NONLINEAR ACOUSTIC ECHO CANCELLATION

A Thesis
Presented to
The Academic Faculty

by

Kun Shi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
December 2008

NONLINEAR ACOUSTIC ECHO CANCELLATION

Approved by:

Professor G. Tong Zhou, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Xiaoli Ma, Co-advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor James Stevenson Kenney
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor David V. Anderson
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor William D. Hunt
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Liang Peng
School of Mathematics
Georgia Institute of Technology

Date Approved: November 6, 2008

To my dear family, friends and my teachers

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation and gratitude to my advisor, Dr. G. Tong Zhou, for offering me the opportunity to pursue my Ph.D. degree at Georgia Tech. She taught me how to strive for excellence in my academic pursuits and helped me to develop critical thinking and presenting skills. What I have learned from her is by no means limited to signal processing techniques, but rather applicable to the entire life.

I would like to thank my co-advisor, Dr. Xiaoli Ma. I appreciate her close guidance and great efforts to prepare me to become a mature researcher. She has always been supportive of me not only when I made progress but also when I made mistakes. I appreciate her guidance, support, endurance, and encouragement during my graduate studies at Georgia Tech.

I would like to thank my dissertation committee members: Dr. James Stevenson Kenney, Dr. David V. Anderson, Dr. Williams D. Hunt, and Dr. Liang Peng for taking the time to serve on my committee and for their helpful suggestions.

My sincere thanks also go to my group members: Hua Qian, Ning Chen, Thao Tran, Vince Emanuele, Bob Baxley, Chunming Zhao, Wei Zhang and Qijia Liu for the insightful discussions and help along the way.

I am also thankful to Dr. Vishu Viswanathan and the fellows in his Speech Technologies Lab: Takahiro Unno, Muhammad Ikram, Ali Erdem Ertan, Jacek Stachurski, and Lorin Netsch, at DSP R&D Center, Texas Instruments. My knowledge was enhanced and my horizon was expanded after each technical conversation with them.

Last but not least, I would like to express my deepest gratitude to my family. I would like to thank my parents and my brother for their unconditional and eternal love and support. Their encouragement is always with me and has given me the strength to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

As voice communication becomes an ever-more important and pervasive part of our everyday lives, the issue of speech quality becomes more critical. One of the reasons for the undesirable quality degradation is the appearance of audible echoes. This kind of quality degradation is inherently from network equipment and end-user devices. To increase speech quality and improve listening experience, it is necessary to design effective acoustic echo cancellation systems.

Echo cancellation has been studied for several decades, and today it is easy to implement echo cancellers on digital signal processors (DSPs). However, certain difficulties still remain to meet the requirements imposed by the echo cancellation standard, and some fundamental challenges still wait for breakthroughs. One of them is the nonlinearity in the acoustic echo path. Nonlinearity usually comes from the price competition in the market of consumer electronics. For economic purposes, the small-sized and low-cost analog components that exhibit nonlinearity, such as loudspeakers and power amplifiers (PAs), are utilized. An echo canceller performs poorly or does not work at all in the system where the net nonlinear distortion is higher than a certain value.

In this dissertation, we address the aforementioned nonlinearity issue in acoustic echo cancellation systems. To sufficiently remove the nonlinear acoustic echo, nonlinear adaptive filters have been proposed in the literature to identify the nonlinear acoustic echo path. The identification is done by minimizing the mean square error (MSE) between the microphone-received signal and estimated echo signal. In this way, the echo signal can be reconstructed and subtracted from the microphone-received signal. However, the issues of stability, convergence rate, and computational complexity inhibit nonlinear acoustic echo cancellers (NAECs) from practical implementation. Thus, we are motivated to design efficient NAECs in terms of stability, fast convergence rate, and low computational complexity. First, we propose to perform nonlinearity identification based on the coherence function,

which guarantees the stability of the nonlinear adaptive system. Later on, we present a general framework for echo cancellation systems using a shortening filter that entails low computational burden and fast convergence rate. Moreover, we develop methods to remove the system nonlinearity based on the coherence function, including the predistortion linearization, nonlinear residual echo suppressor, and Hammerstein-Wiener model-based NAEC.

To design an effective AEC is more than performing an system identification. Another important issue for an AEC is the control logic design of filter adaptation. This problem is caused by the interference at the near-end, including ambient noise and double-talk, when both the far-end and near-end talkers speak at the same time. When double-talk occurs, the adaptive filter may not converge and the identification of the echo path becomes difficult. Double-talk detectors (DTDs) can be utilized to detect the presence of the near-end speech and halt the AEC adaptation, thus to avoid filter divergence. However, DTD designs can be quite complicated since it is often not easy to discriminate between the echo signal and the near-end speech. Moreover, to the best of our knowledge, DTD has not been proposed in conjunction with nonlinear AECs. Unlike double-talk, ambient noise of persistent existence. Therefore, filter adaptation rate needs to be continuously adjusted according to the noise characteristics, rather than being controlled based on carrying out detection. However, few of the learning-rate control algorithms are designed specifically for acoustic echo cancellation applications, which results in the ineffectiveness of these approaches in echo cancellation systems.

In the second part of this dissertation, we focus on the control logic design issue. For double-talk detection, we propose to design a DTD based on the mutual information (MI). We show that the advantage of the MI-based method, when compared with the existing methods, is that it is applicable to both the linear and nonlinear scenarios. Furthermore, we extend the MI-based DTD design to the stereophonic acoustic echo cancellation systems. For learning-rate adjustment, we propose a variable step-size and variable tap-length LMS algorithm. Based on the fact that the room impulse response usually exhibits an exponential decay envelop in acoustic echo cancellation applications, the proposed method finds the

optimal step size and tap length at each iteration. Thus, it achieves faster convergence rate and better steady-state performance.

CHAPTER I

INTRODUCTION

1.1 Motivations

From analog to digital signals, from narrowband to broadband speech, from wireline to wireless terminals, and from circuit-switched to packet-switched networks, there have been tremendous advances in voice telecommunication technologies ever since Alexander Graham Bell invented the telephone in 1876. However, conversation and collaboration using today's voice communication technology are still unnatural and even clumsy. The distraction of holding a superfluous device such as a close-talk microphone and the lack of sensibility of remote speaking environments lead to diminished interaction and productivity, and eventually cause customer dissatisfaction. It is no longer a luxury but truly a rational demand to create a life-like voice communication mode that gives the involved people the impression of being in the same acoustic environment, which is referred to as "immersive experience" in the multimedia communication literature [62]. To achieve this goal, one of the problems that must be addressed is acoustic echo cancellation.

In hands-free telephone systems, Internet phone, and teleconferencing systems, the coupling between the loudspeaker and the microphone on one end of the system causes echoes to occur, which degrades the speech quality for the listener on the other end. For this reason, it is often necessary to implement an acoustic echo canceller (AEC). An AEC greatly enhances speech quality, allows conferences to progress more smoothly and naturally, and prevents listener fatigue. Echo cancellation has been studied for several decades. Most AEC designs seek to remove the echo by reconstructing and subtracting an estimate of the echo signal from the microphone-received signal. This is done by modeling the acoustic echo path using an adaptive filter. The acoustic echo path is tracked by adaptively carrying out system identification. Moreover, the adaptive filter has to work well in the presence of interference, such as ambient noise. Thus, two main design issues for AECs are 1) adaptive

filtering algorithms design, and 2) control logic design for filter adaptation.

The first design issue focuses on filter adaptation. Many adaptive filtering algorithms have been developed to remove echoes while keeping full-duplex communications. One of the well-adopted methods is the least mean square (LMS) algorithm. Several techniques based on the affine projection algorithm (APA) or recursive least squares (RLS) algorithm have been developed to cope with the ill-conditioned input autocorrelation matrix that degrades the LMS performance [41]. However, a recent trend in price-competitive audio consumer products has demanded low-cost and small-sized analog components (such as loudspeakers) that usually exhibit nonlinear characteristics. Research results have shown that linear AECs fail when nonlinearity is present in the acoustic echo path. In [10], it has been shown that the performance of a linear AEC is limited by nonlinear components in the echo path. Also, a statistical study of the LMS algorithm shows that even non-significant saturation could degrade the performance of a linear active noise control system [18]. On the other hand, large reverberation time leads to a long room impulse response. Usually, the finite impulse response (FIR) filter, which models the room impulse response, can occupy several hundred to several thousand taps [41]. This long room impulse response gives rise to slow filter convergence and high computational complexity. The emerging nonlinearity combined with the long room impulse response makes the AEC problem more complicated. Some methods have been proposed in the literature to remove the nonlinear echo [101, 71, 17, 52, 20, 38]. However, there are limitations of the existing methods. First, the stability of nonlinear systems is difficult to be guaranteed. Second, low convergence rate and high computational complexity prevent these methods from being widely used in practical applications. Thus, our research on the efficient nonlinear AEC (NAEC) design is well motivated.

The latter design issue is caused by double-talk or ambient noise. During the double-talk period, since both the near-end speech and the echo signal arrive at the microphone simultaneously, it is difficult to guarantee the convergence of the AEC filter. Consequently, the AEC output consists of both the near-end speech and the uncancelled outgoing echo signal, which is annoying to the far-end listener. One of the well-adopted methods to combat double-talk is to utilize a double-talk detector (DTD), based on which the AEC

filter adaptation is frozen in the presence of the near-end speech [23, 110, 7, 28, 104, 12]. The existing DTDs are developed under the assumption of a linear echo path and do not perform well in the presence of nonlinearity. With respect to the ambient noise, learning-rate adjustment is suggested to achieve optimal convergence rate [11, 56, 3, 66, 57]. The step size is controlled based on the noise characteristics. However, few of the existing methods are specifically designed for acoustic echo cancellation applications and thus some features of echo cancellation are not taken into account. For instance, in acoustic echo cancellation systems, the characteristic of the room impulse response plays an important role in the performance of AECs. Thus, we are motivated to carry out research on control logic designs by taking into account the loudspeaker nonlinearity and room impulse response characteristics.

1.2 Objectives

The objective of this dissertation is to provide a suite of relatively simple but effective solutions to design the nonlinear AEC and its control logic. More specifically, this dissertation focuses on the following topics:

- Nonlinear acoustic echo cancellation using the coherence function
- Double-talk detector (DTD) design using mutual information (MI)
- Step size and tap length control for the LMS algorithm

In the literature, nonlinear acoustic echo cancellers are usually realized by nonlinear adaptive filters. Considering the memoryless nonlinearity from a loudspeaker and/or power amplifier (PA), people use a Hammerstein system to model the nonlinear acoustic echo path and carry out nonlinear system identification by minimizing the mean square error (MSE). However, one issue is that it is difficult to guarantee stability because of a non-quadratic objective function. We address the stability issue by using the coherence function and design an efficient nonlinear AEC in terms of fast convergence rate and low computational complexity. Specifically, we investigate different system structures to remove nonlinear

acoustic echoes, such as predistortion linearization, cascade structure, and post-processing technique.

To be robust to double-talk situations, an AEC employs a DTD to freeze filter adaptation in the presence of near-end speech [29, 7]. Lots of DTD design methods have been proposed in the literature for echo cancellation systems. Among various DTD techniques, the correlation-based method is the most attractive one. However, it does not perform well when the acoustic echo path is nonlinear since the correlation-based criterion captures only the linear relationship between two random processes. In this dissertation, we investigate DTD designs in nonlinear scenarios.

Ambient noise is another interference that may cause the AEC filter to diverge. To be robust to noise, step size is optimized in each filter adaptation to adjust the filter learning rate as a response to noise changes [65, 43]. On the other hand, convergence rate is proved to be governed by the filter tap length. However, to the best of our knowledge, step size and tap length have never been controlled jointly in the literature. Thus, we address a simultaneous control for both step size and tap length to enhance the convergence rate and the steady-state performance.

1.3 Outline

The rest of the dissertation is organized as follows:

In Chapter 2, acoustic echo cancellation system is introduced and some existing algorithms for nonlinear echo cancellation and control logic desgins are reviewed. To remove the nonlinear acoustic echo, we investigate two different system structures: NAEC with cascade nonlinear adaptive filter and nonlinear residual echo suppressor (NRES). Moreover, we discuss the limitations of the existing methods. For control logic design, we first analyze the correlation-based DTD; then we introduce the roles that a step-size control plays in an AEC in the presence of interference; at the end, we point out the deficiencies of the existing design approaches.

In Chapter 3, we investigate different approaches to remove nonlinear acoustic echoes in the system. We focus on the nonlinear acoustic echo cancellation using the coherence

function and investigate different system structures. First, predistortion linearization helps to enhance near-end listener's experience. Then, a cascade structure-based NAEC is proposed to identify the loudspeaker nonlinearity without knowing the room impulse response. Thus, we not only guarantee the system stability but also improve the convergence rate. Moreover, the NAEC with post-processing technique or a shortening filter are proposed to improve the convergence rate. At the end, the Hammerstein-Wiener model-based NAEC is proposed to combat acoustic echoes in the presence of multiple nonlinearities.

In Chapter 4, we focus on control logic designs for echo cancellation systems. First, we propose to design a DTD using MI, which enables the DTD to be applicable in both the linear and nonlinear scenarios. Then, we extend DTD designs into stereophonic systems by using the generalized mutual information (GMI). Compared to MI, the use of GMI not only reduces computational complexity but also facilitates the detection threshold selection. For learning-rate adjustment, we propose a variable step size and variable tap length LMS algorithm for the channel response with an exponential decay envelope.

Finally, in Chapter 5, we summarize this dissertation and suggest topics for future research.

For the reader's convenience, we have attempted to keep every chapter as self contained as possible.

CHAPTER II

BACKGROUND

In this section, we present a literature review to emphasize the necessity of removing the nonlinear acoustic echoes in the system. We start with the traditional AEC structure and the algorithms for implementing it. Next, focusing on the nonlinear acoustic echo path as one of the challenges in the echo cancellation system, we review some AEC approaches for tackling nonlinearity. Finally, we investigate the variable learning-rate adaptive algorithms for robust AEC designs.

2.1 Acoustic Echo Cancellation System

The general setup for acoustic echo cancellation is shown in Fig. 1. The received far-end speech is the output at the near-end loudspeaker, passing through the loudspeaker-enclosure-microphone system (LEMS) to cause the echo signal. The microphone-received signal is composed of the echo signal, near-end speech, noise, and any other distortions. Most AEC designs seek to remove the acoustic echo by reconstructing and subtracting an estimate of the echo signal from the microphone-received signal.

People at the far end of the transmission path are the primary beneficiaries of an AEC. Installed at the near end, an AEC prevents the echo signal from being returned (echoed) through the voice communication system. People speaking at the near end should not be aware of the AEC if it functions properly. Since the person at the far end hears the speech with better quality, the AEC enables the conversation to flow more smoothly and thus benefits both parties. In order for participants at both ends (far and near) to hold a full-duplex hands-free conversation, each end must be equipped with an AEC.

Historically, under the assumption of a completely linear acoustic chain (including a power amplifier, loudspeaker, room impulse response, and microphone), a number of adaptive algorithms based on the gradient theory were developed to remove echoes while keeping full-duplex communication characteristics. Due to its simplicity, the normalized least mean

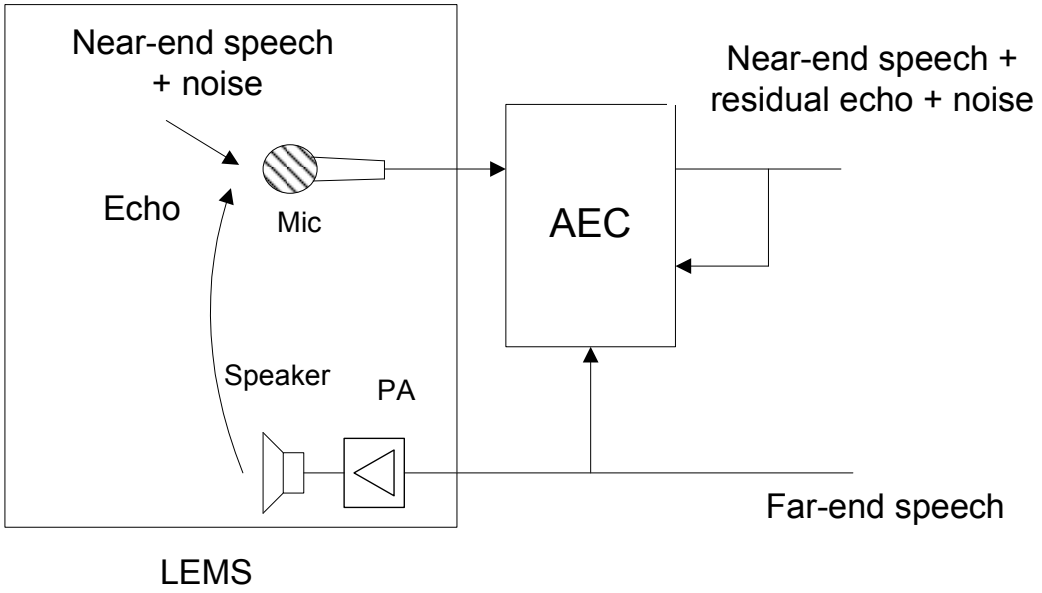


Figure 1: General setup of acoustic echo cancellation.

square (NLMS) algorithm [42] represents a popular approach for the adaptation of AECs. However, the NLMS algorithm suffers from slow convergence for correlated input signals. Therefore, more sophisticated algorithms with decorrelating capability, such as the affine projection algorithm (APA) or the recursive least squares (RLS) algorithm [42], have been proposed to speed up the adaptation of filter coefficients. On the other hand, these approaches increase the computational load remarkably. Consequently, the low-complexity methods that exploit the fast block convolution techniques in the discrete Fourier transform (DFT) domain have been introduced to relieve the computational burden. For example, adaptive DFT-domain algorithms in the so-called constrained and unconstrained versions are presented in [21] and [67], respectively. In these methods, the time-domain linear convolution (used for filtering) and linear correlation (used for adaptation) are efficiently implemented in the frequency domain using the overlap-save algorithm. However, the procedure of data gathering might introduce a long delay. This inherent delay, which is a few hundreds of milliseconds long for typical room acoustic scenarios, is intolerable, as it prevents a natural, full-duplex speech conversation. As a result, a trade-off between the computational complexity and the inherent delay is achieved using the partitioned block

frequency-domain adaptive filter (PBFDAF) algorithm [98]. The PBFDAF splits the time-domain filter into a sequence of non-overlapping partitions. The adaptive filtering is then realized by applying frequency-domain processing to each partition.

2.2 Nonlinear Acoustic Echo Cancellation

A recent trend in consumer electronics is to utilize low-cost and small-sized analog components (such as loudspeakers) for economic considerations. These components usually exhibit nonlinear characteristics, but the hope is to rely on powerful signal processing algorithms to mitigate distortions. The nonlinearities in the LEMS can be roughly divided into two types: nonlinearity with and without memory. Nonlinearity with memory usually occurs in high-quality audio equipment when the time constant of the loudspeaker's electro-mechanical system is large compared to the sampling rate [30]. Memoryless nonlinearity typically occurs in the low-cost power amplifier (PA) or loudspeaker of mobile equipment, where weight constraints call for low supply voltages [101]. With respect to memoryless nonlinearity, the existing methods for nonlinear echo cancellation can be classified into two categories: nonlinear acoustic echo canceller (NAEC)-based and nonlinear residual echo suppressor (NRES)-based methods.

2.2.1 Nonlinear Acoustic Echo Canceller (NAEC)

The general setup of the nonlinear acoustic echo cancellation system is shown in Fig. 2. The received far-end signal $s(n)$ is broadcasted at the near-end loudspeaker, generating the echo signal $c(n)$. The microphone-received signal $y(n)$ is composed of the echo signal $c(n)$ and a signal $v(n)$, representing the background noise and any other signals, such as the near-end speech in a double-talk situation. The goal of an AEC is to subtract the echo signal $c(n)$ from the microphone-received signal $y(n)$. The nonlinear AEC uses a Hammerstein system to model the LEMS. Thus, it consists of a memoryless nonlinear block $u(\cdot;\boldsymbol{\theta})$ and a linear block $h(n)$ corresponding to the PA/loudspeaker nonlinearity and the room impulse response, respectively. The goal is to find an LEMS-equivalent filter to produce $\hat{c}(n)$ such that the energy in the error signal $e(n)$ is minimized. In the following, we introduce a general framework to carry out the nonlinear system identification.

