NONLINEAR ACOUSTIC ECHO CANCELLATION BASED ON A MULTIPLICATIVE TRANSFER FUNCTION APPROXIMATION

Yekutiel Avargel and Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
{kutiav@tx,icohen@ee}.technion.ac.il

ABSTRACT

In this paper, a new nonlinear model for improved acoustic echo cancellation in the short-time Fourier transform domain is introduced. The model consists of a parallel combination of linear and quadratic components. The linear component is represented by multiplicative terms, while the quadratic component is modeled by multiplicative *cross-terms*. We show that for low signal-to-noise ratio (SNR) conditions, a lower mean-square error is achieved by allowing for nonlinear undermodeling and employing only the linear multiplicative transfer function (MTF) model. However, as the SNR increases, the performance can be generally improved by the proposed nonlinear model. A significant reduction in computational cost as well as an improvement in estimation accuracy is achieved over the time-domain Volterra approach. Experimental results demonstrate the advantage of the proposed model for nonlinear acoustic echo cancellation.

Index Terms— Nonlinear acoustic echo cancellation, multiplicative transfer function, short-time Fourier transform, nonlinear undermodeling.

1. INTRODUCTION

Loudspeaker-enclosure-microphone (LEM) system modeling in the short-time Fourier transform (STFT) domain is of major importance in many acoustic echo cancellation applications, especially when long echo paths are considered [1]. The multiplicative transfer function (MTF) approximation [2], which relies on the assumption of a large analysis window length, is widely-used in such applications due to computational efficiency (e.g., [3, 4]). However, in many cases, particularly when small loudspeakers are driven at high volumes, the LEM system often exhibits certain nonlinearities that cannot be sufficiently estimated by the linear MTF model. Volterra filters used for modeling the nonlinear LEM system [5, 6] often suffer from extremely high computational cost due to a large number of parameters. This problem becomes even more crucial when estimating systems with relatively large memory length, which is often the case in acoustic echo cancellation applications.

In this paper, we extend the MTF approximation and introduce a new nonlinear model for improved acoustic echo cancellation in the STFT domain. The proposed model consists of a parallel combination of linear and quadratic components. The linear component is represented by the MTF approximation, while the quadratic component is modeled by multiplicative cross-terms. The quadratic-component model has been recently introduced in [7], and is based on a time-frequency representation of a homogeneous second-order Volterra filter. We consider an off-line echo cancellation scheme based on a least-squares (LS) criterion, and analyze the obtainable mean-square error (mse) in each frequency bin. We mainly concentrate on the error arises due to *nonlinear undermodeling*; that is, when the linear MTF model is utilized for estimating the nonlinear LEM system. We show that for low signal-to-noise ratio (SNR) conditions, a lower mse is achieved by using the MTF model and allowing for nonlinear undermodeling. However, as the SNR increases, the acoustic echo canceller (AEC) performance can be generally improved by employing the proposed nonlinear model. When compared to the conventional time-domain Volterra approach, a significant reduction in computational complexity is achieved by the proposed approach, especially when long-memory systems are considered. Experimental results demonstrate the advantage of the proposed approach for nonlinear acoustic echo cancellation.

The paper is organized as follows. In Section 2, we introduce a new nonlinear STFT model that is based on the MTF approximation. In Section 3, we present an off-line echo cancellation scheme for estimating the model parameters. In Section 4, we derive expressions for the obtainable mse, and investigate the influence of nonlinear undermodeling on the AEC performance. Finally, in Section 5, we present experimental results which support the theoretical derivations.

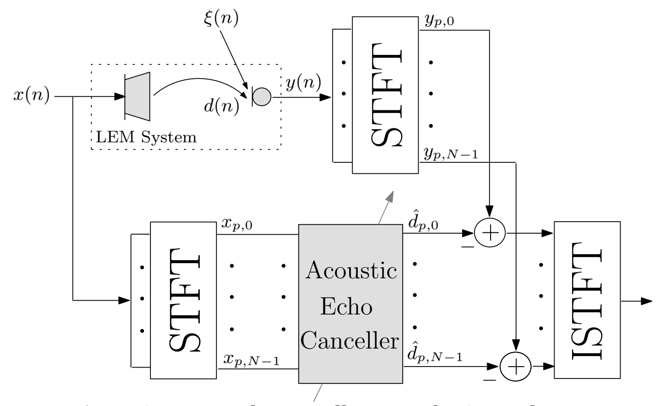


Fig. 1. Acoustic echo cancellation in the STFT domain.

