Generalized Combined Nonlinear Adaptive Filters for Nonlinear Acoustic Echo Cancellation

Wenxia Lu, Lijun Zhang, Jie Chen, Jingdong Chen
Center of Intelligent Acoustics and Immersive Communications
School of Marine Science and Technology, Northwestern Polytechnical University, China
luwenxia121@mail.nwpu.edu.cn zhanglj7385@nwpu.edu.cn dr.jie.chen@ieee.org jingdongchen@ieee.org

Abstract—**This paper proposes a new and generalized algorithm of combination of nonlinear adaptive filters (CNAF) for nonlinear acoustic echo cancellation (NAEC). In contrast to combining filters in a conventional parallel manner, the candidate filters are organized to form in a network structure with two subnetworks. The nodes in each subnetwork serve as linear and nonlinear adaptive filters respectively. A generalized CNAF (GCNAF) is then obtained by linking the nodes in the network and using the diffusion adaptation strategy. The proposed GCNAF algorithm allows information exchange and sharing among the nodes, so as to maximally optimize the performance of the combined filters. Simulations with noise and speech signals demonstrate the effectiveness of the proposed GCNAF for the NAEC problem.**

Index Terms—**Nonlinear acoustic echo cancellation, generalized combination of nonlinear adaptive filters, distributed optimization**

I. Introduction

Along with the widely used of the hands-free mobile phones and video conference applications in our daily lives, acoustic echo cancellation (AEC) has been widely studied in recent years. Echoes are generated acoustically by the coupling between the loudspeaker and the microphone via the impulse response of a room [1]. Removal of these echoes requires the precise knowledge of the impulse response of the acoustic echo path, which can be time varying. There has been a great interest in the use of adaptive filters as acoustic echo cancellers to remove echoes. Moreover, the nonlinear acoustic echo cancellation (NAEC) has received attention due to the inherent nonlinearities in acoustic devices. A variety of structures have been investigated in order to model nonlinear systems in practical applications, including trigonometric expansion [2]–[4], neural networks [5], [6], block-based Wiener-Hammerstein models [7]–[9], polynomial models [10]–[12], etc. A variety of NAEC algorithms are proposed based on these structures.

To enhance the performance and alleviate parameter selection dilemmas of adaptive filters, using the combination of nonlinear adaptive filters (CNAF) is gaining interest in NAEC. For example, collaborative functional link adaptive filters (FLAF) were proposed in [2] based on the adaptive combination of filters in order to improve their robustness against different degrees of nonlinearity. A combination of Volterra filters (CVF) and a combination of kernel (CK) filters were presented in [11]. While both approaches achieve similar performance that is superior to a single VF, the latter is significantly more efficient. An improved solution to the CK scheme was subsequently developed in [12], named D-NLAEC-AZK. However, the performance improvement of the existing CNAF is somehow limited as the design method neglects the internal structure optimization and information interaction between linear filters and nonlinear filters.

Recently, distributed adaptation has emerged as an attractive and challenging research area with the advent of multi-agent networks. There are several useful distributed strategies for data processing over networks including incremental strategies [19]–[22], and diffusion strategies [23]–[27]. The convergence rate of distributed optimization via diffusion strategies is enhanced and adaptive diffusion performs better than a non-cooperative strategy under certain conditions [24], [28]. Nodes in an adaptive network may approach the centralized solution through a continuous process of cooperation and information sharing with neighbors.

Inspired by the work in [11], [24], [29], we propose a new and generalized NAEC scheme of CNAF from the perspective of the distributed optimization based on the diffusion strategy over networks. The proposed GCNAF performs the filtering task via a linear filtering subnetwork and a nonlinear filtering one. In the linear filtering subnetwork, a combination of linear filters is obtained by optimally linking each node that is associated with a linear filter. Likewise, in the nonlinear filtering subnetwork, a combination of nonlinear filters is constructed by the topological linking of each node that is associated with a nonlinear filter. A new scheme is then established by combining the two subnetworks under the guidance of a unified optimization objective function. Simulations with an NAEC problem are conducted to validate the effectiveness of the developed GCNAF.

