigate the impact of strong ambient noise or a time-varying echo channel. The robustness is achieved by properly control the values of the combination factor. We apply a non-linear S-shaped transfer function to the combination factor so that the transient time of adaptation of this combination factor is reduced; meanwhile, in the steady-state, the fluctuation of the combination factor is reduced as well. The other mechanism is to detect the event of the presence of strong ambient noise. Upon the occurrence of such even, we force the contribution to the output of the combined adaptive filter comes from only one of its component filter. The effectiveness of the proposed method is confirmed by computer simulations and an actual hands-free system for automobiles. Our results demonstrate the effectiveness of the proposed convex combined adaptive filtering algorithm for AEC in hostile environments.

The remainder of this paper is organized as follows. Section II describes the system model combining two adaptive filters. Section III details major related works. Section IV presents the proposed convex combined adaptive filtering algorithm. Section V presents simulation results. Conclusions are drawn in Section VI.

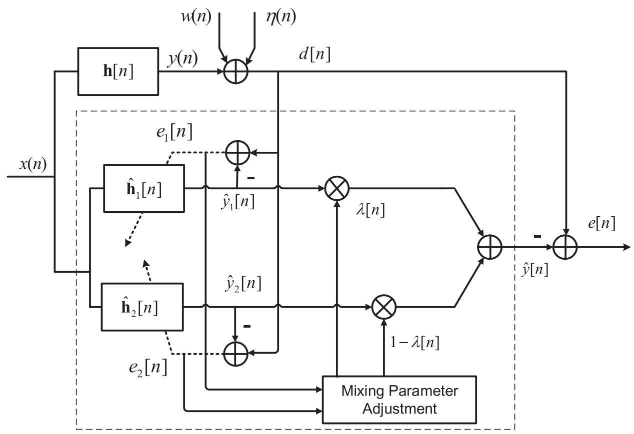

FIGURE 1. System block diagram with the combination of adaptive filters.

II. SYSTEM MODELS

A block diagram of the system used as an acoustic echo canceller is illustrated in Fig. 1. The far-end signal $x(n)$ passes along an unknown echo path $\mathbf{h}(n) = \left[h_0(n), h_1(n), \ldots, h_{N_h-1}(n)\right]^T$ through a room, which is modeled by a finite impulse response (FIR) filter of length N_h, to produce an acoustic echo signal $y(n)$. At the near-end, the echo signal and additive background noise are picked up by a microphone. Thus, the desired signal $d(n)$ can be expressed as follows:

$$d(n) = \mathbf{x}^T(n)\mathbf{h}(n) + w(n) + \eta(n) \tag{1}$$

where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N_h+1)]^T$ is the input regressor vector; superscript T denotes the transpose operation; $w(n)$ is additive white Gaussian noise; and $\eta(n)$ denotes the IN. The output of AEC $\widehat{y}(n)$ is obtained by combining the two adaptive filters, which are denoted as $\widehat{\mathbf{h}}_1(n)$ and $\widehat{\mathbf{h}}_2(n)$, using mixing factor $\lambda^*(n)$. The two filters are adaptively adjusted using the error signals $e_1(n)$ and $e_2(n)$, respectively. The coefficient used to update the recursion of the two adaptive filters can be expressed as follows:

$$\widehat{\mathbf{h}}_i(n+1) = \widehat{\mathbf{h}}_i(n) + \Delta\widehat{\mathbf{h}}_i\left(e_i, \widehat{\mathbf{x}}_i, \mu_i\right)(n), \text{ for } i = 1, 2 \quad (2)$$

where $\Delta\widehat{\mathbf{h}}_i$ denotes the correction term for the ith adaptive filter. Note that this correction term depends on the adaptation algorithm and is a function of the input regressor vector $\widehat{\mathbf{x}}_i$, error signal e_i, and step-size μ_i. Without a loss of generality, we assume that $\widehat{\mathbf{h}}_1(n)$ converges more rapidly than $\widehat{\mathbf{h}}_2(n)$; however, it has a mean-square deviation (MSD) higher than that associated with $\widehat{\mathbf{h}}_2(n)$ (i.e., $\mu_1 > \mu_2$), such that the two are referred to as the fast and the slow adaptive filters, respectively. After combining the outputs of the two adaptive filters, the equivalent outputs $\widehat{y}(n)$ can be expressed as follows:

$$\widehat{y}(n) = \lambda(n)\widehat{y}_1(n) + (1 - \lambda(n))\,\widehat{y}_2(n) \quad (3)$$

Thus, the error signal $e(n)$ can be expressed as follows:

$$\begin{aligned} e(n) &= d(n) - \widehat{y}(n) \\ &= \mathbf{x}^T(n)\mathbf{h}(n) + v(n) - \left[\mathbf{x}^T(n)\left(\lambda(n)\widehat{\mathbf{h}}_{12}(n) + \widehat{\mathbf{h}}_2(n)\right)\right] \\ &= \mathbf{x}^T(n)\left(\widehat{\mathbf{h}}_{o2}(n) - \lambda(n)\widehat{\mathbf{h}}_{12}(n)\right) + v(n) \end{aligned} \quad (4)$$

where $\widehat{\mathbf{h}}_{o2}(n) := \mathbf{h}(n) - \widehat{\mathbf{h}}_2(n)$ denotes the difference between the unknown system $\mathbf{h}(n)$ and the estimated impulse response using the slow adaptive filter; i.e., $\widehat{\mathbf{h}}_2(n)$; $\widehat{\mathbf{h}}_{12}(n) := \widehat{\mathbf{h}}_1(n) - \widehat{\mathbf{h}}_2(n)$. The aggregated noise is denoted by $v(n) = w(n) + \eta(n)$. The problem lies in determining the means by which to systematically identify the mixing parameter $\lambda(n)$, wherein the advantages of both the fast and slow adaptive filters can be obtained simultaneously.

III. RELATED WORKS

The performance and convergence of the combined adaptive filters are outlined in [19]. In that paper, it was pointed out that convex combinations may actually be preferable to the affine combinations. The combining of multiple adaptive filters can be achieved using an hierarchical combination of layers or a single-layer approach. Without a loss of generality, we focus on the convex combination of two adaptive filters in this study.