2. MODELING THE LEM SYSTEM

A typical acoustic echo cancellation scheme in the STFT domain is illustrated in Fig. 1. The far-end signal $x(n)$ is emitted by a loudspeaker, then propagates through the enclosure and received in

This research was supported by the Israel Science Foundation (grant no. 1085/05).

the microphone as an echo signal $d(n)$. Together with a near-end speech signal and local noise [collectively denoted by $\xi(n)$], the microphone signal can be written as $y(n) = d(n) + \xi(n)$. Applying the STFT to $y(n)$, we have in the time-frequency domain

$$y_{p,k} = d_{p,k} + \xi_{p,k} \tag{1}$$

where p is the frame index and k represents the frequency-bin index ($0 \le k \le N-1$). To produce an echo estimate $\hat{d}_{p,k}$ in the time-frequency domain, a proper STFT model for the LEM system is needed. The widely-used MTF approximation [2] assumes a relatively large analysis-window length to approximate the system as multiplicative in the STFT domain, i.e.,

$$\hat{d}_{p,k} = h_k \, x_{p,k} \, . \tag{2}$$

The effectiveness of the MTF approximation in estimating linear systems has been demonstrated in [3]. However, in many acoustic echo cancellation applications, particularly when small loudspeakers are driven at high volumes, the LEM system often exhibits certain nonlinearities that cannot be sufficiently estimated by the conventional MTF model.

For improved nonlinear echo cancellation, we may extend the MTF approximation by incorporating a nonlinear component into the model. To do so, we employ the nonlinear model defined in [7], which is based on the time-frequency representation of homogeneous Volterra filters. Since the nonlinearity of loudspeakers can be assumed to be limited up to the second order [5], we consider here only the quadratic case. Accordingly, the output of the proposed nonlinear AEC is given as a parallel combination of linear and quadratic components in the time-frequency domain as follows:

$$\begin{aligned} \hat{d}_{p,k} =& h_k \, x_{p,k} \\ &+ \gamma \sum_{k' \in \mathcal{F}} x_{p,k'} x_{p,(k-k') \bmod N} c_{k',(k-k') \bmod N} \end{aligned} \tag{3}$$

where $\gamma \in \{0,1\}$, $c_{k',(k-k') \bmod N}$ is referred to as a *quadratic cross-term*, and $\mathcal{F} = \{0, 1, \ldots \lfloor k/2 \rfloor, k+1, \ldots, k+1+\lfloor (N-k-2)/2 \rfloor\}$. The conventional MTF approximation is used in (3) for representing the linear component of the system. The cross-terms $\{c_{k',(k-k') \bmod N} \big| \, k' \in \mathcal{F}\}$, on the other hand, are used for modeling the quadratic component of the system using a sum over all possible interactions between pairs of input frequencies $x_{p,k'}$ and $x_{p,k''}$, such that only frequency indices $\{k', k''\}$, whose sum is k or $k+N$, contribute to the output at frequency bin k. Note that γ controls the nonlinear undermodeling as it determines whether a linear or a nonlinear model is considered. By setting $\gamma = 0$, the nonlinearity is ignored and the linear MTF model is fitted to the data, which may degrade the system estimate accuracy. The influence of the parameter γ on the mean-square performance is investigated in Section 4.

3. OFF-LINE CANCELLATION SCHEME

In this section, we introduce an LS-based off-line algorithm for echo cancellation using the proposed nonlinear STFT model. We denote by P the number of samples in a time-trajectory of $x_{p,k}$. Let $\mathbf{x}_k = \begin{bmatrix} x_{0,k} & x_{1,k} & \cdots & x_{P-1,k} \end{bmatrix}^T$ denote a time-trajectory of x_{pk} at frequency bin k, and let the vectors $\mathbf{d}_k$, $\boldsymbol{\xi}_k$ and $\mathbf{y}_k$ be defined similarly. For notational simplicity, let us assume that k and N are both even, such that according to (3), the number of quadratic cross-terms in each frequency bin is $N/2+1$. Then, let