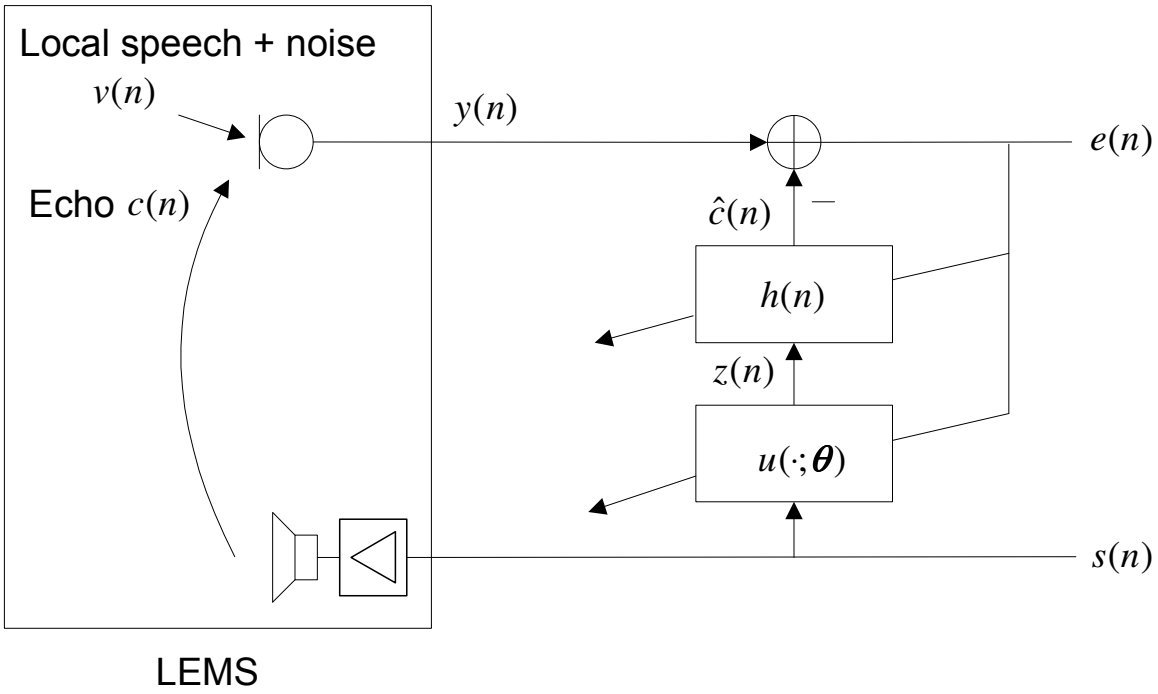


Figure 2: General setup of nonlinear acoustic echo cancellation.

Denote the output of the nonlinear block $u(\cdot;\boldsymbol{\theta})$ by

$$z(n) = u(s(n);\boldsymbol{\theta}(n)). \tag{1}$$

Note that $\boldsymbol{\theta}$ is the parameter of the nonlinear model. Suppose that the AEC filter $h(n)$ has length L_h; we define vectors as

$$\boldsymbol{s}(n) = [s(n), s(n-1), ..., s(n-L_h+1)]^T, \tag{2}$$

$$\boldsymbol{z}(n) = [z(n), z(n-1), ..., z(n-L_h+1)]^T, \text{ and} \tag{3}$$

$$\boldsymbol{h}(n) = [h(n), h(n-1), ..., h(n-L_h+1)]^T. \tag{4}$$

Thus, the estimated echo signal can be expressed as

$$\hat{c}(n) = \boldsymbol{h}^T(n)\boldsymbol{z}(n) = \boldsymbol{h}^T(n)u(\boldsymbol{s}(n);\boldsymbol{\theta}(n)). \tag{5}$$

The estimated error is obtained as

$$e(n) = y(n) - \hat{c}(n) = y(n) - \boldsymbol{h}^T(n)u(\boldsymbol{s}(n);\boldsymbol{\theta}(n)). \tag{6}$$

The LMS-type adaptation for a transversal filter can be derived by forming the gradient of $e^2(n)$ with respect to the transversal filter coefficients. Applying this procedure to the cascaded system described by (6), we obtain the following derivatives:

$$\nabla_h(n) = -2e(n)\boldsymbol{z}(n), \tag{7}$$

$$\nabla_\theta(n) = -2e(n)u'(\boldsymbol{s}(n), \boldsymbol{\theta}(n))^T \boldsymbol{h}(n). \tag{8}$$

If $\boldsymbol{h}(n)$ and $\boldsymbol{\theta}(n)$ are updated with step sizes μ_h and μ_θ, respectively, the LMS-type adaptation algorithm results in

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu_h \boldsymbol{z}(n)e(n), \tag{9}$$

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \mu_\theta u'(\boldsymbol{s}(n), \boldsymbol{\theta}(n))^T \boldsymbol{h}(n)e(n). \tag{10}$$

Based on this framework, a number of algorithms have been proposed in the literature to solve the NAEC problem. Among these approaches, the selection of different nonlinear models to represent the acoustic echo path is widely studied. In [71], a Wiener-Hammerstein system is used to model the acoustic echo path, in which both the hard clipping and soft clipping are suggested to describe the nonlinear characteristic. More general cascade filters and bilinear filters are proposed to compensate for nonlinear echoes in [17]. In [52] and [20], a Hammerstein system is employed to represent the LEMS. An orthogonal polynomial adaptive filter is proposed to accelerate the convergence rate of the nonlinear adaptive filter in [52]. In [20], a nonlinear transform is derived from the raised-cosine function to lower the computational complexity of filter coefficient updates. On the other hand, some studies focus on the mechanism of filter adaptation. For instance, an NLMS-type adaptation algorithm is investigated in [102] that allows simultaneous identification of a polynomial nonlinearity and a linear finite impulse response (FIR) system. In [100], an RLS-type adaptation is derived to speed up the convergence of the polynomial. Moreover, some methods propose more efficient AEC designs in terms of both nonlinear models and filter adaptation schemes, for example, the Volterra model-based [38, 59] and neural network structure-based [83] approaches.

2.2.2 Nonlinear Residual Echo Suppressor (NRES)

As discussed in Section 2.2.1, realizing the AEC as a nonlinear adaptive filter can improve the echo attenuation performance in the presence of nonlinear echo paths. Unfortunately, the convergence rate of the nonlinear adaptive filter achieved by the existing approaches is slow. One way to overcome this drawback is to apply the residual echo suppression technique to further reduce the residual echo that remains after a purely linear AEC. This post-filtering technique for removing the nonlinear residual echo has been studied in [45, 61]. The nonlinear acoustic echo cancellation using an NRES is shown in Fig. 3.

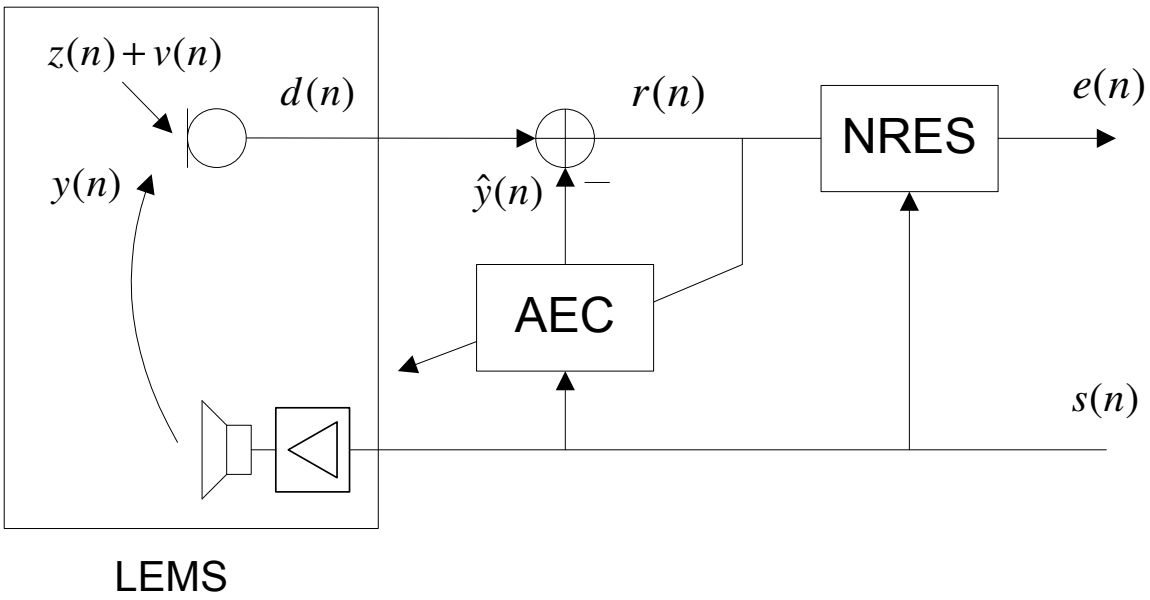


Figure 3: Nonlinear acoustic echo cancellation with an NRES.

Let $s(n)$ denote the far-end signal and $d(n)$ denote the microphone-received signal, which consists of the near-end speech $z(n)$, the background noise $v(n)$, and the acoustic echo $y(n)$. The adaptive AEC tries to identify the LEMS and produce an estimate of the echo signal denoted by $\hat{y}(n)$. The estimated echo is then subtracted from the microphone-received signal to produce the residual signal $r(n)$:

$$r(n) = d(n) - \hat{y}(n) = z(n) + v(n) + y(n) - \hat{y}(n). \tag{11}$$

If the acoustic echo path exhibits nonlinear characteristics, a linear AEC can not completely remove the acoustic echo. We define the nonlinear residual echo $p(n)$ as the difference

between the true echo signal $y(n)$ and its estimate $\hat{y}(n)$:

$$p(n) = y(n) - \hat{y}(n). \tag{12}$$

Similar to the post filter commonly used in the linear echo case, the NRES is a frequency-dependent, real-valued gain filter $C(f)$, realized by the frequency-domain processing on a frame-by-frame basis [39]. Accordingly, for each frame, the NRES output $e(n)$ and the residual signal $r(n)$ are related in the frequency domain as

$$E^{(m)}(f) = C^{(m)}(f)R^{(m)}(f), \tag{13}$$

where m is the frame index; $E^{(m)}(f)$ and $R^{(m)}(f)$ are the DFTs of the m^{th} frame of $e(n)$ and $r(n)$, respectively, at the discrete frequency f. The resulting $E^{(m)}(f)$ is transformed back into the time domain by the inverse DFT (IDFT), and the output signal $e(n)$ is then synthesized with the overlap-save method. One way to design the gain function $C(f)$ is described next. For notational simplicity, we omit the frame index m when feasible from this point on.

The optimal gain $C(f)$ can be derived by minimizing the contribution of the nonlinear residual echo $R(f)$ to the output signal $E(f)$ in the mean square error (MSE) sense. Based on the results obtained from [61, 39], the optimal $C(f)$ is

$$C(f) = \frac{S_r(f) - S_p(f)}{S_r(f)}, \tag{14}$$

where $S_r(f)$ and $S_p(f)$ denote the power spectral density (PSD) functions of $r(n)$ and $p(n)$, respectively. Here, we focus on the suppression of the nonlinear residual echo without attenuating the background noise. If noise reduction is considered, the gain function in (14) can be rewritten as

$$C(f) = \frac{S_r(f) - S_p(f) - S_v(f)}{S_r(f)}, \tag{15}$$

where $S_v(f)$ is the PSD of the background noise $v(n)$. Since we assume $v(n)$ to be a white noise, (14) can be used in place of (15). In (14), $S_r(f)$ can be estimated easily by recursively smoothing $\left|R^{(m)}(f)\right|^2$ as in

$$\hat{S}_r^{(m)}(f) = \lambda \hat{S}_r^{(m-1)}(f) + (1-\lambda)\left|R^{(m)}(f)\right|^2, \tag{16}$$

where $0 < \lambda < 1$ is the forgetting factor. Therefore, calculating of the optimal gain $C(f)$ is reduced to estimating the PSD of $p(n)$. In [61], based on the power filter model of the acoustic echo path, an additional adaptive filter, referred to as the residual echo filter, is used to estimate the nonlinear residual echo, following which $S_p(f)$ is calculated.

2.2.3 Limitations

Although a number of methods have been proposed in the literature to combat the nonlinear acoustic echo, there is still room to explore both NAEC- and NRES-based methods due to the limitations of the existing methods.

For the category of NAEC-based methods, convergence and complexity are the two most important issues. Volterra filter-based methods make use of the linear relationship between the error signal and filter coefficients to guarantee convergence [38, 59], but an adaptive Volterra filter requires high computational complexity [38]. Cascade structures have been proposed to reduce the complexity of the Volterra-based method, but it is hard to assure the convergence to the optimal solution or even guarantee a stable adaptation behavior because of the non-quadratic surface of the objective function [101, 71, 17, 52, 20, 60]. For instance, a smaller step size is used for an adaptive nonlinear filter in a Hammerstein system to ensure convergence in [101]. It has also been recommended not to adapt the nonlinear filter until the linear one has "sufficiently" converged. A strategy of adapting the coefficients of a linear post filter before the nonlinear one in a Wiener-Hammerstein system is employed as a remedy for the convergence issue in [71]. In [60], an adaptive orthogonalized power filter is proposed to improve the convergence rate. The orthogonal basis is updated online in each iteration, and the Gram Schmidt procedure is employed to find the orthogonalized coefficients. As a result, computational complexity is increased. Therefore, it is challenging to achieve satisfactory performance in terms of both convergence rate and complexity.

For the category of NRES-based methods, the post-processing scheme is first proposed in the context of linear echoes to combine the acoustic echo control and noise reduction [39, 70]. However, all these methods require a linear echo path and thus are not applicable when nonlinear distortions are present. Recently, the post-processing technique has been

applied to the nonlinear cases, but some design challenges still remain. The NRES approach proposed in [45] requires a frequency-domain model of the nonlinear residual echo that must be determined in advance. Since this model depends on the hardware components in the echo path, it must be acquired for each hardware setup separately. Similar to the linear case in [70], the NRES in [61] includes a residual echo filter to estimate the nonlinear residual echo. However, the convergence rate and computational complexity of the NRES filter depend on the length of the auxiliary filter. Usually, a desirable length of the auxiliary filter is the length of the room impulse response. Thus, the insufficient knowledge of the room impulse response degrades the residual echo suppression performance.

2.3 Control Logic for the Robust AEC Design

The main objective in an AEC design is to identify the unknown acoustic echo path and hence to subtract an estimate of the acoustic echo from the microphone-received signal. However, when the far-end and near-end talkers speak at the same time, the near-end speech acts as an uncorrelated noise to the adaptive filter and causes the filter to diverge, which results in an annoying audible echo to pass through to the far end. Robust echo cancellation requires a control logic for filter adaptations to account for the interference in the microphone-received signal. The general structure of an AEC with a controller is shown in Fig. 4.

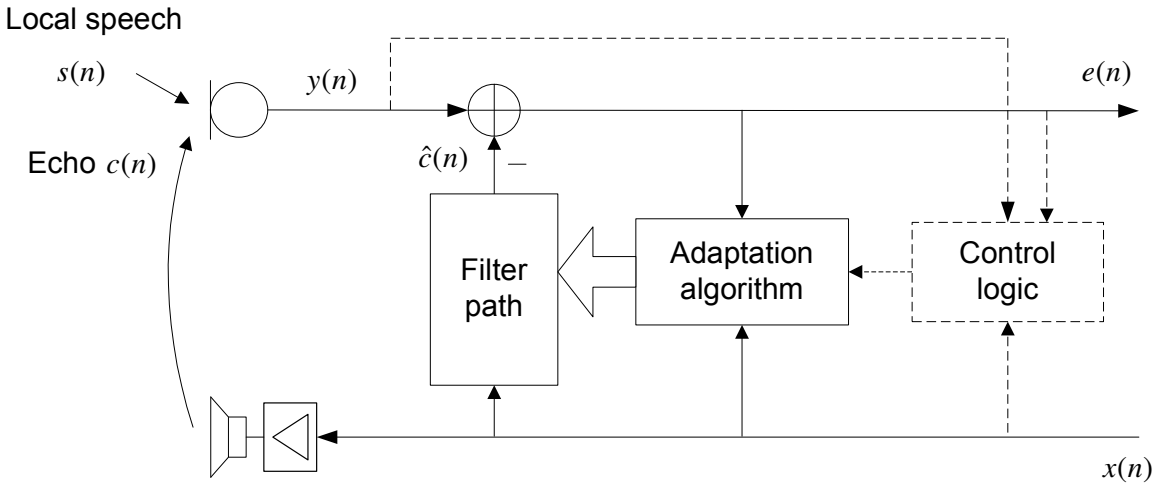


Figure 4: General structure of an AEC with a controller.

The far-end speech $x(n)$ is the output at the near-end loudspeaker, causing the echo signal $c(n)$. The microphone-received signal $y(n)$ is composed of the echo signal $c(n)$ and near-end speech $s(n)$. The AEC employs an adaptive filter to model the acoustic echo path and perform the echo cancellation. The controller adjusts the learning rate of the AEC filter based on the interference (noise and/or near-end speech). The controller can be designed in two different ways: (1) to detect the presence of the near-end speech using a double-talk detector (DTD) and then lock the filter adaptation; (2) to adjust the learning rate continuously without detecting double-talk occurrences.

2.3.1 Double-Talk Detection

Most echo controllers attempt to detect double-talk occurrences and then react by freezing the adaptation of the adaptive filter. A DTD employs available signals or estimates to make the decision of whether or not near-end speech $s(n)$ is present. The DTD decision is then utilized to design the control logic for the AEC filter. In general, double-talk detection is handled in the following way:

1) A detection statistic ξ is formed using available signals, e.g., x, y, e, etc., and the estimated filter coefficients.

2) The detection statistic ξ is compared to a preset threshold T and the double-talk is declared if $\xi > T$.

3) Once a double-talk is declared, the filter adaptation is disabled.

4) If $\xi \leq T$, the comparison of ξ to T and filter adaptation continue.

The Geigel algorithm [23] has been proven effective for line echo cancelers. However, it does not provide reliable performance when applied to AECs. Cross-correlation-based DTD techniques [110, 7] have been proposed, that appear to be suitable for AEC applications. DTDs have also been developed for subband [51] and stereo [54] AEC applications.

Here, the cross-correlation methods in [110, 7] are briefly described as follows.

2.3.1.1 Cross-Correlation Method

The cross-correlation vectors between $x(n)$ and $y(n)$ and between $x(n)$ and $e(n)$ are defined as

$$\boldsymbol{c}_{xy} = \left[c_{xy,0}, c_{xy,1}, ..., c_{xy,L_h-1}\right]^T, \tag{17}$$

$$\boldsymbol{c}_{xe} = \left[c_{xe,0}, c_{xe,1}, ..., c_{xe,L_h-1}\right]^T, \tag{18}$$

where L_h is the length of the AEC filter and

$$c_{xy,i} = \frac{E[x(k-i)y(k)]}{\sqrt{E\left[x^2(k-i)\right]E\left[y^2(k)\right]}}, \tag{19}$$

$$c_{xe,i} = \frac{E[x(k-i)e(k)]}{\sqrt{E\left[x^2(k-i)\right]E\left[e^2(k)\right]}}. \tag{20}$$

The detection statistic ξ can be formed by taking the norm of the cross-correlation vectors [110]. Any scalar metric, such as l_1, l_2, or l_∞ norm, is feasible when determining the norm. For example, the l_∞-based decision statistic results in

$$\xi_{xy} = \left[\max_i \left|c_{xy,i}\right|\right], \tag{21}$$

and

$$\xi_{xe} = \left[\max_i \left|c_{xe,i}\right|\right]. \tag{22}$$

Note that the threshold T is desired to be independent of $x(n)$, $y(n)$, and $e(n)$. Thus, the method in [7] applies a normalization technique in the sense that the detection statistic is equal to one when the near-end signal is absent.

2.3.1.2 Normalized Cross-Correlation Method

The normalized cross-correlation vector is defined as

$$\boldsymbol{c}_{xy} = \left(\sigma_y^2 \mathbf{R}_x\right)^{-1/2} \boldsymbol{r}_{xy}, \tag{23}$$

where σ_y^2 is the variance of $y(n)$, $\mathbf{R}_x$ is the autocorrelation matrix of $\boldsymbol{x}(n)$, and $\boldsymbol{r}_{xy}$ is the cross-correlation between $\boldsymbol{x}(n)$ and $y(n)$. The corresponding detection statistic is obtained by taking the l_2 norm of the normalized cross-correlation vector:

$$\xi_{xy} = \|\boldsymbol{c}_{xy}\|_2. \tag{24}$$

It is shown in [7] that the decision statistic can be expressed as

$$\xi_{xy} = \frac{\sqrt{\boldsymbol{h}^T \mathbf{R}_{xx} \boldsymbol{h}}}{\sqrt{\boldsymbol{h}^T \mathbf{R}_{xx} \boldsymbol{h} + \sigma_{s^2}}}, \tag{25}$$

where σ_s^2 denotes the near-end speech power. It is easily seen that $\xi_{xy} = 1$ when $s(n) = 0$ and $\xi_{xy} < 1$ when $s(n) \neq 0$.

It is known that the existing methods are based on the linear relationship between the far-end signal $x(n)$ and the microphone-received signal $y(n)$:

$$y(n) = \boldsymbol{x}^T(n)\boldsymbol{h}(n) + s(n). \tag{26}$$

However, when nonlinearity is present in the acoustic echo path, e.g., the loudspeaker exhibits nonlinear characteristics, $x(n)$ and $y(n)$ are no longer linearly related:

$$y(n) = g(\boldsymbol{x}(n))^T \boldsymbol{h}(n) + s(n), \tag{27}$$

where $g(\cdot)$ denotes the nonlinearity in the loudspeaker. Thus, the existing DTD algorithms fail to perform well. To the best of our knowledge, a DTD has not been proposed in conjunction with nonlinear AECs. Although [9] derives an optimum log-likelihood ratio test (LRT), the Gaussian assumption no longer holds when nonlinearity is present.