The primary difficulty in obtaining the optimal mixing parameter lies in the fact the *a priori* filter error and the additional measurement noise are unknown beforehand [30]. This necessitates the use of practical approximation methods to approach the optimal mixing parameter. Bershad *et al.* [31] presented the statistical behavior of an affine combination of two LMS adaptive filters when the inputs are white Gaussian signals. They also examined a convex combination method in which the mixing parameter is limited within the interval [0, 1]. As previously stated, the optimal mixing parameter cannot be calculated without the knowledge of system responses beforehand. The authors presented two iterative algorithms to facilitate the adjustment of mixing parameters. The first algorithm is based on stochastic gradient search to estimate the optimal mixing parameter. Unfortunately, this approach requires a reliable estimate of noise power to determine the step-size for updating the mixing parameter. The fact that accuracy of noise power estimation may limit its practicality means that this approximation method is susceptible to failure in hostile environments. The second algorithm is based on the ratio of the error power averaged from the two adaptive filters. It is referred to as the error power based scheme (EPBS), and can be expressed as follows:

$$\lambda_{\text{EPBS}} = 1 - \kappa \operatorname{erf}\left(\frac{\widehat{e}_1^2(n)}{\widehat{e}_2^2(n)}\right) \quad (5)$$

where $\widehat{e}_1^2(n)$ and $\widehat{e}_2^2(n)$ correspond to the estimated error power for the adaptive filter with larger and smaller step-sizes, respectively. The error power is obtained using a uniform sliding time average method. $\operatorname{erf}(\cdot)$ is the Gaussian error function, and the value of κ can be selected empirically. Arenas-García *et al.* [32] proposed the iterative updating of a mixing factor by minimizing the quadratic error of the filter combination. Factor $a(n)$ is then refined via a sigmoid activation function. The resulting mixing parameter can be expressed as

$$\lambda_C(n) = 1/[1 + e^{-a(n)}] \quad (6)$$

where the original mixing factor is iteratively updated as follows:

$$\begin{aligned} a(n+1) = a(n) - \mu_a e(n) \cdot [e_1(n) - e_2(n)] \\ \cdot \lambda_C(n) \cdot [1 - \lambda_C(n)] \end{aligned} \quad (7)$$

where μ_a is the step-size for updating $a(n)$; and $e_1(n)$ and $e_2(n)$ correspond to the error signals for the fast and slow adaptive filters, respectively. Unfortunately, this mixing parameter is sensitive to the effects of IN. The author also proposed a step-by-step transfer of the weight vector from the fast adaptive filter to the slow one in order to accelerate the convergence rate of the slow adaptive filter when an abrupt change appears.

IN or strong background noise can be dealt with by applying the VSS algorithm to the affine projection sign algorithm (APSA) [33]. The main idea is to iteratively pursue the optimal step-size by minimizing the upper bound of the MSD of the $\mathcal{L}_1$-norm of the measurement noise. Instead of combining the output of two adaptive filters, the authors proposed another novel VSS algorithm for APSA in environments with impulsive interference [34]. Their VSS algorithm involves mixing two step-sizes, called combined step-size (CSS), so that the rapid convergence of the large step-size filter can be achieved at the same time as the low steady-state misadjustment of the small step-size filter. Unfortunately, the CSS approach is unable to respond immediately to abrupt changes in impulse responses in an unknown system.

IV. PROPOSED CONVEX COMBINED ADAPTIVE FILTERING ALGORITHM

In this section, we first derive the optimal mixing parameter $\lambda(n)$ under the assumptions of white input signal and independence theory [35], which states that all regression vectors $\mathbf{x}(n)$ are statistically independent. Although this assumption of independence is unrealistic, most previous studies on adaptive filtering have adopted it [36]. We then propose a practical approximation method to obtain the mixing parameters. Finally, we provide a robust combination method to alleviate the impact of the abrupt changes in channel or strong background noise or both.

A. OPTIMAL COMBINATION FACTOR

By minimizing the conditional mean square error (MSE) associated with the system, we find the optimal combination factor $\lambda_o(n)$. Setting the derivative of the conditional MSE, i.e., $\mathbb{E}[e^2(n)|\widehat{\mathbf{h}}_1(n), \widehat{\mathbf{h}}_2(n)]$, with respect to $\lambda(n)$ to zero, we obtain

$$\mathbb{E}\left[-2e(n)\mathbf{x}^T(n)\widehat{\mathbf{h}}_{12}(n)|\widehat{\mathbf{h}}_1(n), \widehat{\mathbf{h}}_2(n)\right] = 0 \tag{8}$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation operation. After substituting (4) into (8) and under the assumption of independence theory [35] and $\mathbb{E}[v(n)] = 0$, we obtain the optimal combination factor as follows:

$$\lambda_o(n) = \frac{\widehat{\mathbf{h}}_{o2}^T(n)\mathbf{R}_\mathbf{x}(n)\widehat{\mathbf{h}}_{12}(n)}{\widehat{\mathbf{h}}_{12}^T(n)\mathbf{R}_\mathbf{x}(n)\widehat{\mathbf{h}}_{12}(n)} \tag{9}$$

where $\mathbf{R}_\mathbf{x}(n) := \mathbb{E}\left[\mathbf{x}(n)\mathbf{x}^T(n)\right]$ denotes the auto-correlation matrix of the input regressor vector. Note that $\widehat{\mathbf{h}}_{o2}^T(n)$ and $\mathbf{R}_\mathbf{x}(n)$ are not available in practical situations. Thus, we require a practical approximation method by which to calculate the combination factor.

B. APPROXIMATION METHOD

After disregarding the aggregated noise $v(n)$, we are left with $d(n) \approx y(n)$. Thus, Eq. (9) can be rewritten as follows:

$$\begin{aligned}\lambda_o(n) &\approx \frac{\mathbb{E}\left[\left(y(n) - \hat{y}_2(n)\right)\left(\hat{y}_1^T(n) - \hat{y}_2^T(n)\right)\right]}{\mathbb{E}\left[\left(\hat{y}_1(n) - \hat{y}_2(n)\right)^2\right]} \\ &= \frac{\mathbb{E}\left[e_2(n)(e_2(n) - e_1(n))\right]}{\mathbb{E}\left[(e_2(n) - e_1(n))(e_2(n) - e_1(n))\right]}.\end{aligned} \tag{10}$$

We then use $\lambda(n)$ to approximate the optimal combination factor as follows:

$$\lambda(n) = \frac{\sigma_{e_2}^2(n) - r_e(n)}{\sigma_{e_1}^2(n) - 2r_e(n) + \sigma_{e_2}^2(n)} \tag{11}$$

where $\sigma_{e_1}^2(n)$ and $\sigma_{e_2}^2(n)$ are the averaged instantaneous error power; $r_e(n) := \mathbb{E}[e_1(n)e_2(n)]$ denotes the cross-correlation between $e_1(n)$ and $e_2(n)$. We use K instantaneous values to calculate the averaged values as follows:

$$\sigma_{e_i}^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_i^2(m) \tag{12}$$

and

$$r_e(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} e_1(m)e_2(m). \tag{13}$$

Note that during the initial adaptation or after an abrupt change in the acoustic echo channel, transferring the fast filter coefficients $(\widehat{\mathbf{h}}_1(n))$ to the slow filter $(\widehat{\mathbf{h}}_2(n))$ (either partially [32] or fully [37]) makes it possible to reduce the transient time associated with the slow adaptive filter. Thus, we modify the adaptation rule for the slow filter as follows: If $\lambda'(n) > \beta$, then

$$\widehat{\mathbf{h}}_2(n) = \gamma\left[\widehat{\mathbf{h}}_2(n-1) + \Delta\widehat{\mathbf{h}}_2(n)\right] + (1-\gamma)\widehat{\mathbf{h}}_1(n) \tag{14}$$

where γ is a parameter close to 1 and the value of $(1-\gamma)$ depends on the weights that the fast filter transferring to the slow filter. Note that β is a threshold value and the correct term $\Delta\widehat{\mathbf{h}}_2(n)$ depends on the underlying adaptation algorithm. Otherwise, no coefficient is transferred from $\widehat{\mathbf{h}}_1(n)$, i.e., if $\lambda'(n) \le \beta$, then

$$\widehat{\mathbf{h}}_2(n) = \widehat{\mathbf{h}}_2(n-1) + \Delta\widehat{\mathbf{h}}_2(n). \tag{15}$$