II. Diffusion LMS algorithm

Notation. Italic letters (e.g., x and X) denote scalars. Boldface small letters (e.g., $\mathbf{x}$) denote column vectors. Boldface capital letters (e.g., $\boldsymbol{X}$) denote matrices. The superscript $(\cdot)^\top$ represents the transpose of a matrix or a vector, and $\|\cdot\|$ denotes Euclidean norm of its vector argument. We denote by $\mathcal{N}_k$ the set of node indices in the neighborhood of node k, including k itself.

This work is supported in part by National Natural Science Foundation of China (NSFC) under Grants 61573288, 61425005, 61671382 and 61811530283.

We consider a distributed network consisting of N nodes. The problem is to estimate the same parameter vector, $\mathbf{h}_{\mathrm{T}}$, of size $L \times 1$. Diffusion adaptive strategies for the distributed estimation of $\mathbf{h}_{\mathrm{T}}$ was derived in [24] by seeking the minimizer of the following global cost function:

$$J^{\mathrm{glob}}(\mathbf{h}) = \sum_{k=1}^{N} J_k(\mathbf{h}), \tag{1}$$

where $J_k(\mathbf{h}) = \mathbb{E}\left\{\left|d_k(i) - \mathbf{x}_{k,i}^\top \mathbf{h}\right|^2\right\}$ denotes the mean-square-error cost at node k, with $\mathbf{x}_{k,i}$ denotes an $L \times 1$ input vector at time instant i, and $d_k(i)$ being the reference signal. Let $\mathbf{h}_{k,i}$ denote the estimate of the minimizer of (1) at node k and time instant i. An adapt-then-combine (ATC) diffusion least mean square (LMS) algorithm that solves (1) consists of the following steps [24]:

$$\boldsymbol{\phi}_{k,i} = \mathbf{h}_{k,i-1} + \mu_k \sum_{\ell \in \mathcal{N}_k} c_{\ell k} \mathbf{x}_{\ell,i} e_\ell(i), \tag{2}$$

$$\mathbf{h}_{k,i} = \sum_{\ell \in \mathcal{N}_k} a_{\ell k}\, \boldsymbol{\phi}_{\ell,i}. \tag{3}$$

where $e_\ell(i) = d_\ell(i) - \mathbf{x}_{\ell,i}^\top \mathbf{h}_{\ell,i-1}$. The cofficient $c_{\ell k}, a_{\ell k}$ are chosen to satisfy:

$$\begin{aligned} c_{\ell k} \geq 0, \ \sum_{\ell=1}^{N} c_{\ell k} = 1, \ \text{and} \ c_{\ell k} = 0, \ \text{if} \ \ell \notin \mathcal{N}_k, \\ a_{\ell k} \geq 0, \ \sum_{\ell=1}^{N} a_{\ell k} = 1, \ \text{and} \ a_{\ell k} = 0, \ \text{if} \ \ell \notin \mathcal{N}_k. \end{aligned} \tag{4}$$

Namely non-negative coefficients $c_{\ell k}$ form to a right-stochastic matrices $\mathbf{C}$ and $a_{\ell k}$ form to a left-stochastic matrix $\mathbf{A}$.

III. The proposed GCNAF

In NAEC problems, combination of nonlinear adaptive filters (CNAF) is gaining interest as it extends the linear model with an extra nonlinear term. In this section, we propose a new and generalized scheme of CNAF from the perspective of the distributed optimization based on the diffusion strategy over networks for NAEC.

A. *Network structure of the GCNAF*

Fig. 1 represents the proposed NAEC scheme, where we extend the parallel structure of filters in conventional filter combination schemes to a networked structure. The approach consists of two concurrent adaptive layers: a diffusion network layer and a combination layer. The former provides building blocks for modeling the system, and is composed of a linear filtering subnetwork and a nonlinear filtering subnetwork. The latter aims to produce a replica of the echo signal $y(i)$ by combining the output of the two subnetworks. For simplicity of exposition, we assume that both the linear and nonlinear subnetworks have the same number of nodes, but the connections of nodes can be different.

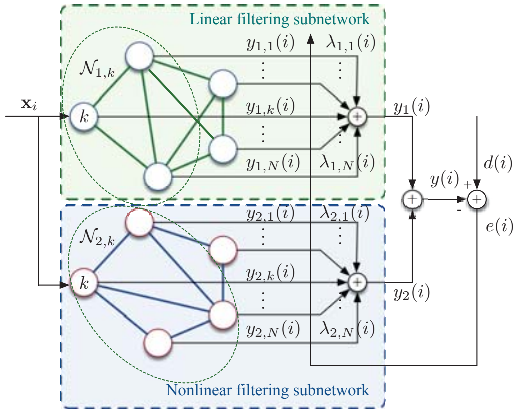


Fig. 1. Structure of the proposed generalized CNAF, consisting of a linear filtering subnetwork and a nonlinear filtering subnetwork.