2.3.1.3 Statistical Analysis

By viewing the DTD design as a binary detection problem, the DTD performance can be evaluated using detection theory concepts that were developed for radar and communication applications [16, 106]. Formulating a binary hypothesis test for a DTD as

- H_0: double-talk is absent ($\xi \geq T$), and
- H_1: double-talk is present ($\xi < T$),

we review the general characteristics of a binary detection scheme:

1. Probability of False Alarm (P_{FA}): The probability of declaring detection when near-end speech is absent:

$$P_{FA} = P[\text{accepting } H_1 | H_0 \text{ is true}] = \int_{-\infty}^{T} f(\xi|H_0) \mathrm{d}\xi. \tag{28}$$

2. Probability of Detection (P_D): The probability of successful detection when near-end speech is present:

$$P_D = P[\text{accepting } H_1 | H_1 \text{ is true}] = \int_{-\infty}^{T} f(\xi|H_1)\mathrm{d}\xi. \tag{29}$$

A "good" detection method should maximize P_D while minimizing P_{FA}. In general, higher P_D is achieved at the cost of higher P_{FA}. To quantify the relationship between P_D and P_{FA}, receiver operating characteristic (ROC) curves are widely used in radar and communication applications. We use a similar approach to evaluate the performance of a DTD.

2.3.2 Learning-Rate Adjustment

In a normal telephone conversation, double-talk occurs approximately 20% of the time [96]. In some AEC scenarios, the background noise is continuously present and the use of a DTD becomes futile because the AEC filter may diverge considerably before a double-talk period is detected. This may be the case, for example, in a noisy teleconferencing application, in an automatic gain adjustment system equipped with an echo canceller, or in an adaptive feedback cancellation (AFC) system [99, 80]. These applications have provided the motivation to improve the robustness of the adaptive algorithm to compensate for the detection lag as well as other DTD imperfections. A variable learning rate has been used without the detection of double-talk occurrences. In [22, 47], a so-called maximum-length correlation estimate replaces the stochastic gradient room impulse response (RIR) estimate whenever a double-talk situation occurs. In [73] and [64], an adaptive cross-spectral technique is employed instead of the standard adaptive algorithm, and it is shown to be robust in double-talk situations. In [107, 108], double-talk robustness is established by taking into account the characteristics of the near-end signal.

Even though many adaptive algorithms are theoretically applicable for AEC designs, in the applications with limited precision and processing power, the least mean square (LMS) algorithm [42] and its modifications (e.g., the normalized LMS, frequency-domain LMS, and subband LMS [6]) are usually used. The performance of the LMS algorithm, in terms of convergence rate, misadjustment, and stability, is governed by the step size. With the stability condition, the selection of the step size reflects a trade-off between fast

convergence rate and good tracking ability on the one hand and low misadjustment on the other hand. To meet these conflicting requirements, the step size needs to be controlled. Thus, a number of variable step size (VSS) NLMS algorithms have been proposed [3, 66, 95, 8, 75]. Although these algorithms can be applied in the context of AEC designs, they consider only the background noise as the interference and are not specifically designed for double-talk situations. Under the assumption of a stationary interference (noise and/or near-end speech), the long-term statistic of filter misadjustment is used to determine step size. In [105], a frequency-domain echo canceller with a variable learning rate is proposed. The optimal learning rate is adjusted as a function of both the interference and the filter misadjustment. This frequency-domain method is briefly described as follows.

The NLMS filter of length L_h is defined as

$$e(n) = y(n) - \hat{\boldsymbol{h}}^T(n)\boldsymbol{x}(n), \tag{30}$$

with the adaptation of filter coefficients

$$\hat{\boldsymbol{h}}(n+1) = \hat{\boldsymbol{h}}(n) + \mu\frac{e(n)\boldsymbol{x}(n)}{\|\boldsymbol{x}(n)\|^2}. \tag{31}$$

Considering the filter misadjustment $\boldsymbol{\delta}(n) = \|\hat{\boldsymbol{h}}(n) - \boldsymbol{h}\|_2^2$, and knowing $y(n) = \boldsymbol{h}^T(n)\boldsymbol{x}(n) + v(n)$, where $v(n)$ denotes the near-end interference, we obtain the expected misadjustment under the assumption that $x(n)$ and $v(n)$ are white signals [105]:

$$E\left[\boldsymbol{\delta}(n+1)|\boldsymbol{\delta}(n), x(n)\right] = \boldsymbol{\delta}(n)\left[1 - \frac{2\mu}{L_h} + \frac{\mu^2}{L_h} + \frac{\mu^2\sigma_v^2}{\boldsymbol{\delta}(n)\|\boldsymbol{x}(n)\|^2}\right], \tag{32}$$

where $E[\cdot]$ denotes the statistical expectation, and σ_v^2 denotes the variance of the signal $v(n)$. Because (32) is a convex function, the expected misadjustment can be minimized with respect to the step size μ by solving $\partial E[\boldsymbol{\delta}(n+1)]/\partial\mu = 0$. This leads to the optimal learning rate

$$\mu_{opt}(n) = \frac{1}{1 + \frac{\sigma_v^2}{\boldsymbol{\delta}(n)\|\boldsymbol{x}(n)\|^2/L_h}}. \tag{33}$$

When there is no near-end interference (σ_v^2), it can be seen that (33) simplifies to $\mu_{opt}(n) = 1$, which is consistent with [46]. However, the white-noise assumption does not hold in acoustic echo cancellation applications. It has been observed that the input signal across consecutive

fast Fourier transform is less uncorrelated than the original time-domain signal. Thus, the optimal step size in (33) is applied to adaptive filter algorithms that operate in the frequency domain [105].

CHAPTER III

NONLINEAR ACOUSTIC ECHO CANCELLATION BASED ON THE COHERENCE FUNCTION

In this chapter, we present several methods to remove the nonlinear acoustic echo in echo cancellation systems. We consider the memoryless nonlinearity, which comes from the loudspeaker and/or PA. Since the LEMS in Fig. 1 consists of a nonlinear PA/loudspeaker followed by a linear subsystem (the room impulse response), the LEMS can be well represented by the Hammerstein model [60, 17]. We adopt nonlinear basis expansion form for nonlinearity modeling and focus on nonlinear acoustic echo cancellation. Specifically, we investigate three different structures: predistortion linearization (Section 3.2), cascade structure (Section 3.3), and post processing (Section 3.4). We apply the coherence function into these structures and discuss the advantages of them. In addition, we consider the issue of computational complexity (Section 3.5) and investigate the echo cancellation in the presence of multiple nonlinearities (Section 3.6).

3.1 Coherence Function and Its Properties

Let $y(n)$ and $z(n)$ be real-valued discrete-time random processes, $n = 0, 1, \ldots, N-1$. Define the discrete-time Fourier transform (DTFT) of $y(n)$ as

$$Y(f) = \sum_{n=0}^{N-1} y(n) e^{-j2\pi fn}, \tag{34}$$

where $-0.5 \leq f \leq 0.5$ is the normalized frequency. The cross-correlation function between $y(n)$ and $z(n)$ at delay m is

$$R_{yz}(m) = E\left[y(n)z(n+m)\right], \tag{35}$$

where $E[\cdot]$ denotes the statistical expectation. The cross-spectral density function between $y(n)$ and $z(n)$ is the DTFT of $R_{yz}(m)$:

$$S_{yz}(f) = \sum_{m=0}^{N-1} R_{yz}(m) e^{-j2\pi fm}. \tag{36}$$

3.1.1 Magnitude Squared Coherence (MSC) Function

Define the (magnitude squared) coherence function (MSC) between $y(n)$ and $z(n)$ at frequency f as [13]

$$C_{yz}(f) = \frac{|S_{yz}(f)|^2}{S_{yy}(f)S_{zz}(f)}, \tag{37}$$

where $S_{yy}(f)$ is the power spectral density (PSD) function of $y(n)$ at frequency f, and similarly for $S_{zz}(f)$. The coherence function has been well studied and applied to many interesting problems, such as system analysis [13, 53], signal-to-noise ratio measurement and noise reduction [63, 24], and time delay estimation [14]. In [26, 72], the coherence function has been used in the blind source separation problem.

It can be shown that [13] $0 \leq C_{yz}(f) \leq 1,\ \forall f$, and that $C_{yz}(f) = 1,\ \forall f$, if and only if $y(n) = a(n) * z(n) + b(n)$, where $*$ denotes the linear convolution. Here, $a(n)$ and $b(n)$ are deterministic quantities; $a(n)$ can be regarded as the impulse response of a linear time-invariant (LTI) system linking $y(n)$ to $z(n)$, and $b(n)$ can be considered as a modeling error or other deterministic error. Thus, the coherence function can be viewed as a measure of the linear relationship between two random processes.

3.1.2 Pseudo-MSC Function and Its Properties

Define the pseudo-MSC function between $y(n)$ and $z(n)$ at frequency f as [27]

$$\tilde{C}_{yz}(f) = \frac{|S_{yz}(f)|^2}{S_{yy}(f)\sigma_z^2}, \tag{38}$$

where $\sigma_z^2 = E\left[z^2(n)\right]$ is the power of $z(n)$. It is clear that $\tilde{C}_{yz}(f) \geq 0,\ \forall f$. The major difference between the pseudo-MSC function in (38) and the MSC function in (37) is in the normalizer, where the power of the signal $z(n)$ (σ_z^2) is used instead of its PSD $S_{zz}(f)$.

The following properties hold for the pseudo-MSC function [90, 92].

Property I: $0 \leq \int_{-0.5}^{0.5} \tilde{C}_{yz}(f)\mathrm{d}f \leq 1$; $\int_{-0.5}^{0.5} \tilde{C}_{yz}(f)\mathrm{d}f = 1$ if and only if $y(n)$ and $z(n)$ are linearly related, i.e., $y(n) = a(n) * z(n) + b(n)$, where $a(n)$ and $b(n)$ are deterministic processes.

Proof: Since $C_{yz}(f) \geq 0,\ \forall f$, it is obvious that

$$\int_{-0.5}^{0.5} \tilde{C}_{yz}(f)\mathrm{d}f \geq 0. \tag{39}$$

Denote by $Y(f)$ and $Z(f)$ the DTFTs of $y(n)$ and $z(n)$, respectively. The covariance between $Y(f)$ and $Z(f)$ can be shown as

$$Cov[Y(f), Z(f)] = NS_{yz}(f).$$

It follows easily that

$$Cov[Y(f), Y(f)] = NS_{yy}(f), \tag{40}$$

$$Cov[Z(f), Z(f)] = NS_{zz}(f). \tag{41}$$

Thus, the pseudo-MSC function between $y(n)$ and $z(n)$ can be rewritten as

$$\begin{aligned}\tilde{C}_{yz}(f) &= \frac{|S_{yz}(f)|^2}{S_{yy}(f)S_{zz}(f)} \cdot \frac{S_{zz}(f)}{\sigma_z^2} \\ &= \frac{|Cov[Y(f), Z(f)]|^2}{Cov[Y(f), Y(f)]Cov[Z(f), Z(f)]} \cdot \frac{S_{zz}(f)}{\sigma_z^2}.\end{aligned} \tag{42}$$

Recalling the Cauchy-Schwartz inequality [90], we infer that

$$\int_{-0.5}^{0.5} \tilde{C}_{yz}(f)\mathrm{d}f \leq \int_{-0.5}^{0.5} \frac{S_{zz}(f)}{\sigma_z^2}\mathrm{d}f = 1, \tag{43}$$

with equality if and only if

$$|Cov[Y(f), Z(f)]|^2 = Cov[Y(f), Y(f)]Cov[Z(f), Z(f)], \tag{44}$$

i.e., when

$$Y(f) = A(f)Z(f) + B(f), \tag{45}$$

where $A(f)$ and $B(f)$ are deterministic constants at frequency f. Transforming (45) into the time domain, we obtain

$$y(n) = a(n) * z(n) + b(n), \tag{46}$$

where $a(n)$ and $b(n)$ are deterministic processes. $a(n)$ can be regarded as the impulse response of an LTI system linking $z(n)$ to $y(n)$, and $b(n)$ can be considered as a modeling error or other deterministic error.

Property II: The metric $\int_{-0.5}^{0.5} \tilde{C}_{yz}(f)\mathrm{d}f$ provides a means for quantifying the linear association between two stationary random processes. This is equivalent to using the normalized

linear minimum mean square error (LMMSE) criterion to quantify the degree of the linear association between two stationary random processes.

Proof: Consider two stationary random processes $y(n)$ and $z(n)$. To measure how closely these two random processes are linearly related, $y(n)$ is designated as the excitation of a linear operator $h(n)$ and $z(n)$ is designated as the target response. The linear operator's output $\hat{z}(n)$ is then specified by

$$\hat{z}(n) = \sum_{\tau} h(\tau) y(n-\tau) \mathrm{d}\tau. \tag{47}$$

To quantify the performance of $h(n)$, the modeling error is introduced as $e(n) = z(n) - \hat{z}(n)$, and the mean square error (MSE) $E\left[e^2(n)\right]$ is to be minimized. It is well known that the LMMSE solution for (47) is

$$H_o(f) = \frac{S_{zy}(f)}{S_{yy}(f)}, \tag{48}$$

where the subscript $_o$ denotes the optimal solution of $h(n)$ in the LMMSE sense. The degree of the linear association between the processes $y(n)$ and $z(n)$ is revealed through the behavior of LMMSE by

$$\begin{aligned} E\left[e_o^2(n)\right] &= \int_{-0.5}^{0.5} \left[S_{zz}(f) - S_{yy}(f)\left|H_o(f)\right|^2\right] \mathrm{d}f \\ &= \sigma_z^2 \left[1 - \int_{-0.5}^{0.5} \tilde{C}_{yz}(f) \mathrm{d}f\right]. \end{aligned} \tag{49}$$

To determine the level of linear association between $y(n)$ and $z(n)$, a more convenient scalar measure is given by the normalized LMMSE:

$$\rho_{yz} = \frac{E\left[e_o^2(n)\right]}{E\left[z^2(n)\right]}. \tag{50}$$

Substitution of (49) into (50) yields

$$\rho_{yz} = 1 - \int_{-0.5}^{0.5} \tilde{C}_{yz}(f) \mathrm{d}f. \tag{51}$$

Therefore, a high degree of the linear association between $y(n)$ and $z(n)$ is revealed when $\int_{-0.5}^{0.5} \tilde{C}_{yz}(f) \mathrm{d}f$ is close to one (or ρ_{yz} is close to zero). On the other hand, little linear association is indicated when $\int_{-0.5}^{0.5} \tilde{C}_{yz}(f) \mathrm{d}f$ is close to zero (or ρ_{yz} is close to one).

3.2 NAEC with Predistortion Linearization

Considering the nonlinearity due to the PA and/or loudspeaker, we propose to linearize the LEMS using a predistorter. Afterwards, we apply the linear AEC as usual (see Fig. 5). Precompensation (in the electric domain) of the PA/loudspeaker nonlinearity is preferable to postcompensation (in the acoustic domain), because the former is easier to implement digitally. Many nonlinear system identification and compensation methods are available in the literature [15, 40, 25, 33, 48, 35]. We propose the MSC function-based criterion to compensate for the nonlinear distortions in a Hammerstein system. We show that the advantage of the coherence function-based method as compared with existing methods is that it is "blind" to the presence of an unknown linear block (e.g., the long room impulse response between the loudspeaker and the microphone) [93, 94].

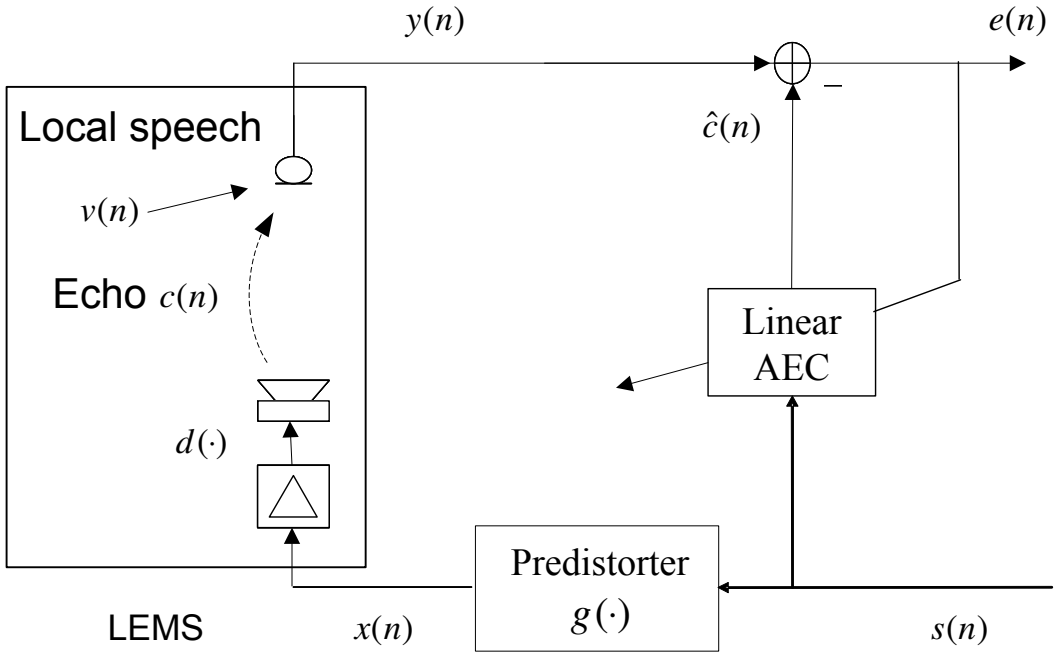


Figure 5: Predistortion architecture for the nonlinear AEC.

3.2.1 Predistorter Design

Consider the Hammerstein system which consists of a memoryless nonlinear mapping $d(\cdot)$ followed by an LTI block with the impulse response $h(n)$. Denote by $g(\cdot)$ a memoryless predistorter function block inserted just before the Hammerstein system to compensate for

the nonlinearity in $d(\cdot)$. We seek a parametric approach for $g(\cdot)$ and denote its parameter vector by $\boldsymbol{\phi}$. The block diagram for this method is shown in Fig. 6. In Fig. 6, $s(n)$ is the system input, $x(n)$ is the output of the predistorter $g(\cdot)$, and $y(n)$ is the output of the Hammerstein system. Let $g(\cdot)$ be a linear combination of the basis functions $b_k(s)$:

$$g(s;\boldsymbol{\phi}) = \sum_{k=1}^{K} \phi_k b_k(s), \quad \boldsymbol{\phi} = [\phi_1, \phi_2, ..., \phi_K]^T, \tag{52}$$

where ^{T} denotes transpose. Correspondingly,

$$x(n;\boldsymbol{\phi}) = g(s(n);\boldsymbol{\phi}), \tag{53}$$

$$y(n;\boldsymbol{\phi}) = d(x(n;\boldsymbol{\phi})) * h(n). \tag{54}$$

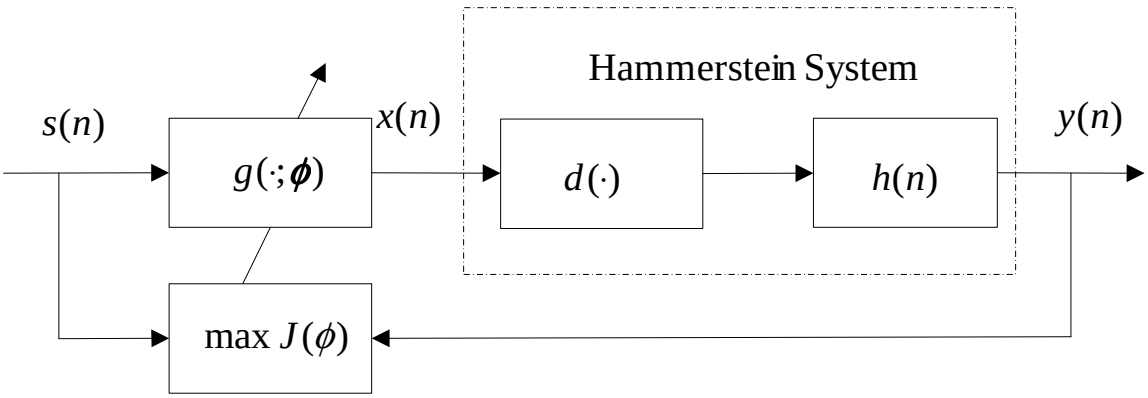


Figure 6: Finding the predistorter $g(\cdot;\boldsymbol{\phi})$ using a MSC-based criterion.

We propose the following criterion to estimate the predistorter parameter vector $\boldsymbol{\phi}$ [94]:

$$\hat{\boldsymbol{\phi}} = \arg\max_{\boldsymbol{\phi}} J_1(\boldsymbol{\phi}), \tag{55}$$

where

$$J_1(\boldsymbol{\phi}) = \int_0^{0.5} \hat{C}_{sy}(f;\boldsymbol{\phi}) df, \tag{56}$$

and $\hat{C}_{sy}(f;\boldsymbol{\phi})$ is the estimated MSC function between $s(n)$ and $y(n)$. The MSC function is estimated according to (37), where the auto- and cross-spectral densities can be estimated using the Welch method with the fast Fourier transform (FFT) [14]. Thus, the MSC function is close to one only at the discrete set of frequencies used by the FFT. Finer frequency resolution in the FFT can be achieved by zero-padding.