$$\mathbf{c}_k = \begin{bmatrix} c_{0,k} & \cdots & c_{\frac{k}{2},\frac{k}{2}} & c_{k+1,N-1} & \cdots & c_{\frac{N+k}{2},\frac{N+k}{2}} \end{bmatrix}^T \tag{4}$$

denote the quadratic cross-terms at the kth frequency bin, and let $\boldsymbol{\Lambda}_k = \begin{bmatrix} \mathbf{x}_{0,k} & \cdots & \mathbf{x}_{\frac{k}{2},\frac{k}{2}} & \mathbf{x}_{k+1,N-1} & \cdots & \mathbf{x}_{\frac{N+k}{2},\frac{N+k}{2}} \end{bmatrix}$ be an $P \times (N/2+1)$ matrix, where $\mathbf{x}_{k,k'} = \mathbf{x}_k \odot \mathbf{x}_{k'}$, and $\odot$ denotes a term-by-term multiplication. Then, the AEC output signal (3) can be written in a vector form as

$$\hat{\mathbf{d}}_{\gamma k}(\boldsymbol{\theta}_k) = \mathbf{x}_k h_k + \gamma \boldsymbol{\Lambda}_k \mathbf{c}_k \triangleq \mathbf{R}_{\gamma k} \boldsymbol{\theta}_k \tag{5}$$

where $\mathbf{R}_{\gamma k} = [\mathbf{x}_k \;\; \gamma \boldsymbol{\Lambda}_k]$, and $\boldsymbol{\theta}_k = \begin{bmatrix} h_k \; \mathbf{c}_k^T \end{bmatrix}^T$ is the model parameters vector. The subscript γ in $\hat{\mathbf{d}}_{\gamma k}(\boldsymbol{\theta}_k)$ indicates the dependence of the echo estimate on the model structure, which can be either linear or nonlinear. Finally, using the above notations, the LS estimate of the model parameters at the kth frequency bin is given by

$$\hat{\boldsymbol{\theta}}_{\gamma k} = \arg\min_{\boldsymbol{\theta}_k} \|\mathbf{y}_k - \mathbf{R}_{\gamma k}\boldsymbol{\theta}_k\|^2 = \mathbf{R}_{\gamma k}^{\dagger} \mathbf{y}_k \tag{6}$$

where $\mathbf{R}_{\gamma k}^{\dagger} = (\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k})^{-1} \mathbf{R}_{\gamma k}^H$ is the Moore-Penrose pseudo inverse matrix of $\mathbf{R}_{\gamma k}$. Substituting (6) into (5), we obtain the best estimate of the echo signal in the STFT domain $\hat{\mathbf{d}}_{\gamma k}(\hat{\boldsymbol{\theta}}_{\gamma k})$ in the LS sense, for a given γ value.

4. MSE ANALYSIS

In this section, we derive expressions for the mse obtainable in the kth frequency bin, and investigate the influence of nonlinear undermodeling (controlled by γ) on the AEC performance. For a tractable analysis, we assume that $x_{p,k}$ and $\xi_{p,k}$ are zero-mean white Gaussian signals with variances σ_x^2 and σ_ξ^2, respectively, and that they are statistically independent.

4.1. Relations Between MSE and SNR

The (normalized) mse is defined by

$$\epsilon_{\gamma k} = \frac{1}{E\left\{\|\mathbf{d}_k\|^2\right\}} E\left\{ \left\| \mathbf{d}_k - \hat{\mathbf{d}}_{\gamma k}(\hat{\boldsymbol{\theta}}_{\gamma k}) \right\|^2 \right\} \tag{7}$$

where $E\{\cdot\}$ denotes expectation. Recall that ϵ_{0k} denotes the mse obtained by using only the linear MTF model, and ϵ_{1k} is the mse achieved by incorporating also a quadratic component into the model [see (3)]. Substituting (5) and (6) into (7), the mse can be expressed as

$$\epsilon_{\gamma k} = 1 + \frac{\epsilon_1 - \epsilon_2}{E\left\{\|\mathbf{d}_k\|^2\right\}} \tag{8}$$