The outputs of node k in the linear and nonlinear subnetworks at instant i are given, respectively, by:

$$y_{1,k}(i) = \mathbf{x}_i^\top \mathbf{h}_{k,i}^{\mathrm{lin}}, \tag{5}$$

$$y_{2,k}(i) = \mathbf{f}(\mathbf{x}_i)^\top \mathbf{h}_{k,i}^{\mathrm{nlin}}, \tag{6}$$

where $\mathbf{h}_{k,i}^{\mathrm{lin}}$ and $\mathbf{h}_{k,i}^{\mathrm{nlin}}$ denote, respectively, the linear and nonlinear filter at node k and time instant i, and $\mathbf{f}(\mathbf{x}_i)$ is a nonlinear map of $\mathbf{x}_i$.

The output of GCNAF is written as

$$y(i) = \sum_{k=1}^{N} \Big[\lambda_{1,k}(i) y_{1,k}(i) + \lambda_{2,k}(i) y_{2,k}(i)\Big], \tag{7}$$

where N denotes the number of nodes and $\lambda_{1,k}(i)$, $\lambda_{2,k}(i)$ are mixing parameters. The error signal, denoted as $e(i)$, is consequently defined by

$$e(i) \triangleq d(i) - y(i). \tag{8}$$

We also define the local errors $e_{1,k}(i)$ and $e_{2,k}(i)$, which are associated to node k of the linear and nonlinear subnetworks, respectively by:

$$e_{1,k}(i) \triangleq d(i) - \mathbf{x}_i^\top \mathbf{h}_{k,i-1}^{\mathrm{lin}} - \sum_{n=1}^{N} \lambda_{2,n}(i) \mathbf{f}(\mathbf{x}_i)^\top \mathbf{h}_{n,i-1}^{\mathrm{nlin}}, \tag{9}$$

$$e_{2,k}(i) \triangleq d(i) - \mathbf{f}(\mathbf{x}_i)^\top \mathbf{h}_{k,i-1}^{\mathrm{nlin}} - \sum_{n=1}^{N} \lambda_{1,n}(i) \mathbf{x}_i^\top \mathbf{h}_{n,i-1}^{\mathrm{lin}}. \tag{10}$$

We will use $e(i)$ to adjust mixing parameters $\lambda_{1,k}(i)$ and $\lambda_{2,k}(i)$, and use $e_{1,k}(i)$ and $e_{2,k}(i)$ to adjust the estimates $\mathbf{h}_{k,i}^{\mathrm{lin}}$ and $\mathbf{h}_{k,i-1}^{\mathrm{nlin}}$.

B. *Adaptation of the linear and nonlinear filters*

The proposed GCNAF scheme using the ATC diffusion LMS algorithm. According to (2), the update of the linear filters can be written as

$$\boldsymbol{\phi}_{k,i}^{\mathrm{lin}} = \mathbf{h}_{k,i-1}^{\mathrm{lin}} + \mu_{1,k} \sum_{\ell \in \mathcal{N}_{1,k}} c_{1,\ell k} \mathbf{x}_i e_{1,\ell}(i), \tag{11}$$

$$\mathbf{h}_{k,i}^{\mathrm{lin}} = \sum_{\ell \in \mathcal{N}_{1,k}} a_{1,\ell k}\, \boldsymbol{\phi}_{\ell,i}^{\mathrm{lin}}. \tag{12}$$

Similarly, the update of the nonlinear filters can be written as

$$\boldsymbol{\phi}_{k,i}^{\mathrm{nlin}} = \mathbf{h}_{k,i-1}^{\mathrm{nlin}} + \mu_{2,k} \sum_{\ell \in \mathcal{N}_{k}} c_{2,\ell k} \mathbf{f}(\mathbf{x}_i) e_{2,\ell}(i), \tag{13}$$

$$\mathbf{h}_{k,i}^{\mathrm{nlin}} = \sum_{\ell \in \mathcal{N}_{1,k}} a_{1,\ell k}\, \boldsymbol{\phi}_{\ell,i}^{\mathrm{nlin}}. \tag{14}$$