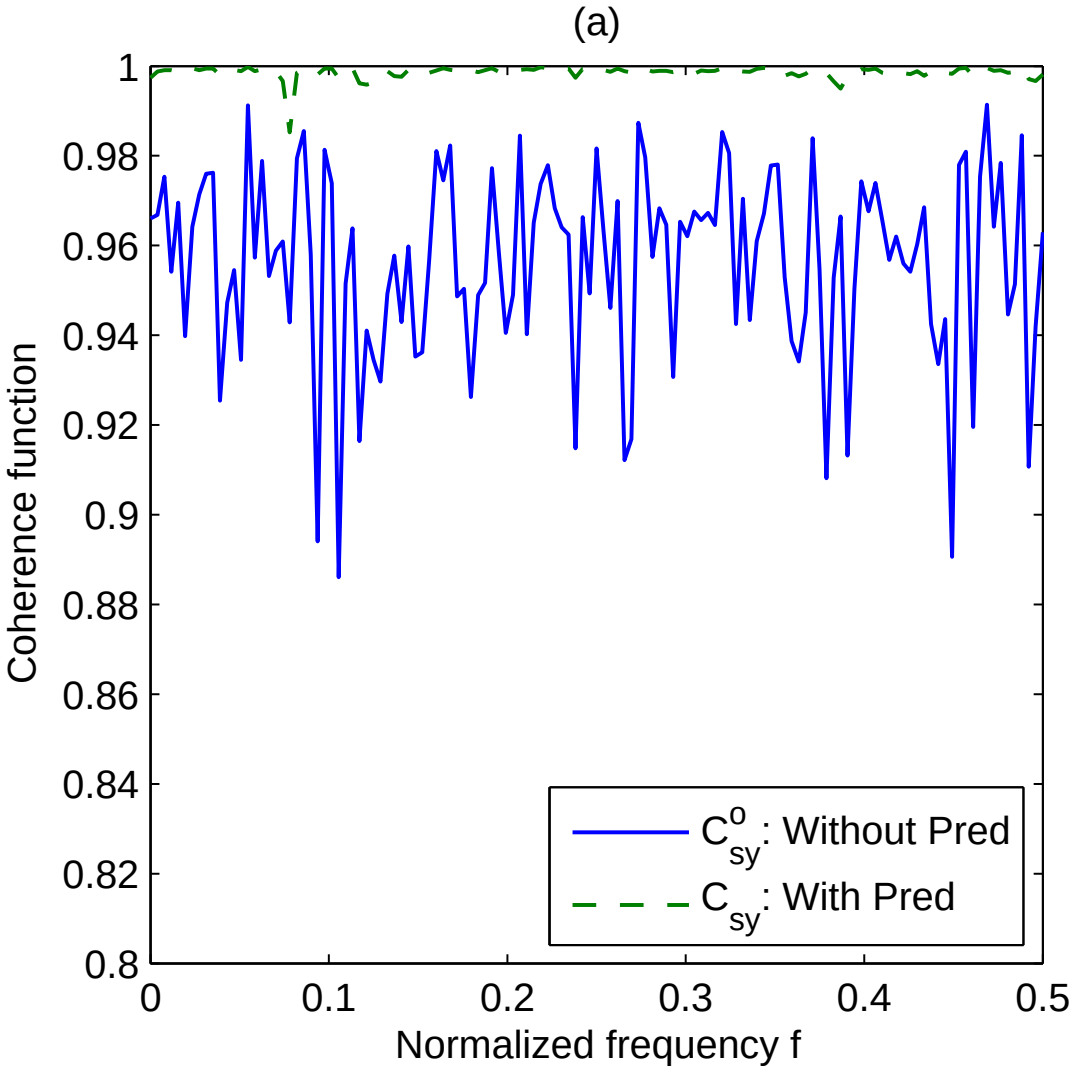
Denote by $s^{(i)}(t)$ and $y^{(i)}(n)$ the i^{th} block of the system input and output, where $y^{(i)}(n)$ is generated according to $y^{(i)}(n) = d(g(s^{(i)}(n); \boldsymbol{\phi}^{(i-1)})) * h(n)$ and the initial $\boldsymbol{\phi}^{(0)}$ is generated such that $x^{(0)}(n) = s^{(0)}(n)$. We adapt the $\boldsymbol{\phi}^{(i)}$ estimate to increase $J_1(\boldsymbol{\phi}^{(i)})$ from block to block. At the convergence, $J_1(\boldsymbol{\phi})$ is maximized at a point where the coherence function $\hat{C}_{sy}(f; \boldsymbol{\phi})$ approaches one at all frequencies. This implies that the overall system between $s(n)$ and $y(n)$ has been linearized.

A Wiener system consists of an LTI block followed by a memoryless nonlinear block, thus the pre-inverse of a Hammerstein system is a Wiener system. The pre-inverse is defined such that its concatenation with the original nonlinear system equals identity. Therefore, the predistorter of a Hammersterin system can also be designed by identifying a Wiener system. Most Wiener system identification methods such as [15, 40] solve for the system parameters of both the linear and the nonlinear parts simultaneously. Compared with those methods, our coherence function-based method uses only a nonlinear block to compensate for the memoryless nonlinearity in the Hammerstein system, and works independently of the subsequent LTI block. Therefore, the proposed method is robust even if the LTI system parameters are unknown. This is a desirable quality for AEC applications since the echo canceller filter can be as long as a few thousand taps. Also, in traditional nonlinear system identification procedures, the computational complexity grows exponentially with the memory length. In contrast, the coherence function-based method is insensitive to the presence of an LTI system, so the length of the LTI system impulse response has no effect on the computational complexity of the nonlinearity identification stage. In the AEC application, once the nonlinear part is compensated for, the residual LTI system can be compensated by a linear AEC as usual.

3.2.2 Simulations

In the simulation examples shown, the source signal $s(n)$ was generated according to an i.i.d. Gaussian distribution. For the nonlinear acoustic echo path, loudspeaker nonlinearity is modeled by $d(s) = \tanh(s) = (e^{2s} - 1)/(e^{2s} + 1)$, where $\tanh(\cdot)$ denotes the hyperbolic tangent function. For the LTI system, the IMAGE method [2] was used to generate a room

impulse response of the length 1024 with the sampling rate 8 kHz. The nonlinearity in the loudspeaker is known to be smooth or mild and can be effectively modeled by polynomials [60]. Thus, it is appropriate to approximate the nonlinearity in the predistorter block $g(\cdot)$ in Fig. 6 using the polynomial basis $b_k(s) = s^k,\ k = 1, 2, ..., K$. K is the highest order of the polynomial basis, empirically selected to be 7. We are thus assuming a modeling error, since neither $d(\cdot)$ nor its inverse is precisely a polynomial function. The simulations were carried out in a noise-free environment. For implementation, we adopted a quasi-Newton method with a mixed quadratic and cubic line search procedure and used as initial estimate, $\boldsymbol{\phi}^{(0)} = [1, 0, \ldots, 0]^T$. Figure 7 (a) shows two estimated coherence functions: $\hat{C}_{sy}(f)$



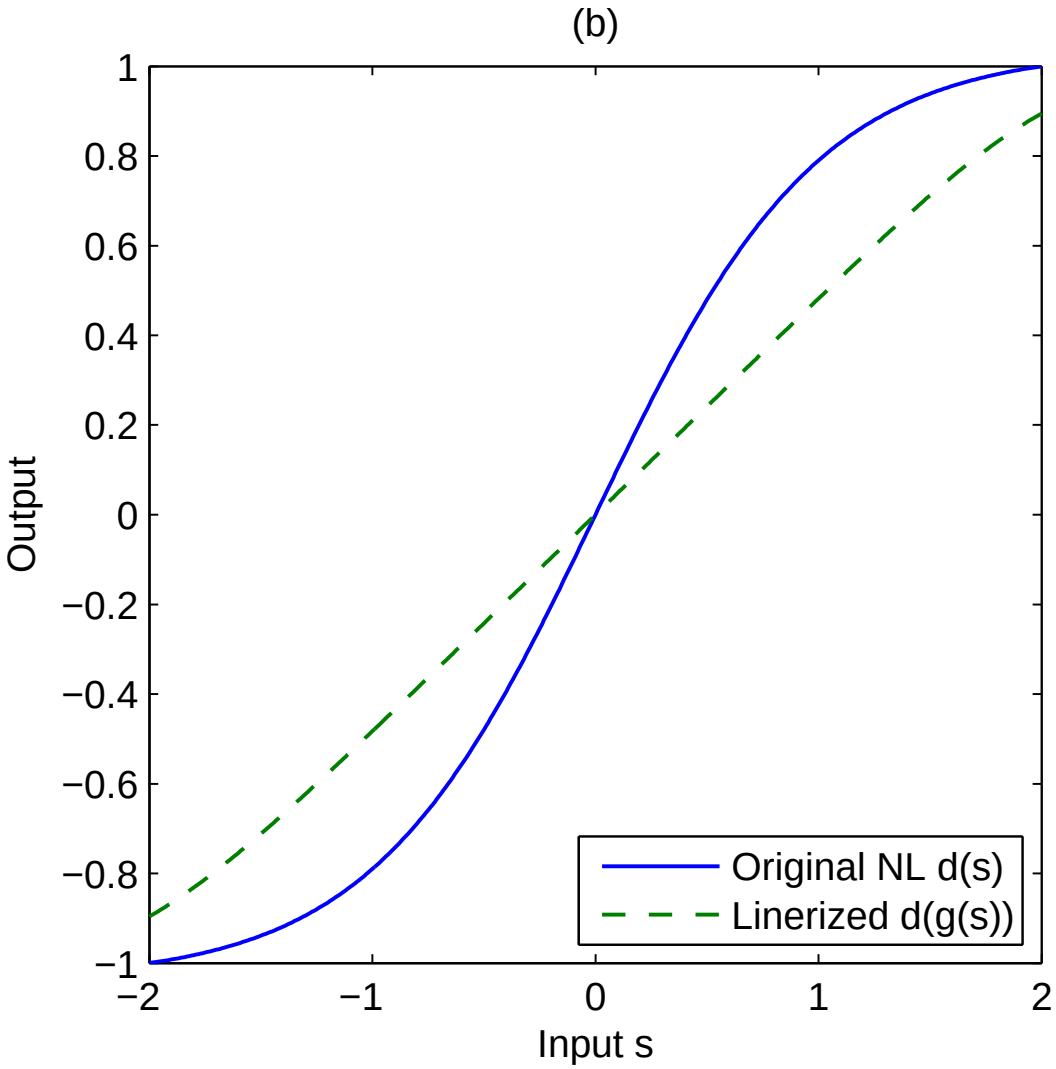


Figure 7: Predistortion: (a) Estimated coherence functions; (b) Linearization performance.

between the system input and the system output with predistortion, and $\hat{C}^o_{sy}(f)$ between the system input and the system output without predistortion. With the predistortion, $\hat{C}_{sy}(f)$ approaches one at each normalized frequency f, indicating that $s(n)$ and $y(n)$ are basically linearly related, and the nonlinearity in the system has been effectively removed. Figure 7 (b) shows the linearization result using the predistorter. It can be seen that the concatenated system consisting of the predistorter $g(\cdot)$ followed by the nonlinear block $d(\cdot)$ has an approximate linear characteristic.

3.3 NAEC with a Cascade Nonlinear Filter

As discussed in Section 3.2, predistrotion can effectively remove the nonlinear echo by linearizing the acoustic echo path. However, the echo cancellation is carried out in two stages and the nonlinearity can not be adaptively linearized. From the perspective of practical applications, an adaptive method is desirable in echo cancellation systems since the LEMS is usually time-varying. Rather than performing the linearization in the predistortion method, we adopt the cascade NAEC structure (see Fig. 2) and propose to adaptively compensate for the loudspeaker nonlinearity. Since the nonlinear AEC uses a cascade nonlinear filter to model the nonlinear LEMS, it consists of a memoryless nonlinear block $u(\cdot;\boldsymbol{\theta})$ and a linear block $h(n)$ corresponding to the PA/loudspeaker nonlinearity $d(\cdot)$ and the room impulse response, respectively. In this section, we first introduce a pseudo-MSC function-based method to identify the nonlinearity in the acoustic echo path. Then, an on-line implementation of the NAEC design is presented.

3.3.1 Nonlinearity Identification Using the Pseudo-MSC Function

Let us model $u(\cdot;\boldsymbol{\theta})$ as a linear combination of nonlinear basis functions $b_k(s)$ with corresponding coefficients θ_k:

$$u(s;\boldsymbol{\theta}) = \sum_{k=1}^{K} \theta_k b_k(s), \quad \boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_K]^T. \tag{57}$$

Given the system shown in Fig. 2, if the function $u(\cdot;\boldsymbol{\theta})$ is a perfect match to the true nonlinearity $d(\cdot)$, then the processes $y(n)$ and $z(n)$ are perfectly linearly related. Thus, the vector $\boldsymbol{\theta}$ in the nonlinear block $u(\cdot;\boldsymbol{\theta})$ can be found as follows [92, 90]:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \tag{58}$$

where

$$J(\boldsymbol{\theta}) = \int_{-0.5}^{0.5} \tilde{C}_{yz}(f;\boldsymbol{\theta}) \mathrm{d}f. \tag{59}$$

The motivation for us to adopt this pseudo-MSC function is twofold: (i) the pseudo-MSC function facilitates the closed-form solution for the nonlinear identification; and (ii) the pseudo-MSC function-based method provides the LMMSE performance.

Define

$$\boldsymbol{b}(n) = [b_1(s(n)), b_2(s(n)), ..., b_K(s(n))]^T . \tag{60}$$

For instance, if the basis functions are polynomials, we have $\boldsymbol{b}(n) = \left[s(n), s^2(n), ..., s^K(n)\right]^T$. The output signal of the nonlinear block can be expressed as (see Fig. 2)

$$z(n;\boldsymbol{\theta}) = u\left(s(n);\boldsymbol{\theta}\right) = \boldsymbol{\theta}^T\boldsymbol{b}(n). \tag{61}$$

From (61), we infer that

$$\sigma_z^2 = \boldsymbol{\theta}^T E\left[\boldsymbol{b}(n)\boldsymbol{b}^T(n)\right]\boldsymbol{\theta}, \tag{62}$$

$$S_{yz}(f;\boldsymbol{\theta}) = \boldsymbol{\theta}^T\boldsymbol{s}_{y\boldsymbol{b}}(f), \tag{63}$$

where $\boldsymbol{s}_{y\boldsymbol{b}}(f)$ is a vector with the k^{th} element being the cross-spectral density function between $y(n)$ and $b_k(s(n))$, $1 \le k \le K$. Substituting (62) and (63) into (38), we can rewrite the objective function in (59) as

$$J(\boldsymbol{\theta}) = \frac{\boldsymbol{\theta}^T\mathbf{R}_1\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{R}_2\boldsymbol{\theta}}, \tag{64}$$

where

$$\mathbf{R}_1 = \int_{-0.5}^{0.5} S_{yy}^{-1}(f)\boldsymbol{s}_{y\boldsymbol{b}}(f)\boldsymbol{s}_{y\boldsymbol{b}}^H(f)\mathrm{d}f, \tag{65}$$

$$\mathbf{R}_2 = E\left[\boldsymbol{b}(n)\boldsymbol{b}^T(n)\right], \tag{66}$$

and ^{H} denotes the Hermitian transpose. The ratio in (64) is known as the generalized Rayleigh quotient whose solution $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathbf{R}_1\hat{\boldsymbol{\theta}} = \lambda_{\max}\mathbf{R}_2\hat{\boldsymbol{\theta}}, \tag{67}$$

where $\lambda_{\max}$ is the largest generalized eigenvalue for the pair ($\mathbf{R}_1$, $\mathbf{R}_2$).

We have shown that the identification of the nonlinear block $u(\cdot;\boldsymbol{\theta})$ can be carried out independently from the linear part $h(n)$. This approach sets itself apart from other conventional methods, whereby the estimations of $u(\cdot;\boldsymbol{\theta})$ and $h(n)$ are coupled using the MMSE criterion.

3.3.2 Adaptive NAEC Using the Pseudo-MSC Function

In this part, we introduce an on-line implementation of the pseudo-MSC function-based nonlinearity identification method. Consider the objective function in (64). From the Rayleigh-Ritz theorem [32], all the generalized eigenvectors of matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ are the stationary points of $J(\boldsymbol{\theta})$ and the generalized eigenvalues are the values of $J(\boldsymbol{\theta})$ evaluated at the corresponding stationary points. This is because when the first-order derivative of $J(\boldsymbol{\theta})$ is set to zero,

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\boldsymbol{\theta}^T\mathbf{R}_2\boldsymbol{\theta}(2\mathbf{R}_1\boldsymbol{\theta}) - \boldsymbol{\theta}^T\mathbf{R}_1\boldsymbol{\theta}(2\mathbf{R}_2\boldsymbol{\theta})}{\left(\boldsymbol{\theta}^T\mathbf{R}_2\boldsymbol{\theta}\right)^2} = \boldsymbol{0}, \tag{68}$$

we obtain

$$\mathbf{R}_1\boldsymbol{\theta} = \frac{\boldsymbol{\theta}^T\mathbf{R}_1\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{R}_2\boldsymbol{\theta}}\mathbf{R}_2\boldsymbol{\theta} = J(\boldsymbol{\theta})\mathbf{R}_2\boldsymbol{\theta}. \tag{69}$$

Therefore, the following updating equation can be used to compute the principal generalized eigenvector [77]

$$\boldsymbol{\theta}(n) = \frac{\boldsymbol{\theta}^T(n-1)\mathbf{R}_2(n)\boldsymbol{\theta}(n-1)}{\boldsymbol{\theta}^T(n-1)\mathbf{R}_1(n)\boldsymbol{\theta}(n-1)}\mathbf{R}_2^{-1}(n)\mathbf{R}_1(n)\boldsymbol{\theta}(n-1). \tag{70}$$

In the following, we present an on-line algorithm to find $\mathbf{R}_1(n)$ and $\mathbf{R}_2(n)$. Define the m^{th} segment of signals $y(n)$ and $b_k(n)$ $(k = 1, 2, ..., K)$ as

$$\begin{aligned} y^{(m)}(l) &= y(mP + l),\ 0 \le l \le L-1, \\ b_k^{(m)}(l) &= b_k(mP + l),\ 0 \le l \le L-1,\ k = 1, 2, ..., K, \end{aligned}$$

where m and L are the index and the length of the data segment, respectively; P is the window sliding step size. The discrete time index n is related to the segment index m by $m = \lfloor n/P \rfloor$, where $\lfloor \cdot \rfloor$ rounds a number towards minus infinity. For the m^{th} segment vector, define

$$\begin{aligned} \boldsymbol{y}^{(m)} &= \left[y^{(m)}(0), ..., y^{(m)}(L-1)\right]^T, \\ \boldsymbol{b}_k^{(m)} &= \left[b_k^{(m)}(0), ..., b_k^{(m)}(L-1)\right]^T,\ k = 1, 2, ..., K. \end{aligned}$$

To obtain an unbiased estimate of the auto- and cross-spectral densities, we use the following recursive estimators:

$$\boldsymbol{s}_{y\boldsymbol{b}}^{(m)}(f_i) = \rho \boldsymbol{s}_{y\boldsymbol{b}}^{(m-1)}(f_i) + (1-\rho) Y^{(m)}(f_i) \left[\boldsymbol{B}^{(m)}(f_i)\right]^*,$$
$$S_{yy}^{(m)}(f_i) = \rho S_{yy}^{(m-1)}(f_i) + (1-\rho) \left|Y^{(m)}(f_i)\right|^2,$$

where * denotes conjugation; ρ is a forgetting factor with constraint $0 < \rho < 1$; $Y^{(m)}(f_i)$ is the discrete Fourier transform (DFT) of $\boldsymbol{y}^{(m)}$ at the i^{th} frequency bin, and $\boldsymbol{B}^{(m)}(f_i)$ is a $K \times 1$ vector with the k^{th} entry being the DFT of $\boldsymbol{b}_k^{(m)}$ at the i^{th} frequency bin. When the segment length L is short, we zero-pad each segment to length $Q \geq L$ to ensure sufficient resolution of the frequency axis. Therefore, we estimate $\mathbf{R}_1$ in (65) as follows:

$$\mathbf{R}_1(n) = \sum_{i=1}^{N} \boldsymbol{s}_{y\boldsymbol{b}}^{(m)}(f_i) \left[\boldsymbol{s}_{y\boldsymbol{b}}^{(m)}(f_i)\right]^H \Big/ S_{yy}^{(m)}(f_i)\,. \tag{71}$$

From (66), we form the following estimate of $\mathbf{R}_2$ at time n

$$\mathbf{R}_2(n) = \rho \mathbf{R}_2(n-1) + (1-\rho)\boldsymbol{b}(n)\boldsymbol{b}^T(n). \tag{72}$$

According to the Sherman-Morrison-Woodbury matrix inversion lemma [32, p. 50], $\mathbf{R}_2^{-1}(n)$ can be estimated recursively via

$$\mathbf{R}_2^{-1}(n) = \frac{1}{\rho}\mathbf{R}_2^{-1}(n-1) - \frac{1-\rho}{\rho}\frac{\mathbf{R}_2^{-1}(n-1)\boldsymbol{b}(n)\boldsymbol{b}^T(n)\mathbf{R}_2^{-1}(n-1)}{\rho + (1-\rho)\boldsymbol{b}^T(n)\mathbf{R}_2^{-1}(n-1)\boldsymbol{b}(n)}. \tag{73}$$

Based on (61), we express the n^{th} output of the nonlinear filter as $z(n) = \boldsymbol{\theta}^T(n-1)\boldsymbol{b}(n)$. Denote

$$\alpha(n) = \boldsymbol{\theta}^T(n-1)\mathbf{R}_2(n)\boldsymbol{\theta}(n-1). \tag{74}$$

Using (72), the recursive estimator of $\alpha(n)$ can be obtained as

$$\begin{aligned}\alpha(n) &= \rho\boldsymbol{\theta}^T(n-1)\mathbf{R}_2(n-1)\boldsymbol{\theta}(n-1) + (1-\rho)\boldsymbol{\theta}^T(n-1)\boldsymbol{b}(n)\boldsymbol{b}(n)^T\boldsymbol{\theta}(n-1) \\ &\approx \rho\alpha(n-1) + (1-\rho)z^2(n),\end{aligned} \tag{75}$$

where we assume that the consecutive values of $\boldsymbol{\theta}$ are approximately the same. Thus, we reduce the computational complexity by avoiding matrix multiplications. Substitution of (74) into (70) yields

$$\boldsymbol{\theta}(n) = \alpha(n)\frac{\mathbf{R}_2^{-1}(n)\mathbf{R}_1(n)\boldsymbol{\theta}(n-1)}{\boldsymbol{\theta}^T(n-1)\mathbf{R}_1(n)\boldsymbol{\theta}(n-1)}. \tag{76}$$

Table 1: On-line algorithm of NAEC based on the pseudo-MSC function.