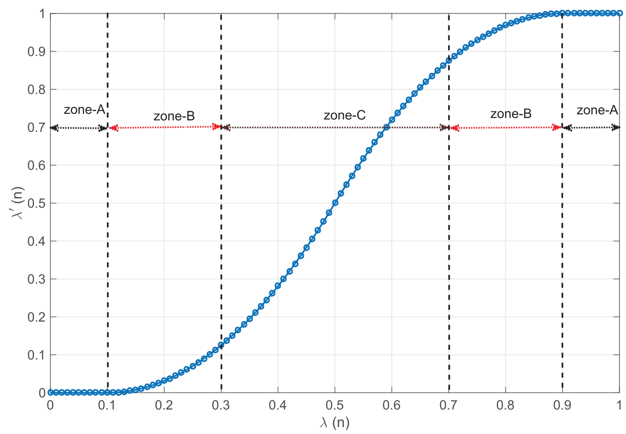

FIGURE 2. The S-shaped membership function.

C. ENHANCED ROBUST COMBINATION METHOD

To establish a robust combination method, we cannot ignore the impact of background noise or channel variations. Thus, we must modify Eq. (11). We apply a nonlinear transformation to $\lambda(n)$, such that its output $\lambda'(n)$ is confined with $[0, 1]$, i.e., convex combination. This transformation can be expressed by an S-shaped membership function, as follows:

$$\lambda'(n) = S(\lambda(n), \tau_1, \tau_2) = \begin{cases} 0, & \lambda(n) < \tau_1 \\ 2\left(\dfrac{\lambda(n)-\tau_1}{\tau_2-\tau_1}\right)^2, & \tau_1 \le \lambda(n) < \dfrac{\tau_1+\tau_2}{2} \\ 1-2\left(\dfrac{\lambda(n)-\tau_2}{\tau_2-\tau_1}\right)^2, & \dfrac{\tau_1+\tau_2}{2} \le \lambda(n) < \tau_2 \\ 1, & \lambda(n) \ge \tau_2 \end{cases} \tag{16}$$

where τ_1 and τ_2 are set to 0.1 and 0.9, respectively. The resulting mapping curve is shown in Fig. 2. We can divide the domain of the S-shaped membership function into three zones. In zone-B, the slope of the non-linear transformation is gradually decreased to zero (zone-A), such that fluctuations in the mixing parameter becomes less pronounced. In contrast, the slope of $S(\cdot)$ in zone-C is greater than one, which means that it is sufficient to reduce the transient time for the mixing parameter to settle.

In cases of strong IN, we have to set the combination factor $\lambda'(n)$ to 0 as follows:

$$\lambda'(n) = \begin{cases} 0, & \hat{\sigma}_d^2(n) \geq \rho\hat{\sigma}_x^2(n) \\ \lambda'(n), & \text{otherwise} \end{cases} \quad (17)$$

where

$$\hat{\sigma}_d^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} d^2(m) \quad (18)$$

and

$$\hat{\sigma}_x^2(n) = \frac{1}{K}\sum_{m=n-K+1}^{n} x^2(m) \quad (19)$$

represent the averaged instantaneous power for $d(n)$ and $x(n)$, respectively. Note that the value of ρ depends on the sensitivity of the combination factor respect to the IN. As the averaged power of IN is too strong, i.e., $\hat{\sigma}_d^2(n) \geq \rho\hat{\sigma}_x^2(n)$, we prefer forcing $\lambda'(n) = 0$. This means the fast adaptive filter, which is vulnerable to IN, is temporary disabled. Thereby, we can mitigate the impact of IN on the combined adaptive filter. Finally, we use the following recursion to smooth $\lambda'(n)$:

$$\lambda^*(n) = \alpha\lambda^*(n-1) + (1-\alpha)\lambda'(n) \quad (20)$$

where $0 \ll \alpha < 1$.

D. COMPUTATIONAL COMPLEXITY

The extra computational complexity incurs by the proposed robust convex combination method mainly comes from calculating the averaged instantaneous error power in (11) and (17). To obtain the approximation of the optimal combination factor in (11), we use K instantaneous values to calculate the averaged instantaneous error power $\sigma_{e_1}^2(n)$ and $\sigma_{e_2}^2(n)$; the averaged cross-correlation $r_e(n)$ is obtained by K instantaneous values of $e_1(n)$ and $e_2(n)$. These computation necessitates $3 \times K$ multipliers, $3 \times (K-1)$ adders, and 3 dividers. To obtain the the averaged instantaneous power for $d(n)$ and $x(n)$ in (17), we need $2 \times K$ multipliers, $2 \times (K-1)$ adders, and 2 dividers. Note that the S-shaped membership function defined in (16) can be implemented by using a pre-calculated table. Thus, the extra computation complexity associated with our combination method is affordable.

V. SIMULATION RESULTS

The simulation results were obtained from an ensemble average of 50 independent runs. Two performance metrics were used to evaluate the performance of the combination algorithms. One metric was normalized MSD (NMSD), which is defined as follows:

$$\text{NMSD} = 10\log_{10}\left(\frac{\|\mathbf{h} - \widehat{\mathbf{h}}(n)\|_2^2}{\|\mathbf{h}\|_2^2}\right) \quad (21)$$

where $\mathbf{h}$ is the unknown impulse response to be identified and $\widehat{\mathbf{h}}(n)$ is the estimated impulse response at time n. The other performance metric is called echo return loss enhancement (ERLE), which is defined as follows:

$$\text{ERLE} = 10\log_{10}\left(\frac{\mathbb{E}[d^2(n)]}{\mathbb{E}[e^2(n)]}\right) \quad (22)$$

where $d(n)$ is the desired signal defined in (1) and $e(n)$ is the error signal defined in (4).

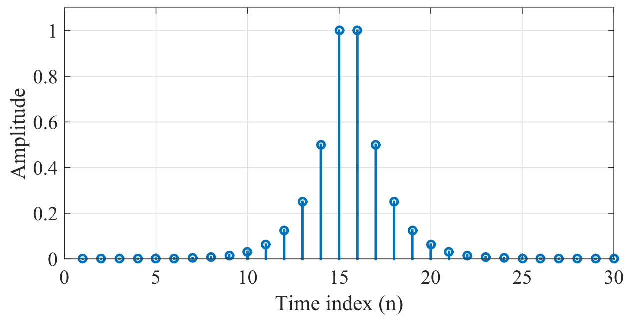

FIGURE 3. **The impulse response of an unknown system in the simple scenario, which has been exemplified in [31, Fig. 2].**

A. SIMPLE SCENARIO

In this simple system identification scenario, we considered the input signals are white Gaussian distributed and the impulse response of an unknown system with tap length N_h in shown in Fig. 3. We compare the proposed approximation method (20) with the optimal combination factor (9) and an existing approximation method referred to as the error-power-based scheme (EPBS) [31]. It should be noted that to enable a fair performance comparison with the other methods, we did not transfer coefficients from the fast filter to the slow filter; i.e., γ in (14) equals one in this sub-section. We also disabled the mechanism used to prevent the adaptive filter from diverging caused due to IN, as defined in (17).