where $\epsilon_1 = E\{\boldsymbol{\xi}_k^H \mathbf{R}_{\gamma k} \mathbf{R}_{\gamma k}^{\dagger} \boldsymbol{\xi}_k\}$ and $\epsilon_2 = E\{\mathbf{d}_k^H \mathbf{R}_{\gamma k} \mathbf{R}_{\gamma k}^{\dagger} \mathbf{d}_k\}$. Using the whiteness assumption for $\xi_{p,k}$ and the property that $\mathbf{a}^H \mathbf{b} = tr(\mathbf{a}\mathbf{b}^H)^*$ for any two vectors $\mathbf{a}$ and $\mathbf{b}$, ϵ_1 can be expressed as

$$\begin{aligned} \epsilon_1 &= tr\left(E\left\{ \boldsymbol{\xi}_k \boldsymbol{\xi}_k^H \right\} E\{\mathbf{R}_{\gamma k} \mathbf{R}_{\gamma k}^{\dagger}\} \right)^* \\ &= \sigma_\xi^2 E\left\{ tr\left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k} \left(\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k} \right)^{-1} \right)^* \right\} \\ &= \sigma_\xi^2 \left[1 + \gamma \left(N/2 + 1 \right) \right] . \end{aligned} \tag{9}$$

For evaluating ϵ_2, let us assume that $x_{p,k}$ is ergodic and that the data length P is sufficiently large. From (5), the inverse of $\mathbf{R}_{\gamma k}^H \mathbf{R}_{\gamma k}$ can be expressed as

$$\left(\mathbf{R}_{\gamma k}^{H}\mathbf{R}_{\gamma k}\right)^{-1}=\begin{bmatrix}\mathbf{x}_{k}^{H}\mathbf{x}_{k} & \gamma\mathbf{x}_{k}^{H}\mathbf{\Lambda}_{k}\\ \gamma\mathbf{\Lambda}_{k}\mathbf{x}_{k}^{H} & \gamma\mathbf{\Lambda}_{k}^{H}\mathbf{\Lambda}_{k}\end{bmatrix}^{-1} \tag{10}$$

where from the ergodicity, the ℓth term of $\mathbf{\Lambda}_k^H\mathbf{x}_k$ may be approximated as $\left(\mathbf{\Lambda}_k^H\mathbf{x}_k\right)_\ell \approx PE\{x_{m,\ell_k}^{*}x_{m,(k-\ell_k)\bmod N}^{*}x_{m,k}\}$ where $\ell_k=\ell$ if $\ell\le k/2$, and $\ell_k=\ell+k/2$ otherwise. Since odd-order moments of a zero-mean complex Gaussian process are zero [8], we get $\left(\mathbf{\Lambda}_k^H\mathbf{x}_k\right)_\ell\approx 0$, and (10) reduces to

$$\left(\mathbf{R}_{\gamma k}^{H}\mathbf{R}_{\gamma k}\right)^{-1}\approx\begin{bmatrix}\left(\mathbf{x}_{k}^{H}\mathbf{x}_{k}\right)^{-1} & \mathbf{0}_{1\times N/2+1}\\ \mathbf{0}_{N/2+1\times 1} & \gamma\left(\mathbf{\Lambda}_{k}^{H}\mathbf{\Lambda}_{k}\right)^{-1}\end{bmatrix} \tag{11}$$

where $\mathbf{0}_{N\times 1}$ is a zero vector of size $N\times 1$. Substituting (11) into the expression for ϵ_2, we obtain

$$\epsilon_2=\epsilon_{12}+\gamma\epsilon_{22} \tag{12}$$

where $\epsilon_{12}=E\{\mathbf{d}_k^H\mathbf{x}_k\mathbf{x}_k^{\dagger}\mathbf{d}_k\}$ and $\epsilon_{22}=E\{\mathbf{d}_k^H\mathbf{\Lambda}_k\mathbf{\Lambda}_k^{\dagger}\mathbf{d}_k\}$. Finally, denoting the SNR by $\eta=\sigma_d^2/\sigma_\xi^2$, where $\sigma_d^2=E\{|d_{p,k}|^2\}$, and substituting (9) and (12) into (8), we obtain