Initialize
$\boldsymbol{\theta}(-1) \in \Re^{K\times 1}$ random vector
$\mathbf{R}_2^{-1}(-1) \in \Re^{K\times K}$ large random matrix
$\alpha(-1) \in \Re = 0$
$\boldsymbol{v}^{(-1)} \in \Re^{N\times 1} = \boldsymbol{0}$
$\boldsymbol{s}_k^{(-1)} \in \mathfrak{C}^{N\times 1} = \boldsymbol{0},\ k = 1, 2, ..., K$
Update
for $n = 0, 1, ..., N$
$z(n) = \boldsymbol{\theta}^T(n-1)\boldsymbol{b}(n)$
$\alpha(n) = \rho\alpha(n-1) + (1-\rho)z^2(n)$
$\mathbf{R}_2^{-1}(n) = \frac{1}{\rho}\mathbf{R}_2^{-1}(n-1)$
$\quad - \frac{1-\rho}{\rho}\frac{\mathbf{R}_2^{-1}(n-1)\boldsymbol{b}(n)\boldsymbol{b}^T(n)\mathbf{R}_2^{-1}(n-1)}{\rho+(1-\rho)\boldsymbol{b}^T(n)\mathbf{R}_2^{-1}(n-1)\boldsymbol{b}(n)}$
$m = \lfloor n/P \rfloor$
$\boldsymbol{Y}^{(m)} = \text{FFT}\left\{\boldsymbol{y}^{(m)}\right\}$
$\boldsymbol{B}_k^{(m)} = \text{FFT}\left\{\boldsymbol{b}_k^{(m)}\right\},\ k = 1, 2, .., K$
$\boldsymbol{v}^{(m)} = \rho\boldsymbol{v}^{(m-1)} + (1-\rho)\left\|\boldsymbol{Y}^{(m)}\right\|^2$
$\boldsymbol{s}_k^{(m)} = \rho\boldsymbol{s}_k^{(m-1)} + (1-\rho)\boldsymbol{Y}^{(m)} \otimes \left[\boldsymbol{B}_k^{(m)}\right]^*, k = 1, 2, ..., K$
$\boldsymbol{u}_i^{(m)} = \left[s_1^{(m)}(i), s_2^{(m)}(i), ..., s_K^{(m)}(i)\right]^T,\ i = 1, 2, ..., Q$
$\mathbf{R}_1(n) = \sum_{i=1}^{i=Q}\boldsymbol{u}_i^{(m)}\left[\boldsymbol{u}_i^{(m)}\right]^H / v^{(m)}(i)$
$\boldsymbol{Q}(n) = \mathbf{R}_1(n)\boldsymbol{\theta}(n-1)$
$\beta(n) = \boldsymbol{\theta}^T(n-1)\boldsymbol{Q}(n)$
$\boldsymbol{\theta}(n) = \frac{\alpha(n)}{\beta(n)}\mathbf{R}^{-1}(n)\boldsymbol{Q}(n)$
$e(n) = y(n) - \boldsymbol{h}(n)^T\boldsymbol{z}(n)$
$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \mu\frac{\boldsymbol{z}(n)}{\|\boldsymbol{z}(n)\|_2^2}e(n)$
end for n

The updating equations (71), (73), and (75) for $\mathbf{R}_1(n)$, $\mathbf{R}_2^{-1}(n)$, and $\alpha(n)$, respectively, give rise to an on-line algorithm for implementing (76).

In acoustic echo cancellation applications, the update of the linear part can be implemented by the NLMS algorithm. Table 1 summarizes our adaptive NAEC algorithm based on the pseudo-MSC function, where $\otimes$ denotes the element-wise product between two vectors. The advantage of the proposed method is that it identifies the nonlinearity without knowing the linear block in the Hammerstein system, which guarantees the stability of cascade nonlinear filter and leads to a faster convergence rate.

3.3.3 Simulations

In this section, the performance of the proposed methods is assessed via computer simulations. In the following simulation examples, the far-end signal $s(n)$ was generated according to an i.i.d. Gaussian distribution or real speech signal. For the acoustic echo path, both the loudspeaker nonlinearity and room impulse response were generated in the same way as in Section 3.2.2. The nonlinear block $u(\cdot;\boldsymbol{\theta})$ in AEC used polynomial basis functions with the highest order K selected to be 7. The block size used in the Welch method is $L = 256$, and the window sliding step size is $P = 192$.

3.3.3.1 Identification Performance

To quantitatively evaluate the system identification performance, misadjustment is taken as a figure of merit. For the nonlinear part misadjustment, we use the normalized mean squared error (NMSE) defined as

$$\text{NMSE (dB)} = 10\log_{10}\frac{\sum_{m=1}^{M}|d(s(m)) - u(s(m))|^2}{\sum_{m=1}^{M}|d(s(m))|^2}, \tag{77}$$

where the number of samples was $M = 65,536$. For the linear part misadjustment, we adopt the distance measure as

$$D_h(\text{dB}) = 10\log_{10}\frac{\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|_2^2}{\|\boldsymbol{h}\|_2^2}, \tag{78}$$

where $\|\cdot\|_2$ denotes the L_2 norm. With the proposed method, the NMSE and D_h were -48.5 dB and -43.1 dB, respectively, indicating that both the nonlinear and linear parts converged to the true values. Thus, the simulation results validated the theoretical analysis. The nonlinear coefficients update in (70) is analogous to the RLS update rule that tracks the Wiener solution and the convergence is proved using the ordinary differential equation (ODE) in [77]. As long as the convergence of the nonlinear part is guaranteed, the convergence of the linear part is reduced to the convergence of the traditional NLMS algorithm, which has been well documented in the adaptive filtering theory literature [42].

3.3.3.2 Stability and Convergence

This part presents the echo cancellation performance with stationary and real speech input, respectively. Echo return loss enhancement (ERLE) is used as a figure of merit for the

performance of AECs:

$$\text{ERLE (dB)} = 10\log_{10}\frac{E\left[y^2(n)\right]}{E\left[e^2(n)\right]}, \tag{79}$$

where $y(n)$ and $e(n)$ represent the microphone-received signal and the residual echo signal, respectively.

Define the signal-to-noise ratio (SNR) as the ratio of the echo signal level to the background noise level:

$$\text{SNR (dB)} = 10\log_{10}\frac{E\left[c^2(n)\right]}{E\left[v^2(n)\right]}. \tag{80}$$

The signal $y(n)$ is generated for a single-talk situation with the additive white Gaussian noise (AWGN) such that an SNR of 35 dB is achieved. For comparison purposes, we also implement the RLS-like algorithm [101]. In [101], the auto-covariance matrix needs to be reinitialized every δ samples to avoid instability. Figure 8 (a) shows the ERLEs of the proposed method and the RLS-like method (for different δ) with a Gaussian signal as input. We notice that the increase of δ leads to the faster convergence, while unlimitedly increasing δ causes the algorithm divergence. Comparatively, our proposed method further improves the convergence rate and guarantees the stability. This is mainly because the proposed method decouples the nonlinear identification from the estimation of the linear part.

Although our analysis is based on the stationary input, we expect the proposed method also works for real speech signals, because we investigate not only static status as ERLE, but also dynamic properties as convergence rate. With an SNR of 30 dB, the ERLEs with speech signal input are depicted in Fig. 8 (b), which also demonstrates the effectiveness of the proposed method.

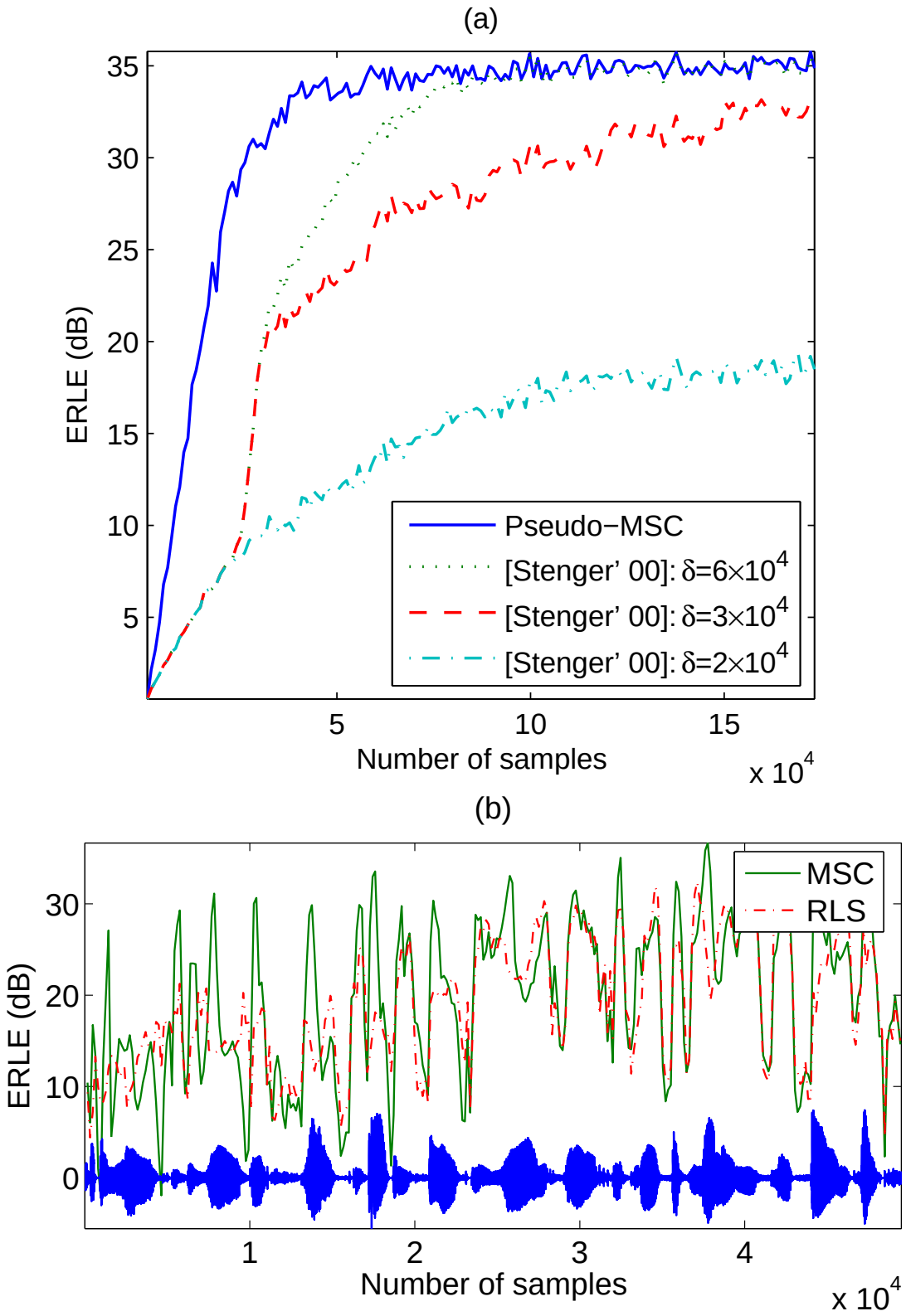


Figure 8: ERLE in a cascade architecture: (a) white Gaussian input; (b) real speech input.

3.3.4 ITU-T G.167 Test

The AEC performance criteria and their related measurement methods are defined by the International Telecommunications Union (ITU)-T Recommendation G. 167 [78]. The Recommendation specifies the performance characteristics and values with which AECs should comply to. The diagram for the AEC performance evaluation is shown in Fig. 9. The AEC performance criteria are referred to four interfaces [78]:

(1) User receive interface (R_{out}): The place(s) where acoustic attributes relating to the characteristics of speech listened to by the local user(s) are measured.

(2) User send interface (S_{in}): The place(s) where acoustic attributes relating to the characteristics of speech produced by the local user(s) are measured.

(3) Network receive interface (R_{in}): A point where the electrical signals received from the network are available.

(4) Network send interface (S_{out}): A point where the electrical signals sent to the network are available.

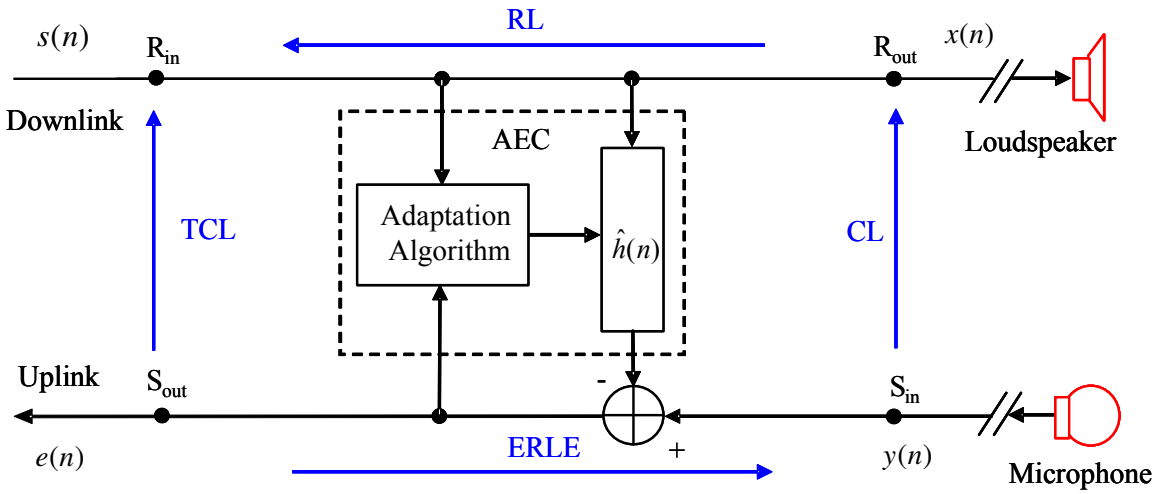


Figure 9: ITU-T G. 167 test of an AEC.

In this test, we consider two types of criteria: the coupling loss and the time adaptivity.

(I) Coupling Loss

The AEC coupling loss parameters include (see Fig. 9):

- Terminal Coupling Loss (TCL or TCLw), defined as

$$\text{TCL (dB)} = 10\log_{10}\frac{E[s^2(n)]}{E[e^2(n)]}. \tag{81}$$

- Receive Loss (RL), defined as

$$\text{RL (dB)} = 10\log_{10}\frac{E[s^2(n)]}{E[x^2(n)]}. \tag{82}$$

- Coupling Loss (CL), defined as

$$\text{CL (dB)} = 10\log_{10}\frac{E[x^2(n)]}{E[y^2(n)]}. \tag{83}$$

- Echo Return Loss Enhancement (ERLE), defined as

$$\text{ERLE (dB)} = 10\log_{10}\frac{E[y^2(n)]}{E[e^2(n)]}. \tag{84}$$

(II) Time Adaptivity

The time adaptivity parameters represent the AEC ability to converge during the initial time, after double-talk situations and echo path variations in a noiseless environment. The AEC time adaptivity parameters include

- Initial convergence time (Tic)
- Recovery time after double-talk (Trdt)
- Recovery time after path variation (Trpv)

According to the Recommendation G. 167 test procedure, we evaluate the terminal coupling loss during the single-talk (TCLwst), double-talk mode (TCLwdt), and echo path variation (TCLwpv). In addition, we measure the time adaptability parameters (Tic, Trdt, Trpv). The test procedure is briefly described in Appendix B. The Recommendation requirements and test results are summarized in Table 2. It can be seen that the proposed NAEC meets all the test requirements. And for every test, it leaves margin for the design of additional components in the system.

Table 2: ITU-T G. 167 test results.

Quantity	Description	Result (dB)	Requirement
TCLwst	Echo loss in single talk	63.3	> 45 dB
TCLwdt	Echo loss in double talk	32.7	> 25 dB
TCLwpv	Echo loss during echo path variation	35.7	> 20 dB
Tic	Initial convergence time	30.2	1 sec, 20 dB
Trdt	Recovery time after double talk	28.2	1 sec, 20 dB
Trpv	Recovery time after echo path variation	42.3	1 sec, 20 dB

3.4 NRES Using the MSC Function

As discussed in Section 2.2.2, in lieu of the NAEC, the NRES technique can be applied to remove the nonlinear acoustic echo. Moreover, it is shown that the optimal gain $C(f)$ of the NRES can be reduced to the estimation of the PSD of the nonlinear residual echo $p(n)$. In this section, we propose to estimate the PSD of $p(n)$ using the multiple coherence function [86]. Compared to the method in [61], the estimation of the PSD of the nonlinear residual echo bypasses the estimation of the additional filter coefficients. Therefore, our proposed method improves the convergence rate and is robust to the length of the acoustic echo path.

3.4.1 Residual Echo Power Estimation

Suppose that the LEMS in Fig. 3 consists of a (memoryless) nonlinear amplifier and/or loudspeaker followed by a linear subsystem (the room impulse response). We model the nonlinearity $f(\cdot)$ in the amplifier/loudspeaker block as a linear combination of basis functions $b_k(\cdot)$ with corresponding coefficients α_k:

$$d(s;\boldsymbol{\alpha}) = \sum_{k=1}^{K} \alpha_k b_k(s), \tag{85}$$

where $b_k(s) = s^k$ and K is the order of nonlinearity. If the room impulse response is modeled by an FIR filter $h(n)$ with length L_h, the nonlinear acoustic echo can be expressed as

$$y(n) = \sum_{l=0}^{L_h-1} h(l) \sum_{k=1}^{K} \alpha_k b_k(s(n-l)). \tag{86}$$

Assuming that the AEC uses an FIR filter $\hat{h}(n)$ (an estimate of $h(n)$) with length L_h', we obtain the estimated echo as

$$\hat{y}(n) = \sum_{l=0}^{L_h'-1} \hat{h}(l)s(n-l). \tag{87}$$

Usually we choose $L_h' < L_h$, since we can decrease the computational complexity by sacrificing some echo cancellation performance when the reverberation time of the room response is too long. Combining (86), (87), and (12), we obtain the nonlinear residual echo:

$$\begin{aligned} p(n) &= \sum_{l=0}^{L_h-1} h(l) \sum_{k=1}^{K} \alpha_k b_k(s(n-l)) - \sum_{l=0}^{L_h'-1} \hat{h}(l)s(n-l) \\ &= \sum_{k=1}^{K} g_k(n) * x_k(n), \end{aligned} \tag{88}$$

where

$$g_k(n) = \begin{cases} \alpha_k h(n) - \hat{h}(n), & k = 1 \\ \alpha_k h(n), & k = 2, ..., K \end{cases} \tag{89}$$

and $x_k(n) = b_k(s(n))$. Therefore, the nonlinear residual echo $p(n)$ can be treated as the output signal of a multiple-input single-output (MISO) system with the input signals being $x_k(n)$ $(k = 1, ..., K)$. Due to the presence of nonlinearity, there is much energy in the residual echo. Thus, we employ a postfilter to further suppress the echo.