The main simulation parameters were adopted from [31]. The adaptation algorithm in this example was the LMS algorithm with $\mu_1 = 1/\left((N_h+2)\sigma_x^2\right)$ and $\mu_2 = \theta\mu_1$, where σ_x^2 is the variance of the white Gaussian input signal and $N_h = 30$ is the tap-length of the unknown system to be identified. In this example, $\eta[n] = 0$ (i.e., there is no IN), the signal-to-noise ratio (SNR) is 30 dB, and the averaging window length $K = 100$. Then, we compare the resulting combination factor and ERLE for various θ values.

In Fig.4, $\mu_2 = 0.5\mu_1$ and the step-size of the fast and slow filters does not make a significant difference. In this case, the optimal combination factor is negative, i.e., an affine combination of the two LMS filters. The evolution of $\lambda(n)$ associated with the EPBS method responds slowly in its approach to the ideal value. Without using the non-linear transformation of $\lambda(n)$ defined in (16), the proposed approximation method also leads to the affine combination results. However, the proposed method confines the combination factor within the

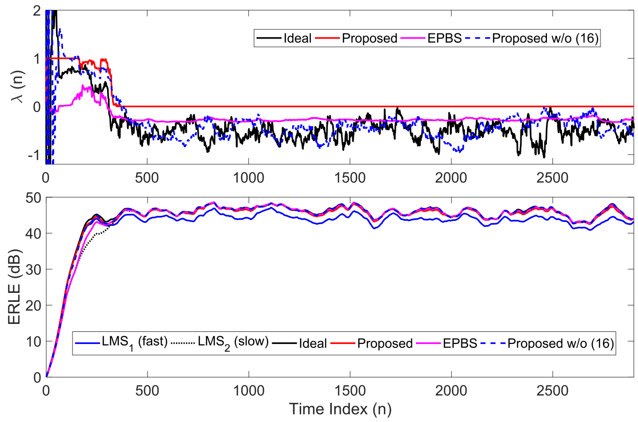

FIGURE 4. A comparison of the time evolution of the combination factor $\lambda[n]$ (top) and ERLE (bottom) for $\theta = 0.5$.

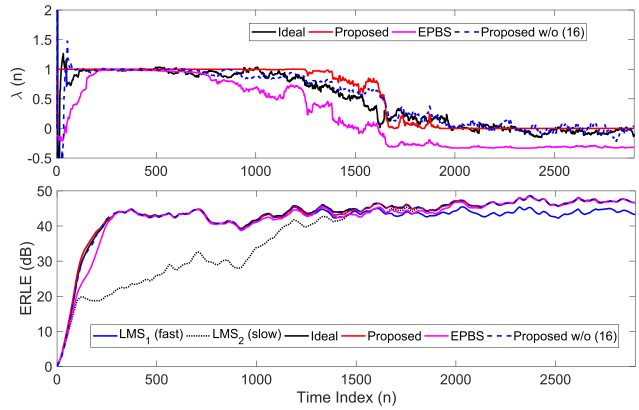

FIGURE 5. A comparison of the time evolution of the combination factor $\lambda[n]$ (top) and ERLE (bottom) for $\theta = 0.08$.

range of [0, 1]. Therefore, the best that can be expected when using the proposed method is set the combination factor to zero. Initially, the combination factor is close to 1, which implies that the main contribution to the output is from the fast adaptive filter. However, as the resulting MSE reduce to a extent, the combination factor is close to 0, which implies that the output is determined primarily by the slow adaptive filter (and it lower MSE). In contrast, the EPBS method, which is an affine combination, has a better approximation than our method. The time evolution of $\lambda[n]$ slowly reached the first steady-state (around $\lambda[n] = 0.4$) at approximately $n = 180$, while our proposed method approached the first steady-state (around $\lambda[n] = 1.0$) in few iterations. This indicated that the EPBS method required more time to reach its ideal value. The EPBS method required approximately 500 samples to transit the value of $\lambda[n]$ from 0.4 to the second steady-state (around $\lambda[n] = -0.3$), whereas the proposed method required approximately 200 samples to transit the from value of $\lambda[n]$ from the first steady-state to the second steady-state (around $\lambda[n] = 0$), which means the response of the proposed method is faster than that of the EPBS method. Note that, as shown in the corresponding ERLE curves, no significant loss of our proposed method was found. The ERLE curve obtained using the proposed method is superior to that of the EPBS method before $n = 300$. Beyond this point, the ERLE curves almost overlap.

As shown in Fig.5, as θ becomes smaller, the optimal $\lambda[n]$ approaches zero in the steady-state. In this case, the approximation obtained using our proposed method is closer to the optimal combination factor than is that obtained using the EPBS method. For the proposed method, $\lambda[n]$ increases initially; however, from $n = 1250$ to $n = 1950$, it gradually drops from one to zero. Beyond this transient period, it remains stationary at approximately zero. The proposed method without Eq. (16) is close to the ideal case, except for the transient period around $n = 1500$ to $n = 1700$, and $n < 100$. However, the ERLE curves remain nearly unchanged regardless of whether or not Eq. (16) is used. When using the EPBS method, the evolution of $\lambda(n)$ exhibits a larger bias during the steady-state (after $n = 2000$). Furthermore, the resulting $\lambda(n)$ may be negative in the case of $\theta = 0.08$. Moreover, the mixing parameter given by the EPBS method increases slowly to one at the beginning and then slowly drops to its steady-state when the MSE of the slow adaptive filter is lower. Even worse, the steady-state value of $\lambda(n)$ when using the EPBS method is approximately -0.3. This bias indicates that the EPBS method does not provide accurate approximations of optimal mixing parameter. The ERLE curves give similar results. Due to bias in $\lambda[n]$, the ERLE curves obtained using EPBS method had loss of approximately 3 dB.

Note that, to acquire the gain of the combination filter, the step-size of the fast and slow filters would be very different; i.e., θ should be quite low. When this is the case, the proposed method comes close to the optimal mixing parameter and surpasses the performance of EPBS approximation. It should also be noted that in this sub-section, the performance of the proposed method is better without the Eq. (16) than it is with (16). However, in the next sub-section, we will show that the non-linear transformation function (16) is advantageous in a holistic environment scenario.