$$\epsilon_{\gamma k}=\frac{\alpha_{\gamma k}}{\eta}+\beta_{\gamma k} \tag{13}$$

where $\alpha_{\gamma k}=1/P+\gamma[N/2+1]/P$ and $\beta_{\gamma k}=1-\epsilon_{12}/(P\sigma_d^2)-\gamma\epsilon_{22}/(P\sigma_d^2)$. We observe from (13) that the mse $\epsilon_{\gamma k}$, for fixed values of γ and k, is a monotonically decreasing function of η. Note that ϵ_{22} can be rewritten as

$$\begin{aligned}\epsilon_{22}&=E\left\{\mathbf{d}_k^H(\mathbf{\Lambda}_k\mathbf{\Lambda}_k^{\dagger})^H\mathbf{\Lambda}_k\mathbf{\Lambda}_k^{\dagger}\mathbf{d}_k\right\}\\&=E\left\{\left\|\mathbf{\Lambda}_k\mathbf{\Lambda}_k^{\dagger}\mathbf{d}_k\right\|^2\right\}\ge 0\,.\end{aligned} \tag{14}$$

Then, following the nonnegativity of ϵ_{22}, it can be verified that $\alpha_{1k}>\alpha_{0k}$ and $\beta_{1k}\le\beta_{0k}$, which implies that $\epsilon_{1k}>\epsilon_{0k}$ for low SNR ($\eta<<1$), and $\epsilon_{1k}\le\epsilon_{0k}$ for high SNR ($\eta>>1$). Accordingly, for low SNR conditions, a lower mse is achieved by allowing for nonlinear undermodeling and employing the conventional linear MTF model in the estimation process. On the other hand, as the SNR increases, the mse performance can be generally improved by incorporating also the nonlinear component into the AEC ($\gamma=1$). These points will be further demonstrated in Section 5.

4.2. Computational Complexity

Forming the normal equations $(\mathbf{R}_{\gamma k}^H\mathbf{R}_{\gamma k})\hat{\boldsymbol{\theta}}_{\gamma k}=\mathbf{R}_{\gamma k}^H\mathbf{y}_k$ in (6), solving them using the Cholesky decomposition and calculating the desired signal estimate (5) for each frequency bin, require $NP[1+\gamma(N/2+1)]^2$ arithmetic operations, where P is assumed sufficiently large, and the computations required for the forward and inverse STFTs and neglected. The computational cost of the proposed approach is therefore $(N/2+1)^2$ times larger than that of the conventional MTF approach ($\gamma=0$). It should be noted here that a time-domain off-line estimation process with a second-order Volterra filter requires $PL\left[N_1+N_2\left(N_2+1\right)/2\right]^2$ arithmetic operations [7], where N_1 and N_2 are the memory length of the linear and quadratic Volterra kernels, respectively, and L is the translation factor of the STFT. For typical values of $N=256$, $L=128$ (i.e., 50% overlap between consecutive windows), $N_1=1024$ and $N_2=60$, the complexity of the proposed approach is reduced by approximately 250, when compared to the complexity of the time-domain Volterra approach.

5. EXPERIMENTAL RESULTS

In this section, we present experimental results that demonstrate the effectiveness of the proposed approach. In the first experiment, we examine the proposed AEC performance for white Gaussian signals, and demonstrate the influence of nonlinear undermodeling by fitting both linear and nonlinear models to the data. The input signal $x(n)$ and the additive noise signal $\xi(n)$ are uncorrelated zero-mean white Gaussian processes. The LEM system is assumed to be represented by a second-order Volterra filter, which relates the input $x(n)$ and output $y(n)$ as follows:

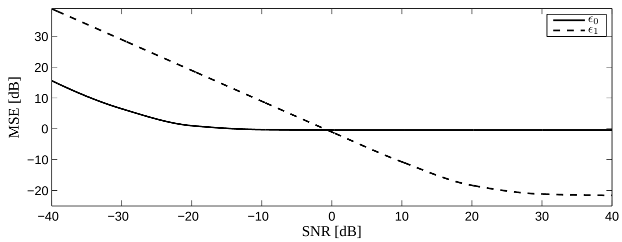
$$\begin{aligned}y(n)&=\sum_{m=0}^{N_1-1}h_1(m)x(n-m) \qquad (15)\\&\quad+\sum_{m=0}^{N_2-1}\sum_{\ell=0}^{N_2-1}h_2(m,\ell)x(n-m)x(n-\ell)+\xi(n)\end{aligned}$$