The Fourier transform of (88) yields

$$P(f) = \sum_{k=1}^{K} G_k(f)X_k(f), \tag{90}$$

where $G_k(f)$ and $X_k(f)$ are the Fourier transforms of $g_k(n)$ and $x_k(n)$, respectively. Define vectors

$$\boldsymbol{G}(f) = [G_1(f), G_2(f), ..., G_K(f)]^T, \tag{91}$$

$$\boldsymbol{X}(f) = [X_1(f), X_2(f), ..., X_K(f)]^T. \tag{92}$$

Using (90), the PSD of $p(n)$ can be expressed as

$$S_p(f) = E\left[|P(f)|^2\right] = \boldsymbol{G}^H(f)\mathbf{S}_{xx}(f)\boldsymbol{G}(f), \tag{93}$$

where

$$\mathbf{S}_{xx}(f) = E\left[\boldsymbol{X}^*(f)\boldsymbol{X}^T(f)\right] = \begin{bmatrix} S_{11}(f) & \cdots & S_{1K}(f) \\ S_{21}(f) & \cdots & S_{2K}(f) \\ \vdots & & \vdots \\ S_{21}(f) & \cdots & S_{KK}(f) \end{bmatrix} \tag{94}$$

is the autocorrelation matrix of $\boldsymbol{X}(f)$, with its ij^{th} element being the cross-spectral density function between $x_i(n)$ and $x_j(n)$. Furthermore, the linear MMSE solution for $G_k(f)$ of (90) can be calculated as

$$\boldsymbol{s}_{xp}(f) = \mathbf{S}_{xx}\ \boldsymbol{G}(f), \tag{95}$$

where

$$\boldsymbol{s}_{xp}(f) = E\left[\boldsymbol{X}^*(f)P(f)\right] = \left[S_{1p}(f), ..., S_{Kp}(f)\right]^T \tag{96}$$

is the cross-correlation vector between $\boldsymbol{X}(f)$ and $P(f)$, with its i^{th} element being the cross-spectral density function between $x_i(n)$ and $p(n)$. Combining (93) and (95), we obtaine the PSD of the nonlinear residual echo as

$$S_p(f) = \boldsymbol{s}_{xp}^H(f)\mathbf{S}_{xx}^{-1}(f)\boldsymbol{s}_{xp}(f). \tag{97}$$

Note that the nonlinear residual echo $p(n)$ is not accessible, since it is hidden in the microphone signal $r(n)$. Assume that the near-end speech, the background noise, and the far-end speech are mutually independent of each other. Thus, $\boldsymbol{s}_{xp}(f) = \boldsymbol{s}_{xr}(f)$, and correspondingly (97) can be rewritten as

$$S_p(f) = \boldsymbol{s}_{xr}^H(f)\mathbf{S}_{xx}^{-1}(f)\boldsymbol{s}_{xr}(f). \tag{98}$$

Since the signals $x_i(n)$ and $r(n)$ are known, the recursive estimate of the i^{th} entry in $\boldsymbol{s}_{xr}^H(f)$ and the ij^{th} entry in $\mathbf{S}_{xx}$ can be given, respectively, as

$$\left[\hat{\boldsymbol{s}}_{xr}^{(m)}(f)\right]_i = \lambda\left[\hat{\boldsymbol{s}}_{xr}^{(m-1)}(f)\right]_i + (1-\lambda)[X_i^{(m)}(f)]^* R(f), \tag{99}$$

$$\left[\hat{\mathbf{S}}_{xx}^{(m)}(f)\right]_{ij} = \lambda\left[\hat{\mathbf{S}}_{xx}^{(m-1)}(f)\right]_{ij} + (1-\lambda)[X_i^{(m)}(f)]^* X_j^{(m)}(f). \tag{100}$$

To avoid high computational complexity associated with the matrix inversion, $\mathbf{S}_{xx}^{-1}(f)$ can be calculated recursively according to the Sherman-Morrison-Woodbury matrix inversion

lemma [44]:

$$(\mathbf{S}_{xx}^{(m)}(f))^{-1} = \frac{1}{\lambda}(\mathbf{S}_{xx}^{(m-1)}(f))^{-1} - \left(1-\frac{1}{\lambda}\right) \cdot \frac{(\mathbf{S}_{xx}^{(m-1)}(f))^{-1}(\boldsymbol{X}^{(m)}(f))^{*}(\boldsymbol{X}^{(m)}(f))^{T}(\mathbf{S}_{xx}^{(m-1)}(f))^{-1}}{(\boldsymbol{X}^{(m)}(f))^{T}(\mathbf{S}_{xx}^{(m-1)}(f))^{-1}(\boldsymbol{X}^{(m)}(f))^{*}}. \quad (101)$$

Therefore, the PSD estimate of the nonlinear residual echo $\hat{S}_p(f)$ can be obtained by substituting (99), (100), and (101) into (98). Correspondingly, the nonlinear gain $C(f)$ can be found using (14).

Remark: We recognize that (98) can be rewritten as

$$S_p(f) = \frac{\boldsymbol{s}_{xr}^{H}(f)\mathbf{S}_{xx}^{-1}(f)\boldsymbol{s}_{xr}(f)}{S_r(f)} \cdot S_r(f) = \Gamma_{x_1 \ldots x_K, r}(f) \cdot S_r(f), \quad (102)$$

where $\Gamma_{x_1 \ldots x_K, r}(f)$ is the so-called multiple coherence function [74]. It can be shown that $0 \leq \Gamma_{x_1 \ldots x_K, r}(f) \leq 1, \ \forall f$; and the multiple cohere function $\Gamma_{x_1 \ldots x_K, r}(f)$ indicates the fraction of the power in the signal $r(n)$ that is attributed to the linear combination of $x_1(n), ..., x_K(n)$. Therefore, (102) extracts the power of the signal that is related to $x_1(n), ..., x_K(n)$ from the signal $r(n)$. This is exactly the PSD of the nonlinear residual echo signal.

3.4.2 Simulations

The performance of the proposed method is assessed via computer simulations. The nonlinearity of the power amplifier/loudspeaker is modeled by a third-order polynomial function:

$$d(s) = -0.0325s^3 - 0.0003s^2 + 0.4824s. \quad (103)$$

The room impulse response was generated according to

$$h(n) = \begin{cases} \beta(n)e^{-\alpha n}, & 4 \leq n \leq L_h \\ 0, & \text{otherwise} \end{cases} \quad (104)$$

where $\beta(n)$ was i.i.d. standard Gaussian distributed; $L_h = 512$ and $\alpha = 0.004$. A white noise $v(n)$ was added and the resulting SNR was 30 dB. Here, the linear AEC is implemented using the frequency-domain NLMS algorithm [41], since the subsequent NRES is also performed in the frequency domain.

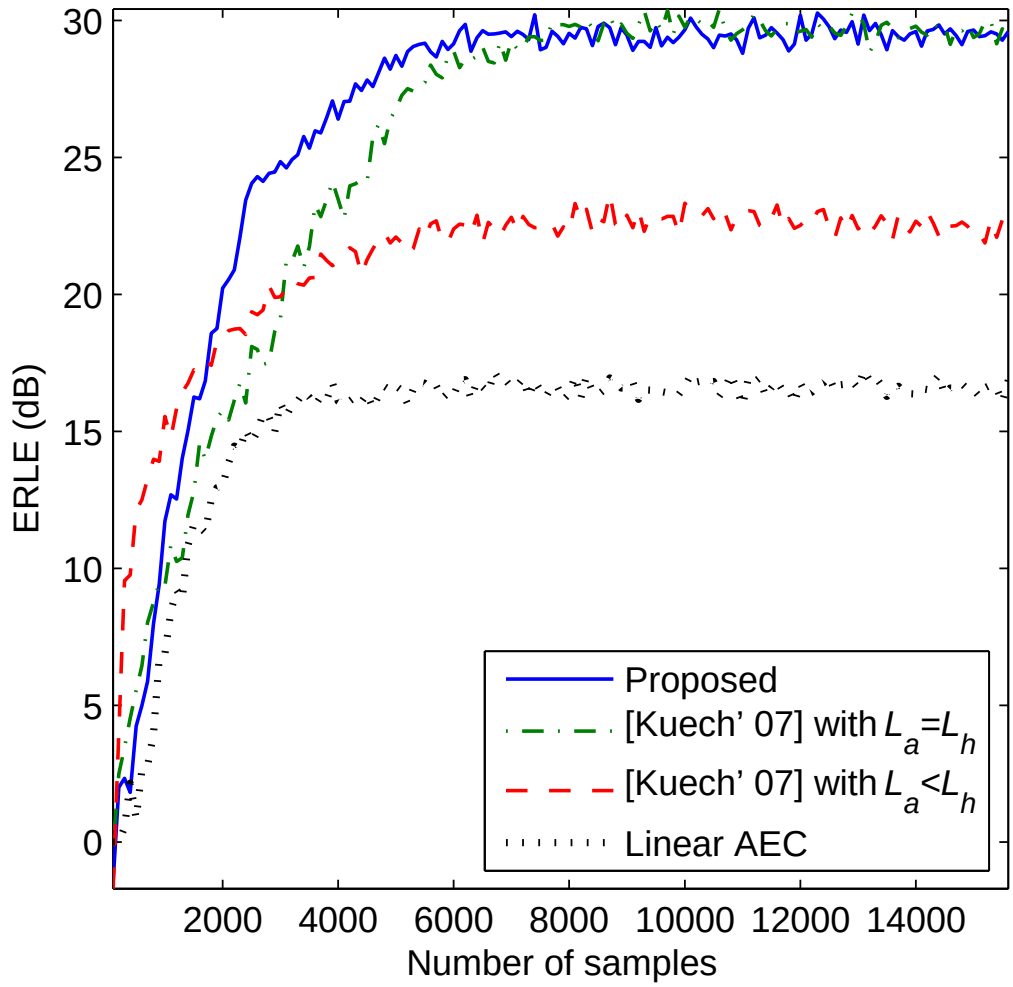


Figure 10: Performance of the linear AEC with/without the NRES with a white Gaussian noise as the input signal.

In the first experiment, we used a white Gaussian noise for the far-end signal. For comparison purposes, we also implemented the method of [61]. The ERLEs obtained for different approaches are shown in Fig. 10. We can see that both the nonlinear approaches remarkably improved the echo attenuation performance compared to the purely linear AEC. The proposed method outperforms the method of [61] in terms of the convergence rate. This is because the proposed method estimates $S_p(f)$ directly, whereas the estimate of $S_p(f)$ in [61] depends on the convergence of another filter with length L_a. The major advantages of the proposed method are that it bypasses the estimation of the additional filter coefficients

and requires no knowledge of the room impulse response length L_h. This can also be seen from Fig. 10, where the method in [61] uses $L_a = 350$ ($< L_h$) for the additional adaptive filter to estimate $S_p(f)$. It is seen that inadequate filter length gives rise to a large bias in the estimate of $S_p(f)$, and correspondingly the performance of the algorithm in [61] is degraded, whereas the proposed method is not affected.

Next, we evaluated the performance of the proposed method using speech data as the input signal. In Fig. 11, we show the ERLEs obtained with the proposed NRES and with a linear RES (LRES) in [39]. We notice that the nonlinear approach provides a consistent increase in echo attenuation throughout the data frame.

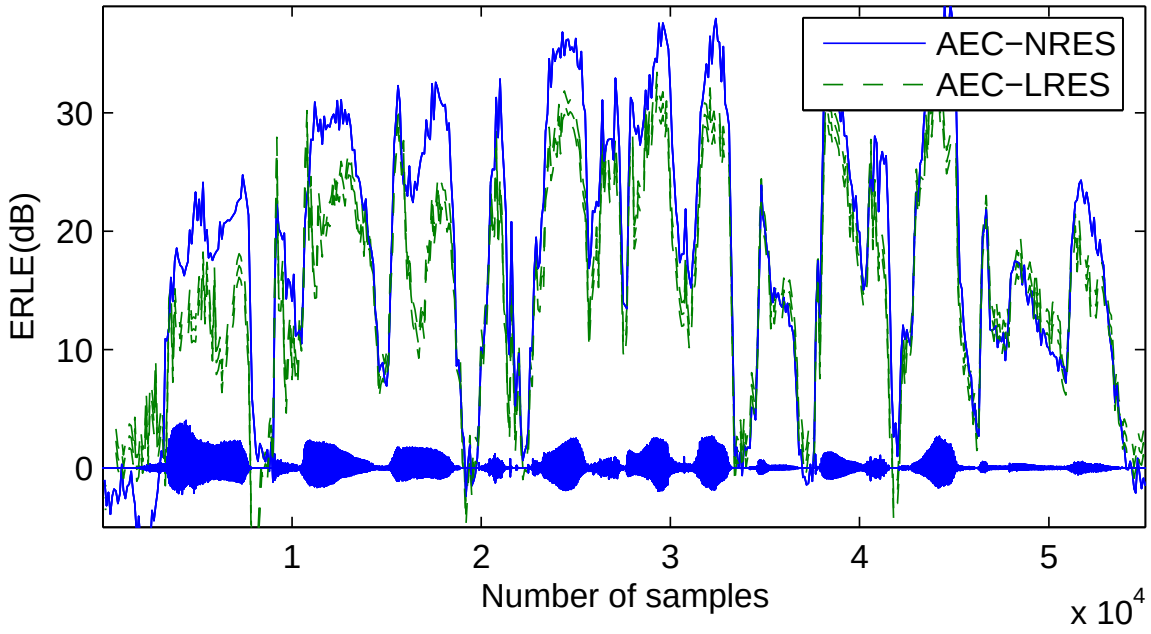


Figure 11: Performance of the AEC with LRES and NRES using speech as the input signal.

In the last experiment, we evaluated the performance of the proposed method in the double-talk situation. The far-end speech $s(n)$ is shown in Fig. 12(a). The near-end speech $z(n)$ is depicted in Fig. 12(b). In Fig. 12(c), the NRES output signal $e(n)$ is shown. It can be seen that the near-end speech is hardly distorted, while the echo signal has been sufficiently suppressed.

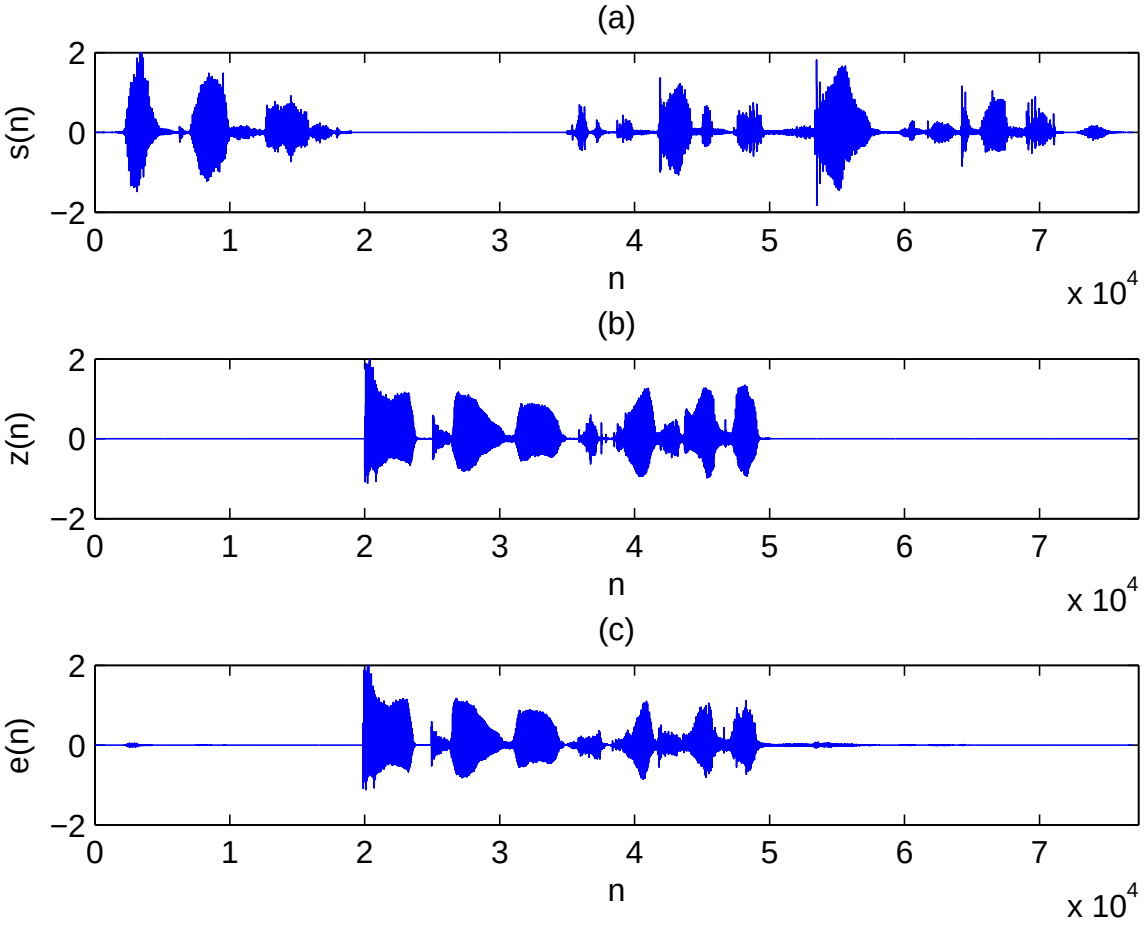


Figure 12: Different signals for nonlinear acoustic echo cancellation: (1) far-end speech $s(n)$, (2) near-end speech $z(n)$, (3) NRES output $e(n)$.

3.5 Cascade NAEC with a Shortening Filter

In this section, we design an efficient AEC that exhibits fast convergence rate and low complexity by using a shortening filter. Figure 13 shows the architecture of our proposed nonlinear AEC. We introduce an FIR filter $w(n)$ after the acoustic echo path (see Fig. 13) [91]. Let the LEMS consist of a (memoryless) nonlinear PA/loudspeaker followed by a linear subsystem (the room impulse response); it can be well represented by a Hammerstein model. The linear room impulse response followed by an FIR filter $w(n)$ is still a linear system. The purpose of introducing $w(n)$ is to make the "effective" channel, which is the convolution of the room impulse response and $w(n)$, have a smaller number of dominant

taps. Although in theory, the convolution of two sequences yields a longer sequence, when designed carefully, a more compact sequence can result [1]. The shortening filter design has also been proposed in other applications. For example, a shortening filter has been used to shorten the long impulse response of twisted-pair telephone lines, and it is more commonly referred to as a channel shortening equalizer (CSE) [69].

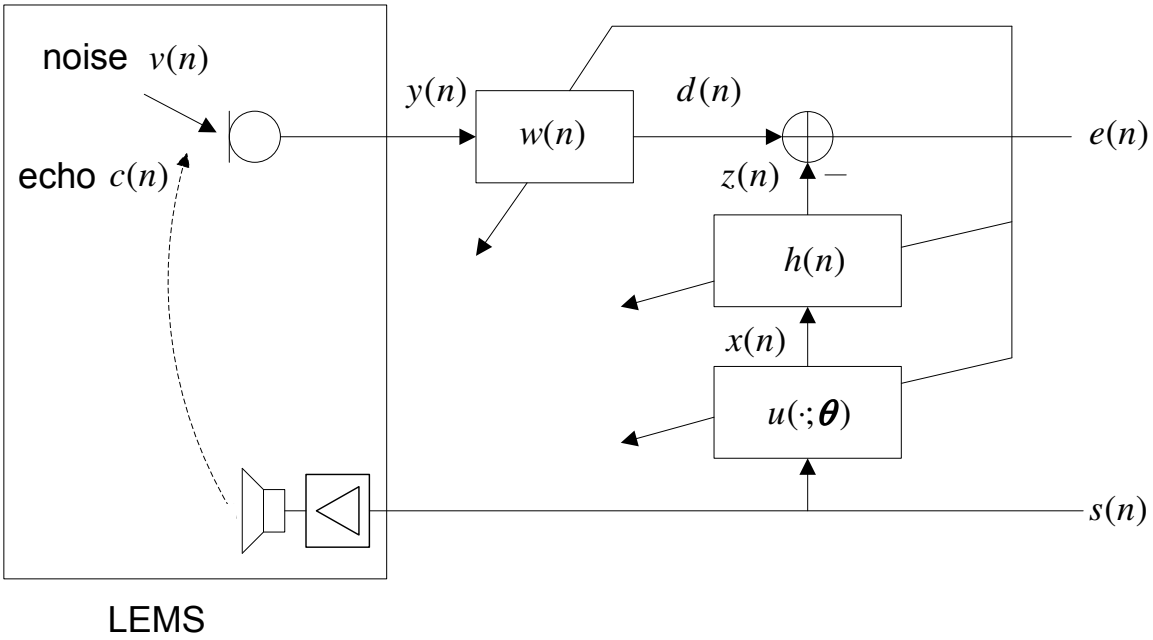


Figure 13: Nonlinear AEC structure with a shortening filter.

Suppose that the shortening filter $w(n)$ and the AEC filter $h(n)$ have lengths L_w and L_h, respectively. Define vectors

$$\boldsymbol{w}(n) = [w_0(n), w_1(n), ..., w_{L_w-1}(n)]^T, \tag{105}$$

$$\boldsymbol{y}(n) = [y(n), y(n-1), ..., y(n-L_w+1)]^T. \tag{106}$$

The reference signal $d(n)$ can be expressed as

$$d(n) = \boldsymbol{w}^T(n)\boldsymbol{y}(n). \tag{107}$$

For the AEC branch, we use the nonlinear basis expansion to model the nonlinear function

$u(\cdot;\boldsymbol{\theta})$ as in (57). Define vectors

$$\boldsymbol{b}(n) = [b_1(s(n)), b_2(s(n)), ..., b_K(s(n))]^T, \tag{108}$$

$$\boldsymbol{h}(n) = [h_0(n), ..., h_{L_h-1}(n)]^T, \tag{109}$$

$$\boldsymbol{x}(n) = [x(n), ..., x(n - L_h + 1)]^T, \tag{110}$$

and a matrix

$$\mathbf{S}(n) = [\boldsymbol{b}(n), \boldsymbol{b}(n-1), ..., \boldsymbol{b}(n - L_h + 1)]. \tag{111}$$

The output of the AEC branch is

$$z(n) = \boldsymbol{h}^T(n)\boldsymbol{x}(n) = \boldsymbol{\theta}^T(n)\mathbf{S}(n)\boldsymbol{h}(n). \tag{112}$$

Our goal is to design $w(n)$, $u(\cdot;\boldsymbol{\theta})$ and $h(n)$ such that $d(n)$ and $z(n)$ approximately cancel each other in a single-talk scenario (i.e., the near-end speech is not present). The purpose of the shortening filter $w(n)$ is to reduce the required number of taps in $h(n)$ and thus reduce the complexity and improve the convergence rate of the AEC.