B. COMPLICATED SCENARIO

In simulating hostile environments, we consider the following complicated scenarios. We first assume that the unknown system to be identified has long taps and an impulse response $\mathbf{h}[n]$ that changes abruptly to $-\mathbf{h}[n]$ at time $n = 3.5 \times 10^4$. Fig. 6 presents the original and changed impulse responses. Second, we assume that the input signal is no longer a white Gaussian input. On the contrary, it could be a color Gaussian input or a speech input [38]. A color Gaussian input is generated by filtering a white Gaussian input through a first-order autoregressive system with a pole at 0.7. Third, we evaluate the resulting performance when additive IN $\eta(n)$ is generated using a Bernoulli-Gaussian (BG) IN model or a recording of creaking door. The BG IN is given by

$$\eta(n) = b(n) \cdot g(n) \tag{23}$$

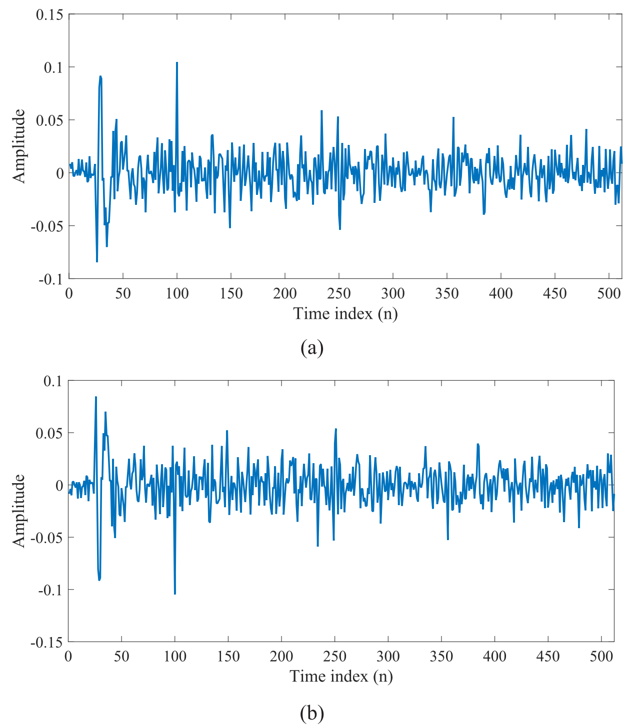

FIGURE 6. **The impulse response of an unknown system in the complicated scenario. (a) Original and (b) changed impulse responses.**

where $b(n)$ denotes the Bernoulli process with probability of success p, and $g(n)$ is the white Gaussian process with zero mean and variance of σ_w^2. Two parameters are used to represent the intensity of the IN; i.e., the occurrence probability of IN p and the Gaussian-to-impulsive-noise ratio (GINR) $\Gamma := \sigma_w^2/\sigma_\eta^2$, where σ_η^2 denotes the variance of the BG IN. A higher p value and smaller Γ value indicate a situation with greater additive IN. In the following simulation, we selected $p = 0.05$ and $\Gamma = 10^{-3}$.

The adaptation algorithm used for the fast and slow adaptive filters is ASPA. The weight updating equation with APSA algorithm of the ith adaptive filter can be expressed using the method outlined in [39], as follows:

$$\hat{\mathbf{h}}_i(n) = \hat{\mathbf{h}}_i(n-1) + \mu_i \frac{\mathbf{X}(n)\mathrm{sgn}(\mathbf{e}_i(n))}{\|\mathbf{X}(n)\mathrm{sgn}(\mathbf{e}_i(n))\|_2 + \delta} \tag{24}$$

where $0 < \delta \ll 1$ is a regularization parameter to prevent numerical divergence during the adaptation process, $\mathrm{sgn}\{\cdot\}$ is the sign function, and $\mathbf{X}(n)$ is an $N_h \times P$ matrix as follows:

$$\mathbf{X}(n) = [\mathbf{x}(n), \mathbf{x}(n-1), \ldots, \mathbf{x}(n-P+1)]^T \tag{25}$$

where $\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-N_h+1)]^T$ denotes the input regressor vector and P is the projection order. The error vector $\mathbf{e}_i(n) = [e_i(n), e_i(n-1), \ldots, e_i(n-P+1)]^T$ corresponds to the i-th adaptive filter can be expressed as follows:

$$\mathbf{e}_i(n) = \mathbf{d}(n) - \mathbf{X}^T(n)\hat{\mathbf{h}}_i(n-1) \tag{26}$$

where $\mathbf{d}(n) = [d(n), d(n-1), \ldots, d(n-P+1)]^T$ is the desired signal vector. In the following simulation, the value of SNR is 30 dB, the projection order $P = 4$, and the regularization parameter $\delta = 10^{-6}$. For cases involving color Gaussian or speech inputs, we selected $\mu_1 = 10^{-2}$ and $\mu_2 = 10^{-3}$. For cases involving conversation inputs or actual measured echoes, we selected $\mu_1 = 10^{-3}$ and $\mu_2 = 10^{-4}$. The size of uniform sliding window K used to calculate the averaged power in Eqs. (12), (13), (18), and (19) may depend on the input signal. We respectively selected K values of 200 and 400 for the color Gaussian signal and other signals. We chose 0.999 and 0.9 as the values of γ and β, respectively, in (14). The smoothing factor α used in Eq. (20) was 0.9 for color Gaussian inputs and 0.999 for all other cases. The proposed method was compared with C-APSA [32], VSS-APSA [33], and CSS-APSA [34].

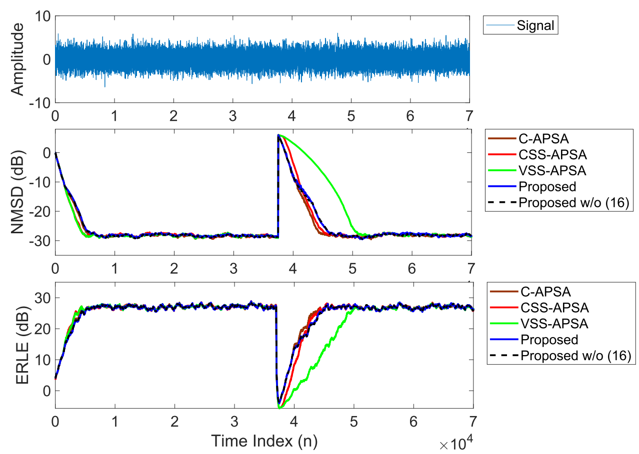

FIGURE 7. **The comparison of the NMSD and ERLE curves for different mixing parameters with color Gaussian input.**

1) COLOR GAUSSIAN INPUT

Fig. 7 illustrates the time evolution of the NMSD and ERLE curves when using a color Gaussian input. In cases involving abrupt channel changes, the time required to reach a steady-state (the same NMSD level) is as follows: VSS-APSA (17,000 samples) and CSS-APSA, C-APSA, and the proposed method (8500 samples). We obtained similar results for the ERLE curves, as shown in the bottom panel of Fig. 7. It should be noted that the proposed method without Eq. (16) produced fluctuations in the ERLE and NMSD curves when in a steady-state (around $n \in [0.8 - 3.5] \times 10^4$ and $n \in [4.8 - 7] \times 10^4$. Even worse, when echo path abruptly changed at $n = 3.5 \times 10^4$, it too longer to reach a steady-state without the aid of (16). The root cause of these results can be found in Figs. 2 and 8. With the aid of (16), mixing parameter $\lambda(n)$ is clipped when it exceeds τ_2 or smaller than τ_1, and the slope of this transformation less than 1 for $\tau_1 \leq \lambda(n) \leq 0.3$ and $0.7 \leq \lambda(n) \leq \tau_2$. Thus, (16) can be used to stifle fluctuations in the mixing parameter when the behavior of the combined adaptive filter tends toward one of the two component filters. We also observed that the slope of this transformation is slight greater than 1 in the case where $0.3 < \lambda(n) < 0.7$, which accelerates the convergence