C. Adaptation of the mixing parameters

Updating the mixing parameters $\lambda_{1,k}(i)$ and $\lambda_{2,k}(i)$, $k = 1, 2, \cdots, N$, can be performed by minimizing the squared error $e^2(i)$. In order to address the constraints on $\lambda_{1,k}(i)$ and $\lambda_{2,k}(i)$, we reparameterize them with a softmax activation function, i.e.:

$$\lambda_{\ell,k}(i) = \frac{\exp(\alpha_{\ell,k}(i))}{\sum_{j=1}^{N} \exp(\alpha_{\ell,j}(i))}, \quad \text{for}\ \ell = 1, 2, j = 1, \cdots, N, \tag{15}$$

which guarantees that $0 < \lambda_{\ell,k}(i) < 1$ and $\sum_{k=1}^{N} \lambda_{\ell,k}(i) = 1$. Applying the gradient descent adaptation with respect to $\alpha_{\ell,k}(i)$ leads to

$$\alpha_{l,k}(i) = \alpha_{l,k}(i-1) + \frac{\eta_{l,k}}{r_{l,k}(i)} e(i)(e(i) - e_{l,k}(i))\lambda_{l,k}(i), \quad \text{for}\ \ \ell = 1, 2, \tag{16}$$

where $\eta_{\ell,k}$ are step size parameters, and $r_{l,k}(i) = \beta r_{l,k}(i-1) + (1-\beta)(e(i) - e_{l,k}(i))^2$ are rough low-pass filtered estimates and the parameter β is a smoothing factor. The value of $\alpha_{l,k}(i)$ is kept within [-4,4] for practical reasons [30].

Now considering the NAEC problem of online system estimation, the entire algorithm is summarized in Algorithm 1.

IV. Simulations

In this section, simulations were conducted to illustrate the performance of the proposed GCNAF scheme under different linear-to-nonlinear distortion power ratios (LNLRs) and to compare the results with those of several other algorithms of NAEC.

The reference signal $d(i)$ was considered by using the following input-output relation

$$d(i) = \mathbf{h}_{\mathrm{T}}^{\top} \left[\mathbf{x}_i + \sigma(i)\mathbf{f}(\mathbf{x}_i)\right] + v(i), \tag{21}$$

with $v(i)$ being additive white Gaussian noise. The variance of $v(i)$ was adjusted so that SNR = 30 dB. Considering $\mathbf{h}_{\mathrm{T}}^{\mathrm{lin}} = \mathbf{h}_{\mathrm{T}}$ and $\mathbf{h}_{\mathrm{T}}^{\mathrm{nlin}} = \sigma(i)\mathbf{h}_{\mathrm{T}}$. Note that the parameter $\sigma(i)$ allows us to adjust the LNLR, namely,

$$\mathrm{LNLR} = 10\log_{10}\left(\frac{\mathbb{E}\{\|\mathbf{x}_i\|^2\}}{\mathbb{E}\{\|\sigma(i)\mathbf{f}(\mathbf{x}_i)\|^2\}}\right). \tag{22}$$

Algorithm 1: GCNAF algorithm for NAEC.

Parameters: Preset
- positive step size $\mu_{i,j}$ for all nodes (filters);
- right-stochastic matrices $\mathbf{C}_1$ and $\mathbf{C}_2$;
- left-stochastic matrices $\mathbf{A}_1$ and $\mathbf{A}_2$;

Initialization: Set $\mathbf{h}_{k,0}^{\mathrm{lin}} = \mathbf{0}$, and $\mathbf{h}_{k,0}^{\mathrm{nlin}} = \mathbf{0}$.

Algorithm: At each time instant $i \geq 1$:

- Node k in the linear subnetwork updates $\mathbf{h}_{k,i}^{\mathrm{lin}}$ by:

$$\boldsymbol{\phi}_{k,i}^{\mathrm{lin}} = \mathbf{h}_{k,i-1}^{\mathrm{lin}} + \mu_{1,k} \sum_{\ell \in \mathcal{N}_{1,k}} c_{1,\ell k} \mathbf{x}_i e_{1,\ell}(i), \tag{17}$$

$$\mathbf{h}_{k,i}^{\mathrm{lin}} = \sum_{\ell \in \mathcal{N}_{1,k}} a_{1,\ell k}\, \boldsymbol{\phi}_{\ell,i}^{\mathrm{lin}}, \tag{18}$$

with $e_{1,k}(i) = d(i) - \mathbf{x}_i^{\top}\mathbf{h}_{k,i-1}^{\mathrm{lin}} - \sum_{n=1}^{N}\lambda_{2,n}(i)\mathbf{f}(\mathbf{x}_i)^{\top}\mathbf{h}_{n,i-1}^{\mathrm{nlin}}$.