From Fig. 13, the error signal $e(n)$ can be written as [cf. (107) and (112)]

$$e(n) = d(n) - z(n) = \boldsymbol{w}^T(n)\boldsymbol{y}(n) - \boldsymbol{\theta}^T(n)\mathbf{S}(n)\boldsymbol{h}(n). \tag{113}$$

Then, the MSE can be expressed as

$$J(\boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{w}) = E\left[e^2(n)\right] = E\left[\left(\boldsymbol{w}^T(n)\boldsymbol{y}(n) - \boldsymbol{\theta}^T(n)\mathbf{S}(n)\boldsymbol{h}(n)\right)^2\right]. \tag{114}$$

We propose the following criterion to solve the unknown parameters [91]:

$$[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{h}}, \hat{\boldsymbol{w}}] = \arg \min_{\boldsymbol{\theta},\boldsymbol{h},\boldsymbol{w}} J(\boldsymbol{\theta}, \boldsymbol{h}, \boldsymbol{w}), \quad \text{subject to } \|\boldsymbol{\theta}\|_2 = 1, \ \|\boldsymbol{h}\|_2 = 1, \tag{115}$$

where the constraints are added to avoid trivial solutions and $\|\cdot\|_2$ denotes the l_2 norm.

3.5.1 Filter Coefficients Update

3.5.1.1 Adaptive Algorithm to Update the Linear Filters

Since the error signal in (113) is a linear function of $\boldsymbol{h}(n)$ and $\boldsymbol{w}(n)$, the update equations can be derived using the LMS algorithm [42] by finding the partial derivatives of $e^2(n)$. For

the AEC filter $\boldsymbol{h}(n)$, we obtain

$$\begin{aligned}\boldsymbol{h}(n+1) &= \boldsymbol{h}(n) + \mu_h e(n) \nabla_{\boldsymbol{h}} \left\{ \boldsymbol{\theta}^T(n) \mathbf{S}(n) \boldsymbol{h}(n) \right\} \\ &= \boldsymbol{h}(n) + \mu_h e(n) \mathbf{S}^T(n) \boldsymbol{\theta}(n), \end{aligned} \tag{116}$$

where $\nabla_{\boldsymbol{h}}$ denotes the partial derivative over $\boldsymbol{h}$ and μ_h is the step size. Similarly, the update equation for the shortening filter $\boldsymbol{w}(n)$ is [cf. (114)]

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \mu_w e(n) \boldsymbol{y}(n), \tag{117}$$

where μ_w is the step size. Note that the idea of the shortening filter is similar to the one in [1]. The update equations (116) and (117) do not ensure stability unless a strong condition is imposed on the step sizes μ_h and μ_w. The optimum step size for the LMS algorithm that guarantees stability and fast convergence leads to the so-called NLMS algorithm [42]:

$$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \frac{\mu_h}{\|\mathbf{S}^T(n) \boldsymbol{\theta}(n)\|_2^2} e(n) \mathbf{S}^T(n) \boldsymbol{\theta}(n), \tag{118}$$

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \frac{\mu_w}{\|\boldsymbol{y}(n)\|_2^2} e(n) \boldsymbol{y}(n). \tag{119}$$

3.5.1.2 Adaptive Algorithm to Update the Nonlinear Filter

We introduce three schemes to update the nonlinear filter coefficients [84].

(1) *NLMS Adaptation.* Following a similar procedure as that developed for the coefficients of the linear filters, the update equation for the nonlinear coefficients $\boldsymbol{\theta}(n)$ can be derived as

$$\boldsymbol{\theta}(n+1) = \boldsymbol{\theta}(n) + \frac{\mu_\theta}{\|\mathbf{S}(n) \boldsymbol{h}(n)\|_2^2 + \delta} e(n) \mathbf{S}(n) \boldsymbol{h}(n), \tag{120}$$

where the regulation term δ is a small positive constant to avoid divergence at the initial stage when $\boldsymbol{h}(0) = \mathbf{0}$.

(2) *RLS Adaptation.* The adaptation of the nonlinear coefficients $\boldsymbol{\theta}(n)$ can also be performed by the RLS algorithm [101].

(3) *Coherence Adaptation.* An alternative method to update the nonlinear coefficients is based on the pseudo-MSC function, see [92] for details.

So far, we have introduced the adaptive schemes for both the linear and nonlinear parts. The entire step-by-step algorithm is summarized in Table 3.

Table 3: Adaptive algorithm for the nonlinear AEC with a CSE.

Update $\boldsymbol{\theta}(n+1)$ by nonlinearity identification algorithms
$e = d(n) - \boldsymbol{\theta}^T(n+1)\boldsymbol{u}(n)$
$\boldsymbol{h}(n+1) = \boldsymbol{h}(n) + \frac{\mu_h}{\|\mathbf{S}^T(n)\boldsymbol{w}(n))\|_2^2} e(n)\mathbf{S}^T(n)\boldsymbol{w}(n)$
$\boldsymbol{h}(n+1) = \frac{\boldsymbol{h}(n+1)}{\|\boldsymbol{h}(n+1)\|_2}$
$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \frac{\mu_w}{\|\boldsymbol{y}(n)\|_2^2} e(n)\boldsymbol{y}(n)$

3.5.2 Performance Analysis

Residual Echo Power

In this section, we analyze the residual echo power, which is an important figure of merit to measure the performance of AECs [84]. To separate the effect of the linear and nonlinear filters, we assume that there is no model mismatch of the loudspeaker nonlinearity, i.e., the nonlinearity in the loudspeaker can be modeled using $u(\cdot;\boldsymbol{\theta})$ with the perfect knowledge of $\boldsymbol{\theta}$. A block diagram for this analysis is given in Fig. 14. The LEMS consists of the loudspeaker nonlinearity $u(\cdot;\boldsymbol{\theta})$ and room impulse response $h_r(n)$. In the following, for any quantity ξ, $\hat{\xi}$ stands for its corresponding estimate.

Consider the background noise at the microphone, i.e., $y(n)$ is the corrupted version of $c(n)$ by the noise $v(n)$. Suppose the original room impulse response $h_r(n)$ has length L_o. Define vectors

$$\tilde{\boldsymbol{x}}(n) = [x(n), x(n-1), ..., x(n-L_w-L_o+2)]^T, \tag{121}$$

$$\boldsymbol{v}(n) = [v(n), v(n-1), ..., v(n-L_w+1)]^T, \tag{122}$$

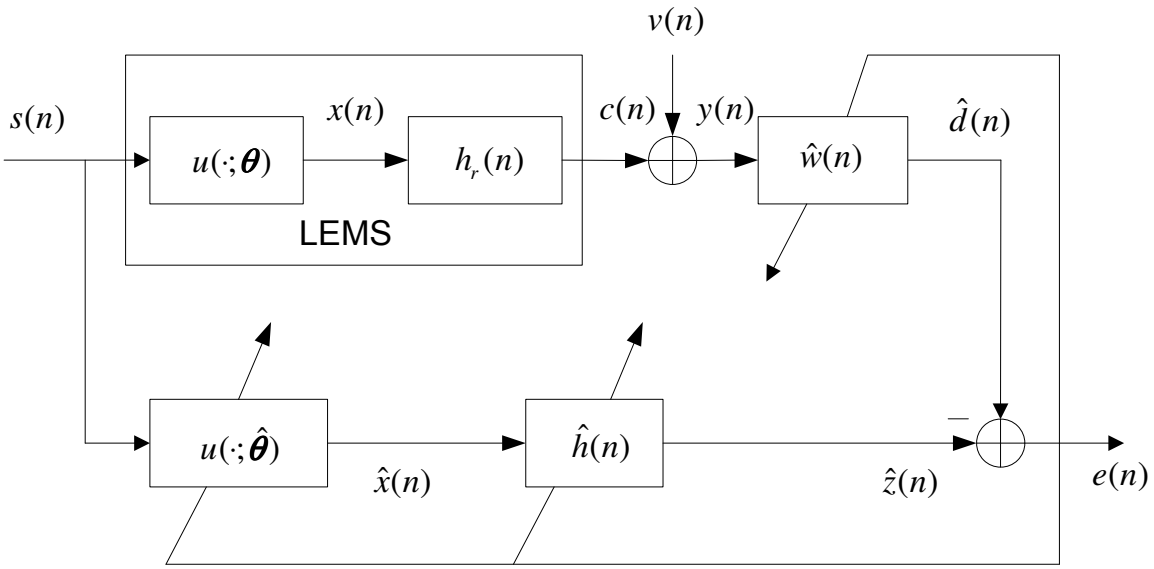


Figure 14: System structure for the performance analysis.

and a matrix

$$\mathbf{H}=\begin{bmatrix} h_r(0) & \cdots & h_r(L_o-1) & 0 & \cdots & 0 \\ 0 & h_r(0) & \cdots & h_r(L_o-1) & 0 & \cdots \\ \vdots & & & & & \vdots \\ 0 & \cdots & 0 & h_r(0) & \cdots & h_r(L_o-1) \end{bmatrix}. \tag{123}$$

Over a block of L_w output symbols, the input-output relationship of $h_r(n)$ can be cast in the matrix form:

$$\boldsymbol{y}(n) = \mathbf{H}\tilde{\boldsymbol{x}}(n) + \boldsymbol{v}(n). \tag{124}$$

Define the correlation matrices

$$\mathbf{R}_{xx} = E\left[\tilde{\boldsymbol{x}}(n)\tilde{\boldsymbol{x}}^T(n)\right], \tag{125}$$

$$\mathbf{R}_{yy} = E\left[\boldsymbol{y}(n)\boldsymbol{y}^T(n)\right], \tag{126}$$

$$\mathbf{R}_{vv} = E\left[\boldsymbol{v}(n)\boldsymbol{v}^T(n)\right], \tag{127}$$

$$\mathbf{R}_{xy} = \mathbf{R}_{yx}^T = E\left[\tilde{\boldsymbol{x}}(n)\boldsymbol{y}^T(n)\right]. \tag{128}$$

Based on the convergence analysis (see Appendix A), the nonlinear coefficients vector converges to its true value, i.e., $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$. Then, the optimal solution for $\hat{\boldsymbol{h}}$ based on (115) is the eigenvector of matrix the $\mathbf{R}_\Delta$ corresponding to the smallest eigenvalue [1]:

$$\mathbf{R}_\Delta \hat{\boldsymbol{h}} = \lambda_{\min}\hat{\boldsymbol{h}}, \tag{129}$$

and the corresponding optimal solution for the shortening filter $\boldsymbol{w}$ is [1]

$$\hat{\boldsymbol{w}} = \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\hat{\boldsymbol{h}}, \tag{130}$$

where

$$\Delta = L_o + L_w - 1 - L_h, \tag{131}$$

$$\mathbf{R}_{\Delta} = [\mathbf{I}_{L_h} \quad \mathbf{0}_{L_h \times \Delta}] \cdot \mathbf{R}_{x/y} \cdot [\mathbf{I}_{L_h} \quad \mathbf{0}_{L_h \times \Delta}]^T, \tag{132}$$

$$\mathbf{R}_{x/y} = \mathbf{R}_{xx} - \mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx} = \mathbf{R}_{xx}^{-1} + \mathbf{H}^T\mathbf{R}_{vv}^{-1}\mathbf{H}. \tag{133}$$

Define the "filtered" room impulse response as

$$\hat{g}(n) = h_r(n) * \hat{w}(n), \tag{134}$$

with its vector form as

$$\hat{\hat{\boldsymbol{g}}}(n) = [\hat{g}_0(n), \hat{g}_1(n), ..., \hat{g}_{L_o+L_w-2}(n)]^T. \tag{135}$$

Therefore, the residual echo signal can be written as

$$e_{res}(n) = x(n) * \hat{g}(n) - x(n) * \hat{h}(n). \tag{136}$$

The MMSE (i.e., residual echo power) can be obtained [1]:

$$E\left[e_{res}^2(n)\right] = E\left[\Big(x(n) * \hat{g}(n) - x(n) * \hat{h}(n)\Big)^2\right] = \lambda_{\min}. \tag{137}$$

Computational Complexity

Table 4 shows the computational complexity in terms of the number of multiplications and additions required per iteration by the proposed and existing algorithms. For all the algorithms, the auxiliary nonlinear block $u(\cdot;\boldsymbol{\theta})$ uses the same order K. For the algorithms without a shortening filter (NLMS, RLS, and MSC), the AEC filter $h(n)$ adopts the same length L_o as the original room impulse response. The computational complexity of these algorithms depends on K and L_o. For the algorithms with a shortening filter (NLMS-CSE, RLS-CSE, MSC-CSE), the computational complexity is given in terms of K, L_h, and L_w.

Usually, the order of the nonlinear model K is small, while the length of the room impulse response L_o can be several hundreds or even close to a thousand. The goal of the

Table 4: Computational complexity comparison of the proposed and existing methods.

Algorithms	Multiplications	Additions
NLMS	$2KL_o + 2L_o + 3K + 4$	$2KL_o + L_o + 2K - 2$
RLS	$2KL_o + 3K^2 + 2L_o + 5K + 2$	$2KL_o + 2K^2 + L_o + 2K - 1$
MSC	$2KL_o + 2(K+1)\log_2 N + 8K^2$ $+2L_o + 17K + 4$	$2KL_o + 3(K+1)\log_2 N + 8K^2$ $+L_o + 6K - 3$
NLMS-CSE	$2KL_h + 4L_h + 5K + 3L_w + 6$	$2KL_h + 2L_h + 3K + 3L_w - 6$
RLS-CSE	$2KL_h + 3K^2 + 4L_h + 7K + 3L_w + 4$	$2KL_h + 2K^2 + 2L_h + 3K + 3L_w - 5$
MSC-CSE	$2KL_h + 2(K+1)\log_2 N + 8K^2$ $+4L_h + 17K + L_w + 4$	$2KL_h + 3(K+1)\log_2 N + 8K^2$ $+2L_h + 6K + L_w - 5$

shortening filter is to use an FIR filter $w(n)$ with a short length to "squeeze" most of the room response power to a certain portion of the channel taps. Thus, the AEC filter $h(n)$ with a much smaller length L_h can be used to generate the echo signal. The dominant factor of computational complexity for an AEC without a shortening filter resides in the terms L_o and KL_o. When L_h and L_w are much smaller than L_o, the computational complexity of the proposed method is reduced considerably relative to the existing ones.

3.5.3 Simulations

Shortening Effect

In the simulations, the nonlinearity of the PA/loudspeaker obeys the same model as in Section 3.2.2. The room impulse response was generated by an FIR filter whose coefficients were obtained via

$$h_r(n) = \begin{cases} \beta(n)e^{-\alpha n}, & 4 \le n \le L_o \\ 0, & \text{otherwise} \end{cases} \tag{138}$$

with $\beta(n)$ following the standard normal distribution; $L_o = 300$ and $\alpha = 0.02$. We set the filter length $L_h = 100$ and $L_w = 300$, respectively. The far-end signal $s(n)$ was generated according to an i.i.d. Gaussian distribution. The signal $y(n)$ was generated with the SNR set at 30 dB.

Figure 15 shows the results of the NLMS-based algorithms with/without the shortening filter. It can be seen that the ERLE of the NLMS algorithm without the CSE can reach 29 dB, while the ERLE-CSE saturated at around 25 dB. This effect is ascribed to the residual

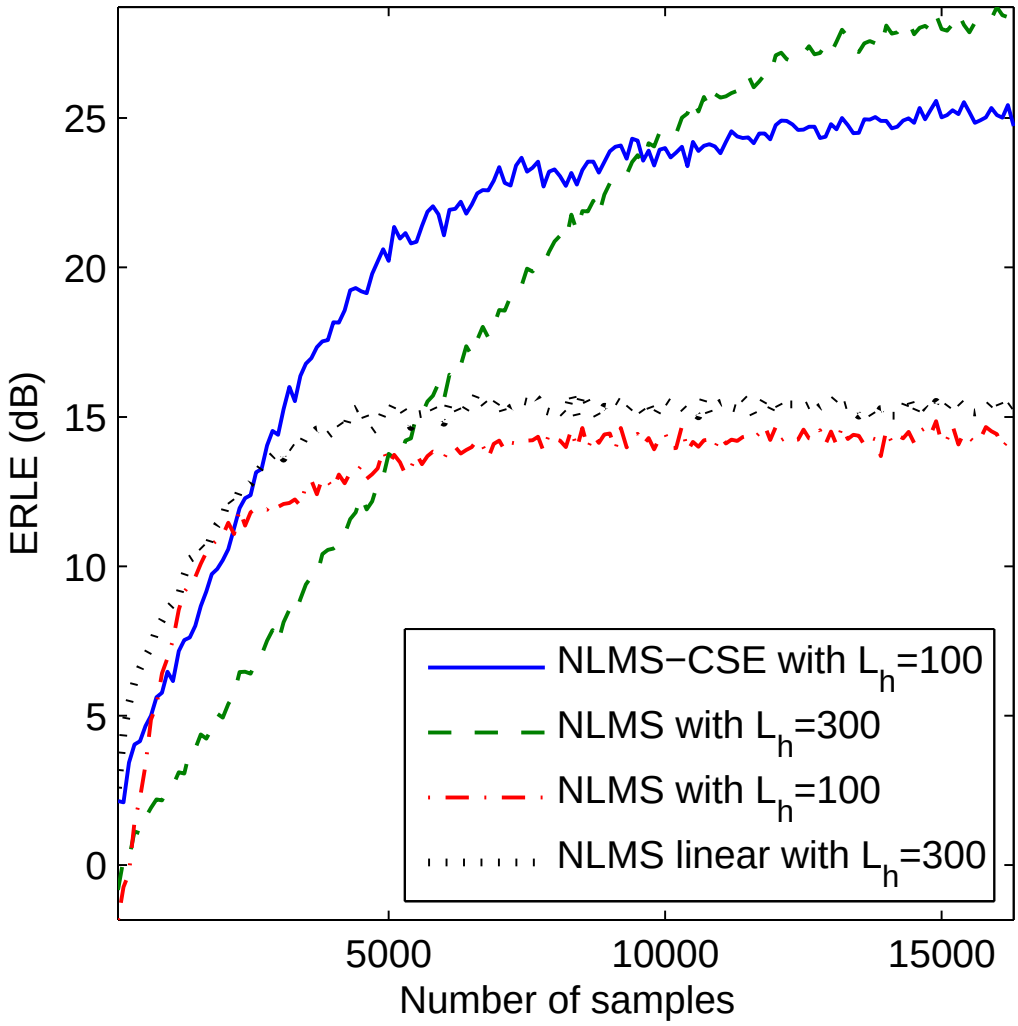


Figure 15: NLMS-based algorithms with/without the CSE.

error in (137). To analytically justify the 4dB performance loss of our proposed methods, we calculated the theoretical residual error power in (137) by finding the minimum eigenvalue of the matrix $\mathbf{R}_\Delta$ in (133). With the SNR of 30 dB, the theoretical results give us the maximum ERLE as 25.3 dB, which is consistent with the simulated result. Although the proposed method has a 4 dB loss in ERLE, it increases the convergence rate a lot while achieving considerably good echo cancellation performance. Moreover, the proposed method reduces the computational complexity significantly (see Table 4). In Fig. 15, we also show the ERLEs of the NLMS algorithm with a shorter filter length ("NLMS with $L_h = 100$")

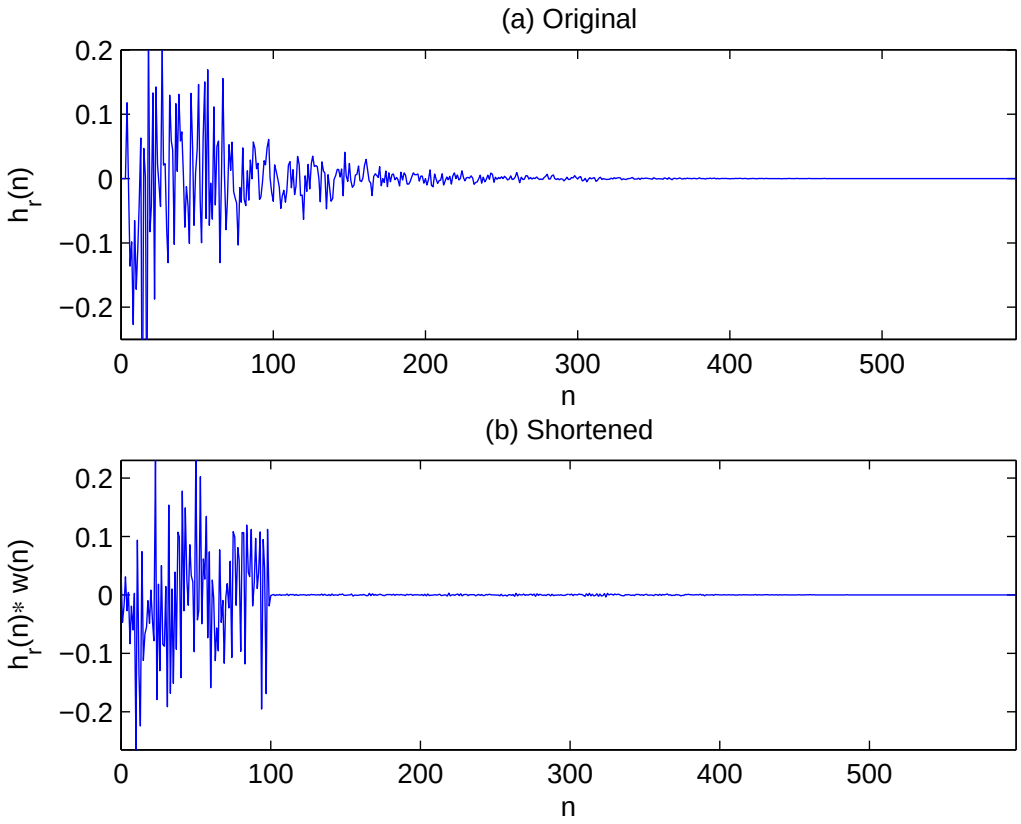


Figure 16: Impulse response: (a) The original room impulse response $h_r(n)$ has length $L_o = 300$. (b) In theory, $h_r(n) * w(n)$ would have length $L_o + L_w - 1 = 599$; the actual effective duration of $h_r(n) * w(n)$ is $L_h = 100$, illustrating the effect of the channel shortening filter.

and the traditional linear AEC ("NLMS linear with $L_h = 300$") as benchmarks. None of them can achieve reasonable performance. This simulation shows the effectiveness of the proposed method.