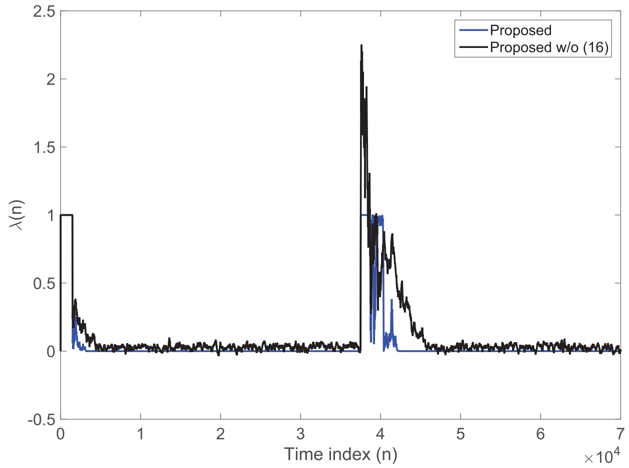

FIGURE 8. The comparison of the evolution of the mixing parameters with and without Eq. (16).

of the mixing parameter. Fig. 8 illustrated that the proposed method with Eq. (16) is able to fast adapt the value of $\lambda(n)$ to its steady-state and to reduce the fluctuations of $\lambda(n)$ in the steady-state than that without the aid of Eq. (16). Thus, we implemented the proposed method in conjunction with Eq. (16) in the following simulations.

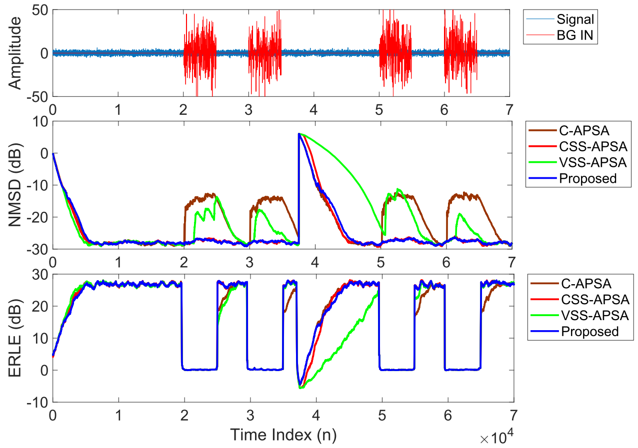

FIGURE 9. The comparison of the NMSD and ERLE curves for different mixing parameters with color Gaussian input and BG INs (GINR $\Gamma = 10^{-3}$ and $p = 0.05$).

2) COLOR GAUSSIAN INPUT WITH ADDITIVE IN

In addition to white Gaussian background noise, we also considered situations involving IN with the BG model. The upper panel of Fig. 9 presents a trace file of BG IN. Notice that the IN appears in four regions: $n \in [2-2.5] \times 10^4$, $n \in [3-3.5] \times 10^4$, $n \in [5-5.5] \times 10^4$, and $n \in [6-6.5] \times 10^4$. Moreover, there was an abrupt channel change at $n = 3.5 \times 10^4$. In these four regions, the BG IN seriously degraded the performance of the C-APSA and VSS-APSA algorithms. CSS-APSA and the proposed method have comparable performance to against the IN; however, in the event of an abrupt channel change, the convergence rate of the CSS-APSA (in returning to a steady-state) was slower than that of our proposed method.

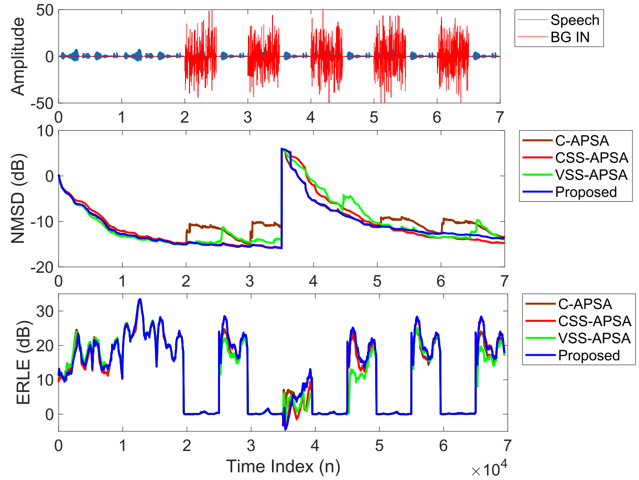

FIGURE 10. The comparison of the NMSD and ERLE curves for different mixing parameters with speech input and BG INs (GINR $\Gamma = 10^{-3}$ and $p = 0.05$).

3) SPEECH INPUT WITH BG IN

In this case, we consider a speech input signal corrupted by BG IN. The trace file is shown in the upper panel of Fig. 10. In the NMSD curves, we can see that the C-APSA is robust to channel variations but vulnerable to BG IN. In contrast, VSS-APSA is robust to the BG IN; however, when a speech input follows BG IN, it suddenly diverges as indicated by a jump in the NMSD curve. The CSS-APSA algorithm suffers from the channel variations only. Overall, the proposed method outperformed all three of these algorithms.

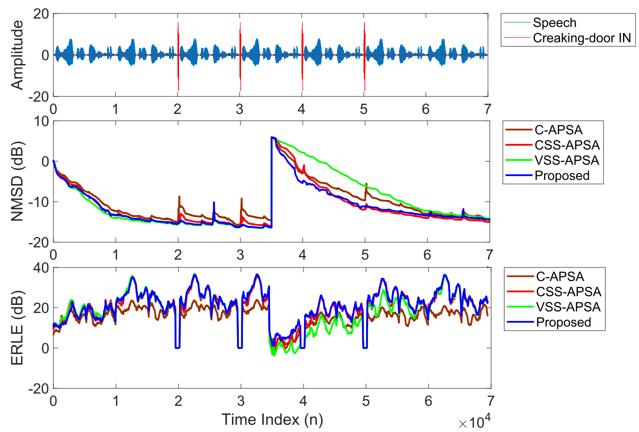

FIGURE 11. The comparison of the NMSD and ERLE curves for different mixing parameters with a single-talk speech input and a creaking-door IN.