where $h_1(m)$ and $h_2(m,\ell)$ are the linear and quadratic Volterra kernels, respectively, and N_1 and N_2 are their corresponding memory lengths. The quadratic kernel is modeled as a unit variance zero-mean white Gaussian process, whereas the linear kernel is modeled as a stochastic process with an exponential decay envelope, i.e., $h(n)=u(n)\beta(n)e^{-0.009n}$ [where $u(n)$ is the unit step function and $\beta(n)$ is a unit-variance zero-mean white Gaussian process]. The memory lengths are set to $N_1=50$ and $N_2=40$. To maintain the large analysis-window support assumption, a Hamming analysis window of length $N=8N_1$ with 50% overlap is employed. The AEC performance is evaluated by the time-domain mse, defined by

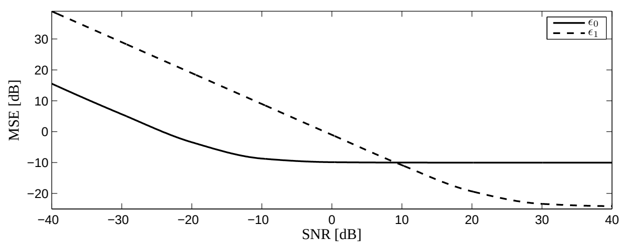
$$\epsilon_\gamma=\frac{1}{E\left\{|d\left(n\right)|^2\right\}}E\left\{\left|d\left(n\right)-\hat{d}_\gamma\left(n\right)\right|^2\right\} \tag{16}$$

where $d(n)$ is the clean output signal [i.e., $d(n)=y(n)-\xi(n)$], and $\hat{d}_\gamma(n)$ is the inverse STFT of the AEC output signal $\hat{d}_{p,k}$ [see (3)], as obtained for a given γ value. Figure 2 shows the resulting mse curves ϵ_0 and ϵ_1 as a function of the SNR, as obtained for nonlinear-to-linear ratios (NLRs) of 10 dB and -10 dB. The NLR represents the power ratio between the output signals of the quadratic and linear components of the true system. The results confirm that for relatively low SNR values, a lower mse is achieved by using the linear MTF model ($\gamma=0$) and allowing for nonlinear undermodeling. For instance, Fig. 2(a) shows that for a -20 dB SNR, employing only a linear model reduces the mse by approximately 18 dB, compared to that achieved by the nonlinear model ($\gamma=1$). However, for high SNR values, the proposed model is considerably more advantageous, as it enables a substantial decrease of 20 dB in the mse for an SNR of 20 dB. A comparison of Figs. 2(a) and (b) indicates that as the NLR decreases, the two curves intersect at a higher NLR value. This implies that when the nonlinearity of the LEM system becomes weaker (i.e., the NLR decreases), higher SNR values should be considered to justify the estimation of the nonlinear component. Moreover, one can observe that the relative improvement achieved by the proposed model at high SNR values becomes larger when increasing the NLR. Specifically for an SNR of 30 dB, the proposed model improves the mse of the linear MTF model by 13 dB for a -10 dB NLR [Fig. 2(b)]; whereas a larger improvement of 21 dB is achieved for a 10 dB NLR [Fig. 2(a)].

In the second experiment, we demonstrate the proposed ap-



(a)



(b)

Fig. 2. MSE curves as a function of the SNR for white Gaussian signals, as obtained by the MTF approach (ϵ_0) and the proposed approach (ϵ_1). (a) Nonlinear-to-linear ratio (NLR) of 10 dB (b) NLR of -10 dB.