- Node k in the nonlinear subnetwork updates $\mathbf{h}_{k,i}^{\mathrm{nlin}}$ by:

$$\boldsymbol{\phi}_{k,i}^{\mathrm{nlin}} = \mathbf{h}_{k,i-1}^{\mathrm{nlin}} + \mu_{2,k} \sum_{\ell \in \mathcal{N}_{k}} c_{2,\ell k} \mathbf{f}(\mathbf{x}_i) e_{2,\ell}(i), \tag{19}$$

$$\mathbf{h}_{k,i}^{\mathrm{nlin}} = \sum_{\ell \in \mathcal{N}_{1,k}} a_{1,\ell k}\, \boldsymbol{\phi}_{\ell,i}^{\mathrm{nlin}}, \tag{20}$$

with $e_{2,k}(i) = d(i) - \mathbf{f}(\mathbf{x}_{k,i})^{\top}\mathbf{h}_{k,i-1}^{\mathrm{nlin}} - \sum_{n=1}^{N}\lambda_{1,n}(i)\mathbf{x}_i^{\top}\mathbf{h}_{n,i-1}^{\mathrm{lin}}$.

- Update the combination weights $\{\lambda_{i,j}\}$ by (15) and (16).

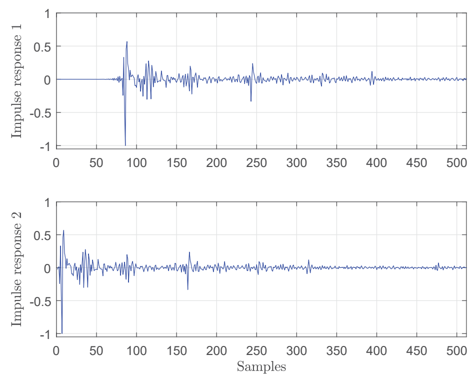


Fig. 2. The impulse responses of echo path used for $\mathbf{h}_{\mathrm{T}}$ in simulations.

In our work, the echo path was measured in a real room within our laboratory, using the sampling rate of 8 kHz, and the measured impulse responses were then truncated to 512 samples and used as $\mathbf{h}_{\mathrm{T}}$ in simulations, as shown in Fig. 2.

The proposed scheme with two different types of inputs were tested: a stationary white Gaussian process and a speech signal. Note that speech signals are highly non-stationary and self-correlated. For comparison, the linear normalized least-mean-square (LNLMS) algorithm and the CK algorithm [11] were also evaluated. The echo return loss enhancement (ERLE) was used to evaluate the performance, which is defined

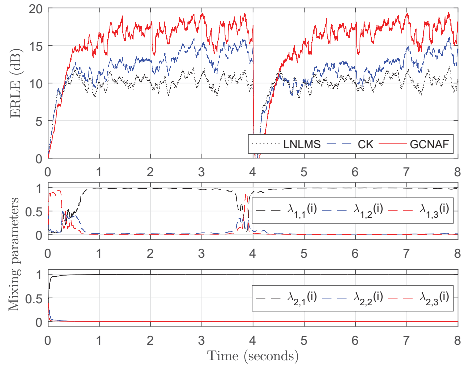


Fig. 3. Performance illustration with $\mathbf{h}_\mathrm{T}$ abruptly changed at $i = 4$ s with LNLR= 10 dB and SNR= 30 dB. From top to bottom: ERLE evolution of GCNAF, CK and LNLMS; evolution of the GCNAF mixing parameters $\{\lambda_{1,k}\}$ and $\{\lambda_{2,k}\}$.