4) SPEECH INPUT WITH CREAKING-DOOR IN

We introduced a short-duration sound of a creaking door to confirm the effectiveness of our proposed method in a realistic environment. The trace file of this IN is shown in the upper panel of Fig. 11. We observed consistent results in the NMSD and ERLE curves, respectively appearing in the middle and bottom panels in Fig. 11. Note that VSS-APSA was insensitive to IN of the short duration and low intensity, such as the creaking door; however, observed slow convergence after an abrupt change in the echo channel. C-APSA achieved satisfactory performance except in the presence of IN. CSS-APSA presented a slight drop in performance degradation in the presence of IN; however, it still suffered from the variations in the echo channel. Once again, our proposed method achieved the best performance among the four parameter selection schemes.

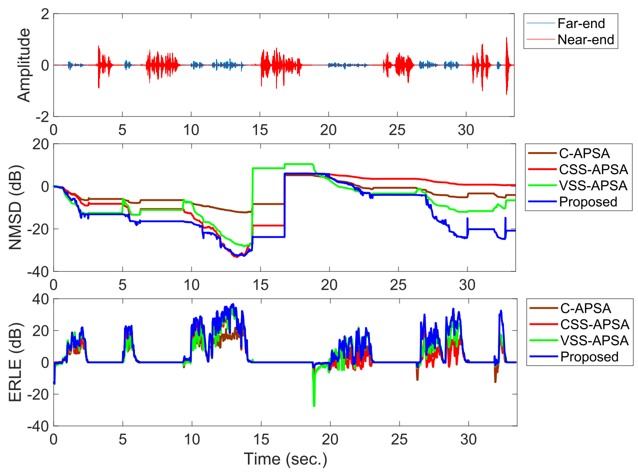

FIGURE 12. The comparison of the NMSD and ERLE curves for different mixing parameters with a conversation speech inputs.

5) CONVERSATION INPUT

In this scenario, we mimic a conversation between two men. The trace file of this conversation is shown in the top panel of Fig. 12. The duration of the conversation is 33.5 s and the echo path changed at 16.75 s. Notice that no double-talk occurred in this case. In this scenario, the VSS-APSA algorithm was sensitive to the near-end speech as well as changes in the echo path. The CSS-APSA algorithm is robust to near-end speech; however, it is unable to deal with changes in the echo path. Variations in the echo channel caused the rate of convergence to slow down significantly. Overall, our proposed method can achieve the lowest NMSD in this experiment.

6) REAL MEASURED ECHO

In this scenario, we used a hands-free smart phone to speak to a person in the far-end in a compact car. Note that the build-in echo cancellation function of the smart phone are disabled in this experiment. All trace files were recorded at a sampling rate of 8 kHz. The far-end signal is shown in the top panel in Fig. 13. The measured microphone inputs, which include the echo signal results from the hand-free communication and the near-end speech, are illustrated in the middle panel in Fig. 13. Notice that the signals between approximately $n = 1.3 \times 10^5$ and $n = 2.1 \times 10^5$ comprise only pure near-end speech. Beyond $n = 2.1 \times 10^5$, there occurs double-talk.

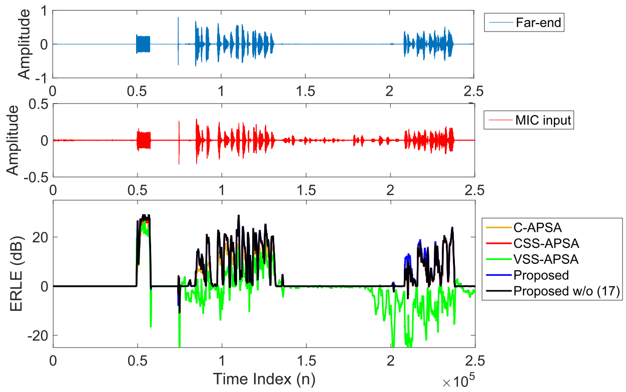

FIGURE 13. The comparison of the ERLE curves for different mixing parameters with a real measured echo signals.

The resulting ERLE curves are shown in the bottom panel in Fig. 13. With no far-end voice activity, the ideal ERLE value should be zero. VSS-APSA was sensitive to signal variations and achieved the worst performance in this comparison. VSS-APSA was unable to converge in dealing with double-talk. C-APSA presented low tracking capability when $n \in (0.9 - 1.4) \times 10^5$. CSS-APSA presented performance comparable to that of our proposed method when there was no double-talk; however, our method achieved better ERLE results in the presence of double-talk. The proposed method, with or without (17), achieved the same ERLE curve, except during periods of double-talk. During the period of double-talk, the proposed method with (17) achieved a higher ERLE.

VI. CONCLUSIONS

This paper presents a novel convex combination adaptive filtering algorithm for AEC applications in difficult environments. We propose a method by which to approximate the optimal convex combination factor, which is unavailable due to the presence of unknown system impulse responses. This combination factor was made more robust through the following ideas. First, we apply a non-linear transformation to the original combination factor such that this factor has less fluctuations in the steady-state and has less transient time during the evolution of the combination factor. Second, we provide a simple method to detect the presence of strong ambient IN. In this case, we will force the combination factor to zero, which increases the weighting of the slow filter. Therefore, it is possible to mitigate the impact of the IN on the combined adaptive filter. Simulation results have confirmed that our proposed approximation method resembles the optimal combination factor than the EPBS method, particularly when the features of the two adaptive filters differ significantly different; i.e., the step-size used in the fast LMS algorithm is roughly 10-times larger than that of the slow one. The performance of the proposed scheme was verified using color Gaussian input signals with BG IN as well as in the AEC applications using real speech inputs and recordings of a creaking door. Experiment results demonstrate the robustness of the proposed scheme to IN as well as abrupt variations in the echo channel. Our method outperforms other existing methods in terms of the resulting NMSD and ERLE curves.

REFERENCES

[1] S. M. Kuo, H. Chuang, and P. P. Mallela, "Integrated automotive signal processing and audio system," *IEEE Trans. Consum. Electron.*, vol. 39, no. 3, pp. 522–532, Jun. 1993.

[2] W. L. B. Jeannes, P. Scalart, G. Faucon, and C. Beaugeant, "Combined noise and echo reduction in hands-free systems: A survey," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 808–820, Nov. 2001.

[3] M. Fukui, S. Shimauchi, K. Kobayashi, Y. Hioka, and H. Ohmuro, "Acoustic echo canceller software for voip hands-free application on smartphone and tablet devices," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 461–467, Aug. 2014.

[4] M. Fukui, S. Shimauchi, Y. Hioka, A. Nakagawa, and Y. Haneda, "Double-talk robust acoustic echo cancellation for CD-quality hands-free video-conferencing system," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 468–475, Aug. 2014.

[5] B. N. M. Laska, R. A. Goubran, and M. Bolic, "Improved proportionate subband NLMS for acoustic echo cancellation in changing environments," *IEEE Signal Process. Lett.*, vol. 15, pp. 337–340, Mar. 2008.

[6] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[7] A. W. H. Khong and P. A. Naylor, "Selective-tap adaptive filtering with performance analysis for identification of time-varying systems," *IEEE Trans. Audio, Speech, Language Process.*, vol. 15, no. 5, pp. 1681–1695, Jul. 2007.