proach in a real acoustic echo cancellation scenario using speech signals. We use an ordinary office with a reverberation time T_{60} of about 100 ms. The far-end speech signal is fed into a loudspeaker at high volume (thus introducing non-negligible nonlinear distortion), and received in a microphone, which is located 10 cm away from the loudspeaker. The effective length of the echo path is 100 ms, and the signals are sampled at 16 kHz. In this experiment, we compare the performance of the subband models (both linear and nonlinear) to that of the fullband (second-order) Volterra model, where the parameters of the latter are also estimated off-line. The performance is evaluated in the absence of near-end speech, since in such case a double-talk detector (DTD) is often employed to freeze the estimation process. We use an analysis window length of $N = 1024$ for the linear MTF model in order to validate the large window support assumption. For the proposed model, on the other hand, a smaller length of $N = 256$ is employed in order to maintain a reasonable computational complexity (see Section 4.2). In addition, for the Volterra model, the memory lengths of the linear and quadratic kernels are set to 768 and 60, respectively. Figures 3(a)–(b) show the far-end signal and the microphone signal, respectively. Figures 3(c)–(e) show the residual echo signal $e(n)$ $[= y(n) - \hat{d}(n)]$ obtained by the time-domain Volterra model, the MTF model and the proposed model, respectively. The values of the resulting echo-return loss enhancement (ERLE), defined as $E\{y^2(n)\}/E\{e^2(n)\}$, were also computed, and are given by 18.1 dB (Volterra), 12.6 dB (MTF), and 20.5 dB (proposed). Clearly, the linear MTF model does not provide a sufficient echo attenuation, mainly due to the significant nonlinearity of the echo path. The proposed model, on the other hand, achieves an improvement of 2.4 dB in the ERLE with a lower computational complexity, compared to using the time-domain Volterra model.

6. CONCLUSIONS

Based on the MTF approximation, we have introduced a new nonlinear model for improved acoustic echo cancellation in the STFT

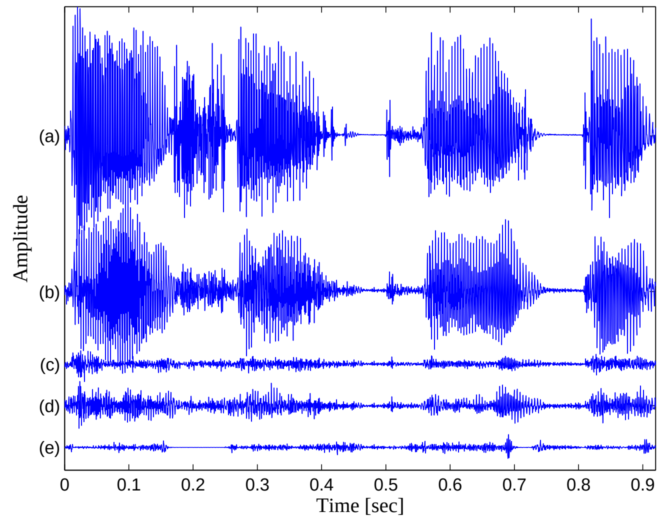


Fig. 3. Temporal waveforms. (a) Far-end signal (b) Microphone signal. (c)–(e) Error signals obtained by a time-domain Volterra model, linear MTF model, and the proposed nonlinear model, respectively.

domain. The proposed model achieves a significant improvement in mse performance over the linear MTF model. Compared to the Volterra approach, the proposed approach provides better estimation accuracy, with a substantially lower computational cost. Future research will concentrate on constructing an adaptive AEC by exploiting the attractive properties of the proposed model.

7. REFERENCES

[1] Y. Lu and J. M. Morris, "Gabor expansion for adaptive echo cancellation," *IEEE Signal Processing Mag.*, vol. 16, pp. 68–80, Mar. 1999.

[2] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Processing Lett.*, vol. 14, no. 5, pp. 337–340, May 2007.

[3] ——, "Adaptive system identification in the short-time Fourier transform domain using cross-multiplicative transfer function approximation," *IEEE Trans. Audio, Speech Lang. Processing*, vol. 16, no. 1, pp. 162–173, Jan. 2008.

[4] C. Faller and J. Chen, "Suppressing acoustic echo in a spectral envelope space," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 13, no. 5, pp. 1048–1062, Sep. 2005.

[5] A. Stenger., L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive Volterra filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Phoenix, USA, Mar. 1999, pp. 877–880.

[6] A. Guérin, G. Faucon, and R. L. Bouquin-Jeannès, "Nonlinear acoustic echo cancellation based on Volterra filters," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 6, pp. 672–683, Nov. 2003.

[7] Y. Avargel and I. Cohen, "Nonlinear systems in the short-time Fourier transform domain–Part I: Representation and identification," *submitted to IEEE Trans. Signal Processing*.

[8] S. Haykin, *Adaptive Filter Theory*, 4th ed. New Jersey: Prentice-Hall, 2002.