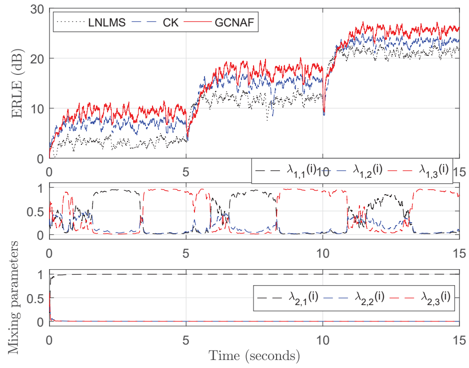


Fig. 4. Performance evolution with SNR= 30 dB. LNLR has been initially set to 0 dB, and then changed to 10 dB and 20 dB at $i = 5$ s and $i = 10$ s, respectively. From top to bottom: ERLE evolution of GCNAF, CK and LNLMS; evolution of the GCNAF mixing parameters $\{\lambda_{1,k}\}$ and $\{\lambda_{2,k}\}$.

by

$$\mathrm{ERLE}(i) \triangleq 10\log_{10}\left(\frac{\mathbb{E}\{d^2(i)\}}{\mathbb{E}\{e^2(i)\}}\right), \tag{23}$$

evaluated by averaging 100 independent Mont-Carlo runs.

A. Simulations with white Gaussian input

We first used the white Gaussian random process as the input. For GCNAF and CK, the number of nodes (filters in each subnetwork) was set to 3, and step sizes were set to $\mu_{1,1} = 0.001$, $\mu_{1,2} = 0.01$, $\mu_{1,3} = 0.5$, $\mu_{2,1} = 0.001$, $\mu_{2,2} = 0.01$, $\mu_{2,3} = 0.5$. The mixing parameters $\{\lambda_{1,k}\}$ and $\{\lambda_{2,k}\}$ were adapted by using $\eta_{1,k} = \eta_{2,k} = 1$ and $\beta = 0.9$.

For GCNAF, we considered a fully connected graph, i.e., all nodes were connected with $\mathcal{N}_{1,k} = \mathcal{N}_{2,k} = \{1, 2, 3\}$ for $\forall k$. The diffusion matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ were set to

$$\mathbf{C}_1 = \mathbf{C}_2 = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}. \tag{24}$$

Matrix $\mathbf{A}_1 = \mathbf{C}_1$ and $\mathbf{A}_2 = \mathbf{C}_2$ were used. Recall that the CK algorithm corresponds to the setting $\mathbf{C}_1 = \mathbf{C}_2 = \mathbf{I}$ and $\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{I}$.

The polynomial nonlinearity [11] settings was used for testing the performance:

$$f_i(\mathbf{x}_i) = x^2(i) + \frac{9}{10}x^2(i-1) + \frac{1}{2}x^3(i-2). \tag{25}$$

In the first simulation, we studied the performance of GCNAF in the NAEC scenario with an abruptly change of the echo path, i.e., $\mathbf{h}_\mathrm{T}$ changed from the impulse response 1 to impulse response 2 in Fig. 3 at $i = 4$ s. The top of Fig. 3 plots the ERLE evolution of GCNAF and the compared algorithms. One can see that GCNAF outperforms the other two algorithms in steady-state performance. The evolution of the mixing parameters is shown in the bottom of Fig. 3. In this NAEC scenario where the linear component is dominant, the linear subnetwork requires a larger step size at the initial phase.

In the second simulation, we studied the performance of GCNAF with different degrees of nonlinearity $\mathbf{f}(\mathbf{x}_i)$. The parameter $\sigma(i)$ in (21) was adjusted so that the LNLR is initially equal to 0 dB, and then changed to 10 dB and 20 dB at $i = 5$ s and $i = 10$ s. We observe that during all the phases, the proposed GCNAF method outperformed the LNLMS and CK algorithms. The LNLMS algorithm yielded the worst performance since it was not able to model the nonlinearity in the system. The CK algorithm showed a significantly better performance compared to LNLMS, but it was inferior to the proposed GCNAF method. This indicates that the information exchange in the proposed networked filters helped enhance the estimation performance. The evolution of the mixing parameters is shown in the bottom of Fig. 4.