[8] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A family of robust algorithms exploiting sparsity in adaptive filters," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 4, pp. 572–581, May 2009.

[9] J. C. Jenq and S. F. Hsieh, "Decision of double-talk and time-variant echo path for acoustic echo cancellation," *IEEE Signal Process. Lett.*, vol. 10, no. 11, pp. 317–319, Nov. 2003.

[10] C. Schuldt, F. Lindstrom, and I. Claesson, "A delay-based double-talk detector," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 6, pp. 1725–1733, Aug. 2012.

[11] Y. Tsao, S.-H. Fang, and Y. Shiao, "Acoustic echo cancellation using a vector-space-based adaptive filtering algorithm," *IEEE Signal Process. Lett.*, vol. 22, no. 3, pp. 351–355, Mar. 2015.

[12] J.-Y. Tourneret, N. J. Bershad, and J. C. M. Bermudez, "Echo cancellation—The generalized likelihood ratio test for double-talk versus channel change," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 916–926, Mar. 2009.

[13] J. Benesty, H. Rey, L. R. Vega, and S. Tressens, "A nonparametric VSS NLMS algorithm," *IEEE Signal Process. Lett.*, vol. 13, no. 10, pp. 581–584, Oct. 2006.

[14] C. Paleologu, J. Benesty, and S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 8, pp. 1466–1478, Nov. 2008.

[15] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 508–518, Sep. 2000.

[16] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2. May 2002, pp. II-1881–II-1884.

[17] O. Hoshuyama, R. A. Goubran, and A. Sugiyama, "A generalized proportionate variable step-size algorithm for fast changing acoustic environments," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 4. May 2004, pp. IV-161–IV-164.

[18] K.-A. Lee, W.-S. Gan, and S. M. Kuo, *Subband Adaptive Filtering: Theory and Implementation*. Hoboken, NJ, USA: Wiley, 2010.

[19] J. Arenas-García, L. A. Azpicueta-Ruiz, M. T. M. Silva, V. H. Nascimento, and A. H. Sayed, "Combinations of adaptive filters: Performance and convergence properties," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 120–140, Jan. 2016.

[20] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-García, and W. Kellermann, "Adaptive combination of Volterra kernels and its application to nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech Language Process.*, vol. 19, no. 1, pp. 97–110, Jan. 2011.

[21] L. A. Azpicueta-Ruiz, A. R. Figueiras-Vidal, and J. Arenas-Garcìa, "A normalized adaptation scheme for the convex combination of two adaptive filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2008, pp. 3301–3304.

[22] L. Shi, Y. Lin, and X. Xie, "Combination of affine projection sign algorithms for robust adaptive filtering in non-Gaussian impulsive interference," *Electron. Lett.*, vol. 50, no. 6, pp. 466–467, Mar. 2014.

[23] L. A. Azpicueta-Ruiz, J. Arenas-García, V. H. Nascimento, and M. T. M. Silva, "Reduced-cost combination of adaptive filters for acoustic echo cancellation," in *Proc. Int. Telecommun. Symp. (ITS)*, Aug. 2014, pp. 1–5.

[24] B. K. Das and M. Chakraborty, "Sparse adaptive filtering by an adaptive convex combination of the LMS and the ZA-LMS algorithms," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 5, pp. 1499–1507, May 2014.

[25] M. Ferrer, M. de Diego, A. Gonzalez, and G. Piñero, "Convex combination of affine projection algorithms," in *Proc. Int. Conf. Eur. Signal Process.*, Aug. 2009, pp. 431–435.

[26] L. Arevalo, J. A. Apolinario, M. L. R. de Campos, and R. Sampaio-Neto, "Convex combination of three affine projections adaptive filters," in *Proc. Int. Symp. Wireless Commun. Syst.*, Aug. 2013, pp. 1–5.

[27] J. Ni and F. Li, "Adaptive combination of subband adaptive filters for acoustic echo cancellation," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1549–1555, Aug. 2010.

[28] Z. Shen, Y. Yu, and T. Huang, "Normalized subband adaptive filter algorithm with combined step size for acoustic echo cancellation," *Circuits, Syst. Signal Process.*, vol. 36, no. 7, pp. 2991–3003, Jul. 2017. [Online]. Available: http://dx.doi.org/10.1007/s00034-016-0429-x

[29] Y. Dong, H. Zhao, and Y. Yu, "Adaptive combination of proportionate NSAF with individual activation factors," *Circuits, Syst. Signal Process.*, vol. 36, no. 4, pp. 1769–1780, 2017. [Online]. Available: http://dx.doi.org/10.1007/s00034-016-0386-4

[30] J. Arenas-García, A. R. Figueiras-Vidal, and A. H. Sayed, "Mean-square performance of a convex combination of two adaptive filters," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1078–1090, Mar. 2006.

[31] N. J. Bershad, J. C. M. Bermudez, and J.-Y. Tourneret, "An affine combination of two LMS adaptive filters—Transient mean-square analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1853–1864, May 2008.

[32] J. Arenas-García, M. Martínez-Ramón, A. Navia-Vázquez, and A. R. Figueiras-Vidal, "Plant identification via adaptive combination of transversal filters," *Signal Process.*, vol. 86, no. 9, pp. 2430–2438, 2006.

[33] J. Yoo, J. Shin, and P. Park, "Variable step-size affine projection sign algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 61, no. 4, pp. 274–278, Apr. 2014.

[34] F. Huang, J. Zhang, and S. Zhang, "Combined-step-size affine projection sign algorithm for robust adaptive filtering in impulsive interference environments," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 5, pp. 493–497, May 2016.

[35] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, no. 5, pp. 963–993, May 1979.

[36] T. Y. Al-Naffouri and A. H. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.

[37] V. H. Nascimento and R. C. de Lamare, "A low-complexity strategy for speeding up the convergence of convex combinations of adaptive filters," in *Proc. ICASSP*, Mar. 2012, pp. 3553–3556.

[38] D. Pearce, H. Hirsch, and E. E. D. Gmbh, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. ISCA ITRW ASR*, 2000, pp. 29–32.

[39] T. Shao, Y. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 327–330, Apr. 2010.

YING-REN CHIEN received the B.S. degree in electronic engineering from the National Yunlin University of Science and Technology, Douliu, Taiwan, in 1999, and the M.S. degree in electrical engineering and the Ph.D. degree in communication engineering from the National Taiwan University, Taipei, Taiwan, in 2001 and 2009, respectively. Since 2012, he has been with the Department of Electrical Engineering, National Ilan University, Yilan, Taiwan, where he is currently an Associate Professor. His research interests include adaptive signal processing theory, digital signal processing in very large scale integration, signal processing on communication system, and interference cancellation.

JIN LI-YOU received the B.S. degree in electronics engineering from National Ilan University, Yilan, Taiwan, in 2013, and the M.S. degree in electrical engineering from the Graduate Institute of Electrical Engineering, National Ilan University, in 2015. He is currently pursuing the Ph.D. degree in communication engineering with the National Taiwan University, Taipei, Taiwan. His research interests include adaptive signal processing and echo cancellation problems.