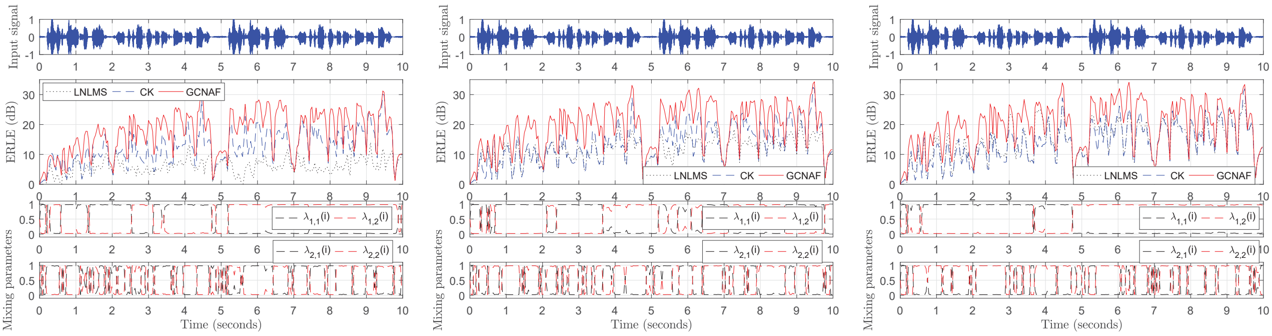
B. Simulations with a speech input

In this part, we examined the performance of the studied algorithms with a segment of speech signal as the system input. The waveform of the speech signal is plotted on Fig. 5. The polynomial nonlinearity setting (25) was used, and the parameter σ was adjusted to achieve LNLR = 0, 10, and 20 dB respectively. For the GCNAF and CK algorithms, the number of linear and nonlinear filters was set to 2, and step sizes of filter were set to $\mu_{1,1} = 0.05$, $\mu_{1,2} = 0.5$, $\mu_{2,1} = 0.05$, $\mu_{2,2} = 0.5$. The parameters for updating $\{\lambda_{1,k}\}$ and $\{\lambda_{2,k}\}$ were set to $\eta_{1,k} = \eta_{2,k} = 1$ and $\beta = 0.9$. We considered the diffusion algorithm with the following doubly stochastic matrices:

$$\mathbf{C}_1 = \mathbf{C}_2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}. \tag{26}$$

Matrix $\mathbf{A}_1 = \mathbf{C}_1$ and $\mathbf{A}_2 = \mathbf{C}_2$ are used.

The middle of Figs. 5(a), 5(b) and 5(c) plot the ERLE results achieved by GCNAF and the other two studied algorithms under three different LNLR levels respectively. One can see from Fig. 5(a) that the LNLMS algorithm is significantly



(a) LNLR= 0 dB (b) LNLR= 10 dB (c) LNLR= 20 dB

Fig. 5. Performance when using a speech signal as the system input with SNR= 30 dB. From top to bottom: input signal; ERLE evolution of GCNAF, CK and LNLMS; evolution of the GCNAF mixing parameters $\{\lambda_{1,k}\}$ and $\{\lambda_{2,k}\}$.

inferior to the other two studied algorithms in performance in the presence of strong nonlinearities. In Fig. 5(c), the performance difference between the GCNAF and LNLMS algorithm becomes smaller since the linear component is now dominant. All the results demonstrate the advantage of GCNAF over the CK and LNLMS algorithms, particularly in the case of LNLR= 0 dB, where the nonlinearity is significant. The time evolution of the mixing parameters are plotted in the bottom of Figs. 5(a), 5(b) and 5(c).

V. Conclusions

In this paper, a new and generalized CNAF (GCNAF) for NAEC was proposed based on the diffusion strategy over networks. In contrast to the existing CNAF, the proposed GCNAF links the nodes via the topological structure of the network in which each node represents either a linear or a nonlinear filter. The adaptation performance was significantly enhanced as compared to CNAF due to the information exchange within the network. Simulations with noise and speech signals demonstrated the effectiveness of the proposed GCNAF for an NAEC problem.

References

[1] M. M. Sondhi, "The history of echo cancellation," *IEEE Signal Process. Mag.*, vol. 23, no. 5, pp. 95–98, 2006.

[2] D. Comminiello, L. A. Azpicueta-Ruiz, M. Scarpiniti, A. Uncini, and J. Arenas-Garcia, "Functional link based architectures for nonlinear acoustic echo cancellation," in *Proc. IEEE HSCMA11*, Edinburgh, U.K., May 2011, pp. 180–184.

[3] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-Garcia, and A. Uncini, "Functional link adaptive filters for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Process.*, Vol. 21, No. 7, pp. 1502–1512, 2013.

[4] D. Comminiello, M. Scarpiniti, L. A. Azpicueta-Ruiz, J. Arenas-Garcia, and A. Uncini, "Nonlinear acoustic echo cancellation based on sparse functional link representations," *IEEE/ACM Trans. Audio, Speech, Language Process.*, Vol. 22, No. 7, pp. 1172–1183, 2014.

[5] A. N. Birkett, and R. A. Goubran, "Acoustic echo cancellation using NLMS-neural network structures," in *Proc. IEEE ICASSP*, Detroit, MI, May 1995, pp. 3035–3038.

[6] J. P. Costa, A. Lagrange, and A. Arliaud, "Acoustic echo cancellation using nonlinear cascade filters," in *Proc. IEEE ICASSP*, HongKong, Apr. 2003, pp. 389–392.

[7] J. Vörös, "An iterative method for Hammerstein-Wiener systems parameter identification," *J. Elect. Eng.*, Vol. 55, pp. 328–331, 2004.

[8] M. Zeller, and W. Kellermann, "Coefficient pruning for higher-order diagonals of Volterra filters representing Wiener-Hammerstein models," in *Proc. IWAENC*, Seattle, WA, Sep. 2008.

[9] V. J. Mathews, and G. L. Sicuranza, "Polynomial signal processing," *New York: Wiley*, 2000.

[10] A. Guérin, G. Faucon, and R. L. Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," in *IEEE Trans. Speech, Audio Process.*, Vol. 11, No. 6, pp. 672–683, 2003.

[11] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, J. Arenas-Garcia, 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, 2011.

[12] L. A. Azpicueta-Ruiz, M. Zeller, A. R. Figueiras-Vidal, W. Kellermann, and J. Arenas-Garcia , "Enhanced adaptive Volterra filtering by automatic attenuation of memory regions and its application to acoustic echo cancellation," *IEEE Trans. Audio, Speech, Language Proce.*, Vol. 61, No. 11, pp. 2745–2713, 2013.

[13] J. Tsitsiklis, and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Trans. Autom. Control*, Vol. 29, No. 1, pp. 42–50, 1984.

[14] L. Xiao, and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 9, pp. 65–78, 2004.

[15] P. Braca, S. Marano, and V. Matta, "Running consensus in wireless sensor networks," in *Proc. Fusion*, Cologne, Germany, 2008, pp. 1–6.

[16] A. Nedic, and A. Ozdaglar, "Distributed subgradient methods formultiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.

[17] S. Kar, and J. M. F. Moura, "Distributed consensus algorithms in sensor networks: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, 2009.

[18] K. Srivastava, and A. Nedic, "Distributed asynchronous constrained stochastic optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 772–790, 2011.

[19] A. Nedic, and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. Optimiz*, vol. 12, no. 1, pp. 109–138, 2001.

[20] M. G. Rabbat, and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE J. Sel. Areas Commun*, vol. 23, no. 4, pp. 798–808, 2005.

[21] D. Blatt, A. O. Hero, and H. Gauchman, "A convergent incremental gradient method with constant step size," *SIAM J. Optimiz*, vol. 18, no. 1, pp. 29–51, 2007.

[22] C. G. Lopes, and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4064–4077, 2007.

[23] A. H. Sayed, S. Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Process. Mag*, vol. 30, no. 3, pp. 155–171, 2013.

[24] A. H. Sayed, "Diffusion adaptation over networks," in *Academic Press Libraray in Signal Processing*, R. Chellapa and S. Theodoridis, Eds. Amsterdam, The Netherlands: Elsevier, 2014, pp. 322–454.

[25] C. G. Lopes, and A. H. Sayed, "Diffusion least-mean squares over

adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, 2008.

[26] F. S. Cattivelli, and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, 2010.

[27] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4129–4143, 2014.

[28] J. Chen, C. Richard, and A. H. Sayed, "Diffusion LMS over multitask networks," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2733–2748, 2015.

[29] D. Jin, J. Chen, J. Chen, "Convex combination of diffusion strategies of distributed newtorks", in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hawaii, U.S., Nov. 2018.

[30] J. Arenas-García, V. Gómez-Verdejo, and A. R. Figueiras-Vidal, "New algorithms for improved adaptive convex combination of LMS transversal filters," *IEEE Trans. Instrum. Measure.*, Vol. 54, No. 6, pp. 2239–2249, 2005.