Figure 16 depicts the original and the shortened impulse response. We see that the method is quite successful in reducing the effective impulse response length. A more complete elimination of the tail of $h_r(n) * w(n)$ can be achieved with a longer AEC filter, but even with the given $h(n)$ of length $L_h = 100$, a fairly high ERLE is achieved (see Fig. 15).

Identification Performance

To quantitatively evaluate the system identification performance, misalignment is taken as a figure of merit. For the linear part misalignment, we use the distance measure defined

as

$$D_h(\text{dB}) = 10\log_{10}\frac{\|\hat{\boldsymbol{g}} - \hat{\boldsymbol{h}}\|_2^2}{\|\hat{\boldsymbol{g}}\|_2^2}. \tag{139}$$

Recall $\hat{\boldsymbol{g}}$ is the first part of the shortened room impulse response. For the nonlinear part misalignment, we adopt the distance measure as

$$D_\theta(\text{dB}) = 10\log_{10}\frac{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2^2}{\|\boldsymbol{\theta}\|_2^2}. \tag{140}$$

For the proposed methods, D_h and D_θ are calculated when the iterative process is terminated and at SNR = 30 dB (see Table 5). The results show that the estimates of the nonlinear coefficients converge to the true values since the misalignment D_θ is very small. The results also illustrate the effectiveness of the shortening filter since the residual energy is negligible relative to the energy of the dominant part after shortening.

Table 5: D_h(dB) and D_θ(dB) of different methods.

	NLMS-CSE	RLS-CSE	MSC-CSE
D_h(dB)	-42.1	-42.8	-43.4
D_θ(dB)	-44.9	-45.5	-45.8

Furthermore, we evaluate the performance of the proposed method for different ratios of the room impulse response length L_o with respect to the AEC filter length L_h and shortening filter length L_w. First, we fix $L_o = 300$ and $L_w = 300$. ERLEs with respect to different L_o/L_h are shown in Table 6. It can be seen that decreasing L_h degrades the echo cancellation performance due to the increase of the uncompensated residual error. Second, ERLEs for different L_o/L_w are shown in Table 7 with $L_o = 300$ and $L_h = 100$. It can be observed that larger L_w achieves better shortening performance, thus leads to better ERLE performance, however more computational burden is incurred.

Convergence in Long Room Impulse Response

We evaluate the performance in more realistic scenarios, where a long room impulse response with length 1024 was generated using the IMAGE method as in Section 3.2.2. To compensate for the ERLE loss of the proposed method, we take advantage of the residual echo suppressor (RES) after the echo canceller. For demonstration purposes, we implement

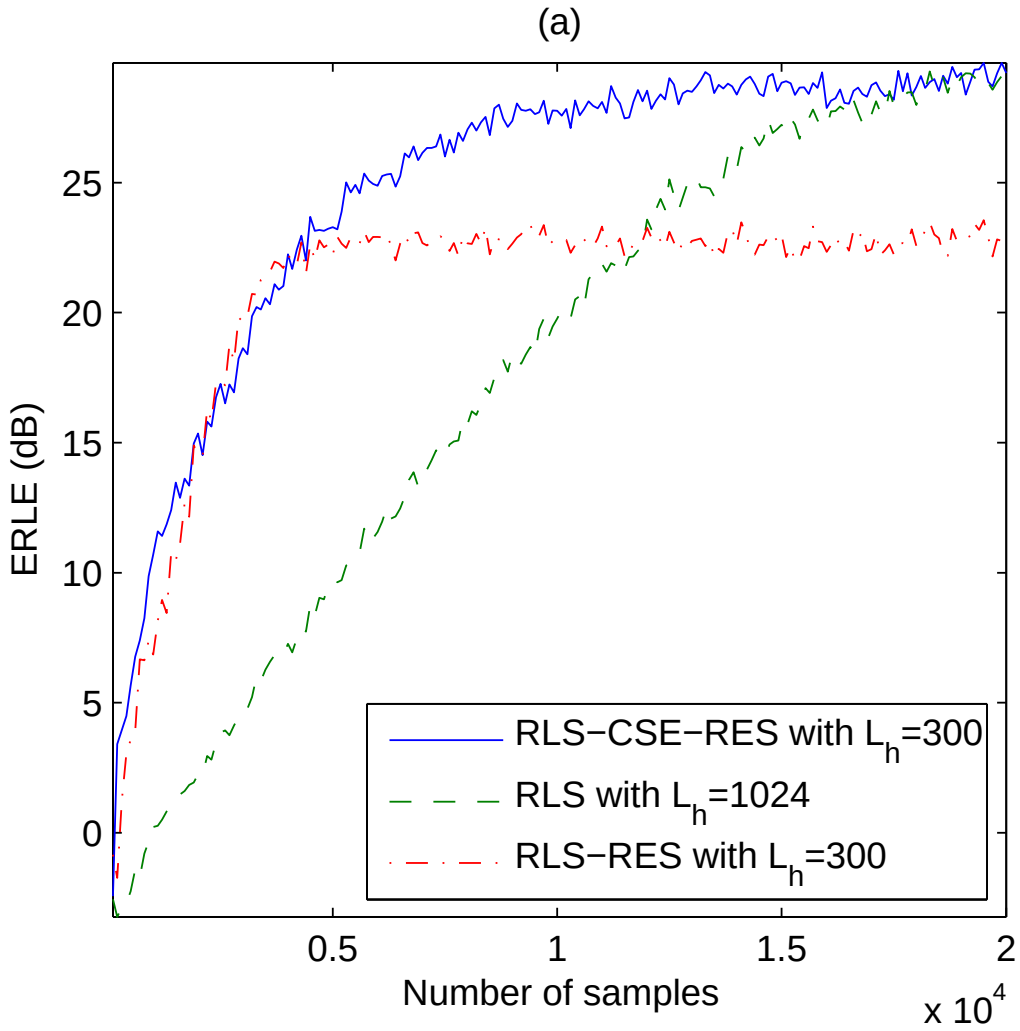
Table 6: ERLE(dB) for different L_o/L_h.

Criterion \ L_o/L_h	1	2	3	4	5	6
NLMS-CSE	25.4	25.2	25.2	22.1	21.4	20.6
RLS-CSE	25.8	25.5	25.3	22.3	21.4	20.8
MSC-CSE	25.6	25.5	25.4	22.7	21.6	20.8

Table 7: ERLE(dB) for different L_o/L_w.

Criterion \ L_o/L_w	0.3	0.5	1	2	3	4	5
NLMS-CSE	27.4	25.7	25.2	21.2	19.6	18.9	18.0
RLS-CSE	28.0	25.9	25.3	21.2	19.8	18.8	18.1
MSC-CSE	27.4	26.3	25.4	21.5	19.8	18.7	18.2

a RES based on [39] by designing a gain filter in the frequency domain. Figure 17 (a) shows the ERLE with an i.i.d. Gaussian far-end signal. It can be seen that the proposed method performs well in the very long impulse response environment and converges much faster than the one without the CSE but using a long linear filter. The performance of the proposed methods is also justified using a real speech signal, which is illustrated in Fig. 17 (b). It is shown that our proposed methods outperform the existing ones.


(a)
ERLE (dB)
25
20
15
10
5
0
0.5
1
1.5
2
Number of samples
x 10^4
RLS−CSE−RES with L_h=300
RLS with L_h=1024
RLS−RES with L_h=300

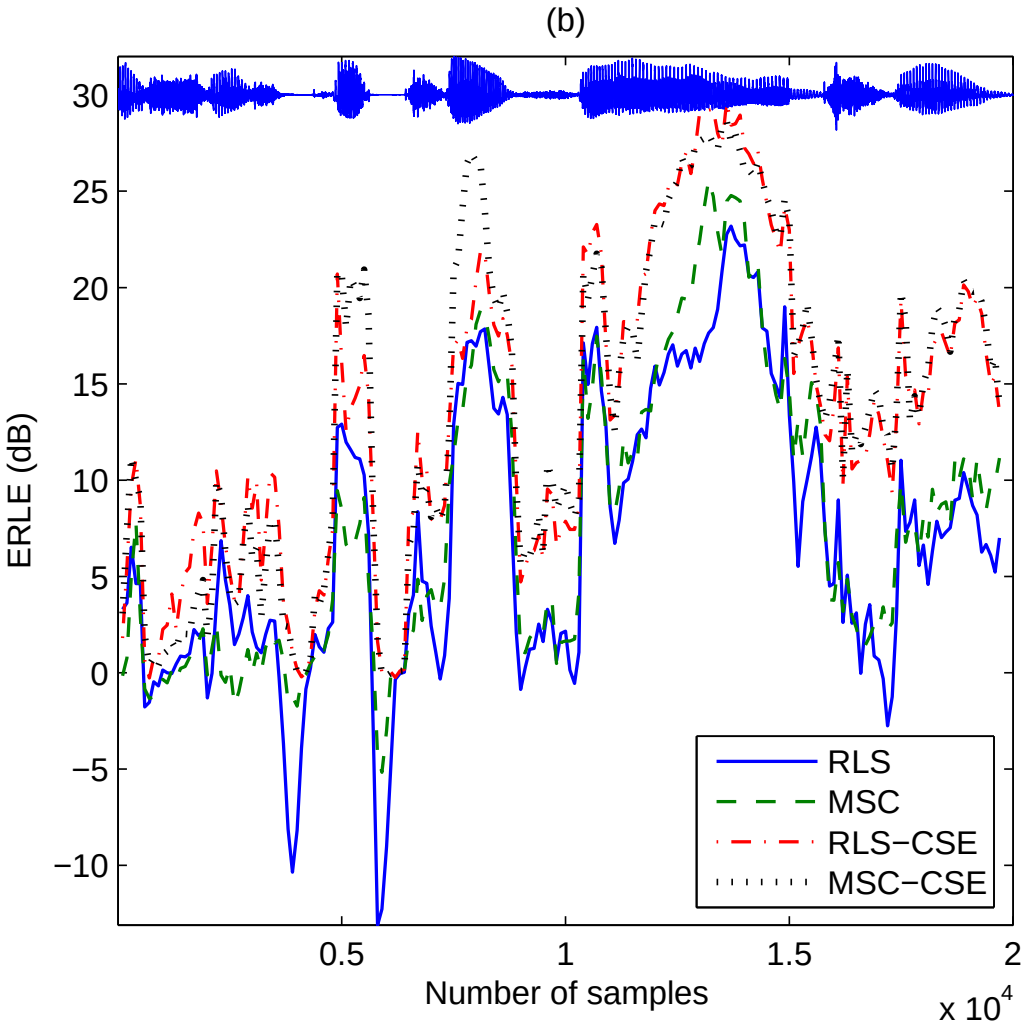


Figure 17: AEC algorithms with/without the CSE using a long room impulse response: (a) i.i.d. Gaussian signal; (b) real speech data.

3.6 Cascade NAEC in the Presence of Multiple Nonlinearities

By considering the memoryless nonlinearity only from the PA/loudspeaker, the LEMS can be well represented by the Hammerstein model. In this section, we take into account the nonlinearities in both the loudspeaker and the microphone, in which case the LEMS can be described by the Hammerstein-Wiener model. Thus, the NAEC design is reduced to the Hammerstein-Wiener system identification. Numerous Hammerstein-Wiener system identification algorithms have been proposed in the literature. In [4], an identification scheme for single-input single-output (SISO) Hammerstein-Wiener systems was developed. A very specific model structure was assumed in [4] which limits its practical applicability. Building upon [4], a more general blind identification technique for SISO systems was proposed in [5]. An iterative method was developed in [112], and a linear subspace intersection algorithm was extended in [31] for the identification of Hammerstein-Wiener systems. However, to the best of our knowledge, none of the existing Hammerstein-Wiener system identification methods are suitable for the NAEC problem on hand, because (1) they are nonadaptive and thus cannot be readily applied to a real-time echo canceller design, and (2) they incur large computational load due to the presence of a long room impulse response.

3.6.1 System Structure and Nonlinearities Identification

We propose a new structure for the NAEC design as shown in Fig. 18 [88]. The adaptive NAEC consists of three blocks. The nonlinear block $\widetilde{g}^{-1}(\cdot;\boldsymbol{\beta})$ models the inverse of the microphone nonlinearity. Thus, the concatenation of the LEMS system with $\widetilde{g}^{-1}(\cdot;\boldsymbol{\beta})$ yields a Hammerstein system. For echo cancellation, we use a nonlinear block $\tilde{f}(\cdot;\boldsymbol{\alpha})$ and an FIR filter $h(n)$ to model the loudspeaker nonlinearity and the room impulse response, respectively. Note that both the memoryless nonlinearity functions $\tilde{f}(\cdot;\boldsymbol{\alpha})$ and $\widetilde{g}^{-1}(\cdot;\boldsymbol{\beta})$ are approximated by a linear combination of nonlinear basis functions. As usual, the goal of the NAEC is to minimize the power of the residual echo signal:

$$e(n) = y(n) - z(n) = \tilde{g}^{-1}(r(n);\boldsymbol{\beta}) - \tilde{f}(s(n);\boldsymbol{\alpha}) * h(n). \tag{141}$$

Next, we introduce a two-stage method for the NAEC design. First, the nonlinearities $\tilde{g}^{-1}(\cdot;\boldsymbol{\beta})$ and $\tilde{f}(\cdot;\boldsymbol{\alpha})$ are identified using the pseudo-MSC function-based method. Afterwards, $h(n)$ can be estimated using the NLMS algorithm.

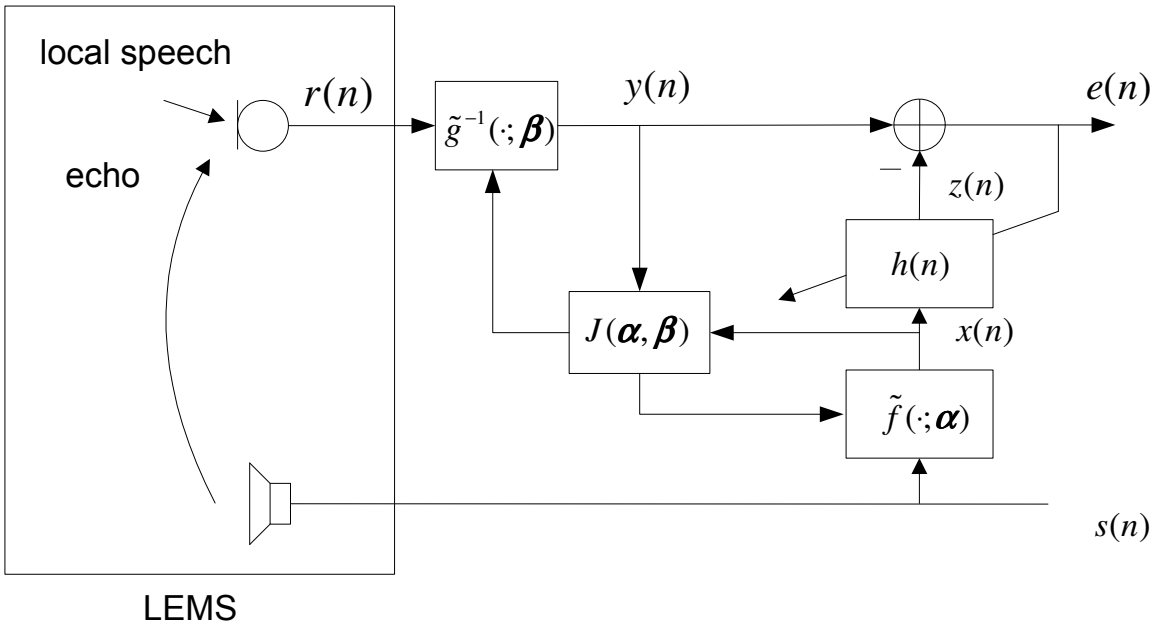


Figure 18: AEC with multiple nonlinearities.

Define vectors

$$\boldsymbol{f}(n) = \left[f_1(s(n)), f_2(s(n)), ..., f_{K_f}(s(n))\right], \tag{142}$$

$$\boldsymbol{g}(n) = \left[g_1(r(n)), g_2(r(n)), ..., g_{K_g}(r(n))\right], \tag{143}$$

where K_f and K_g are nonlinear orders. Thus, the output signals of the nonlinear modules are obtained

$$x(n;\boldsymbol{\alpha}) = \tilde{f}(s(n);\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T\boldsymbol{f}(n), \tag{144}$$

$$y(n;\boldsymbol{\beta}) = \widetilde{g}^{-1}(r(n);\boldsymbol{\beta}) = \boldsymbol{\beta}^T\boldsymbol{g}(n). \tag{145}$$

If $\tilde{f}(\cdot;\boldsymbol{\alpha})$ is a perfect match to $f(\cdot)$ and $\widetilde{g}^{-1}(\cdot;\boldsymbol{\beta})$ is the inverse of $g(\cdot)$ up to a scalar, then the processes $x(n)$ and $y(n)$ will be perfectly linearly related. Since the pseudo-MSC function-based metric $\int_{-0.5}^{0.5}\tilde{C}_{xy}(f)\mathrm{d}f$ provides a means for quantifying the linear association between two stationary random processes, we propose to solve for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the

nonlinear blocks as follows:

$$\left[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}\right] = \arg \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} J(\boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{146}$$

where

$$J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \int_{-0.5}^{0.5} \hat{\hat{C}}_{xy}(f; \boldsymbol{\alpha}, \boldsymbol{\beta}) \mathrm{d}f. \tag{147}$$

Because the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ depend on each other, globally searching for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ incurs high computational complexity. However, given one of the unknown parameters, for instance $\boldsymbol{\beta}$, we can form the signal $y(n)$ according to (145). Thus, the objective function given the parameter vector $\boldsymbol{\beta}$ can be reduced to

$$J(\boldsymbol{\alpha}|\boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^T \mathbf{R}_1 \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{R}_2 \boldsymbol{\alpha}}, \tag{148}$$

where

$$\mathbf{R}_1 = \int_{-0.5}^{0.5} S_{yy}^{-1}(f) \boldsymbol{s}_{y\boldsymbol{f}}(f) \boldsymbol{s}_{y\boldsymbol{f}}^H(f) \mathrm{d}f, \tag{149}$$

$$\mathbf{R}_2 = E\left[\boldsymbol{f}(n) \boldsymbol{f}^T(n)\right], \tag{150}$$

and $y(n)$ is formed given the current $\boldsymbol{\beta}$. A similar form holds for $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$. Therefore, the objective function (147) is a generalized Rayleigh's quotient in $\boldsymbol{\alpha}$ for given $\boldsymbol{\beta}$ and *vice versa*. An alternating parameter estimation procedure is then the following relaxation algorithm [103]:

$$\hat{\boldsymbol{\alpha}}(k) = \arg \max_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}(k-1)), \tag{151}$$

$$\hat{\boldsymbol{\beta}}(k) = \arg \max_{\boldsymbol{\beta}} J(\hat{\boldsymbol{\alpha}}(k), \boldsymbol{\beta}). \tag{152}$$

An adaptive algorithm was also developed in [92] to update the parameter $\boldsymbol{\theta}$ (which can be $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$):

$$\boldsymbol{\theta}(n) = \frac{\boldsymbol{\theta}^T(n-1) \mathbf{R}_2(n) \boldsymbol{\theta}(n-1)}{\boldsymbol{\theta}^T(n-1) \mathbf{R}_1(n) \boldsymbol{\theta}(n-1)} \mathbf{R}_2^{-1}(n) \mathbf{R}_1(n) \boldsymbol{\theta}(n-1). \tag{153}$$

The proposed iterative method for identifying the nonlinear parameters is summarized in Table 8, where L denotes the data segment length. Once the two nonlinear blocks have been identified, the linear block can be found via the least squares method. We point out that the convergence of the proposed iterative method is not guaranteed [103].

However, good initialization usually leads to convergence, which has been demonstrated by simulations. Note that the proposed method decouples the identification of the linear part from the nonlinear part, since the pseudo-MSC function is insensitive to the presence of an unknown linear block [92]. This feature is desirable for the NAEC problem, in which case the length of the room impulse response has no effect on the computational complexity for the nonlinearity identification.

Table 8: Iterative method to estimate parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ in the nonlinear blocks.

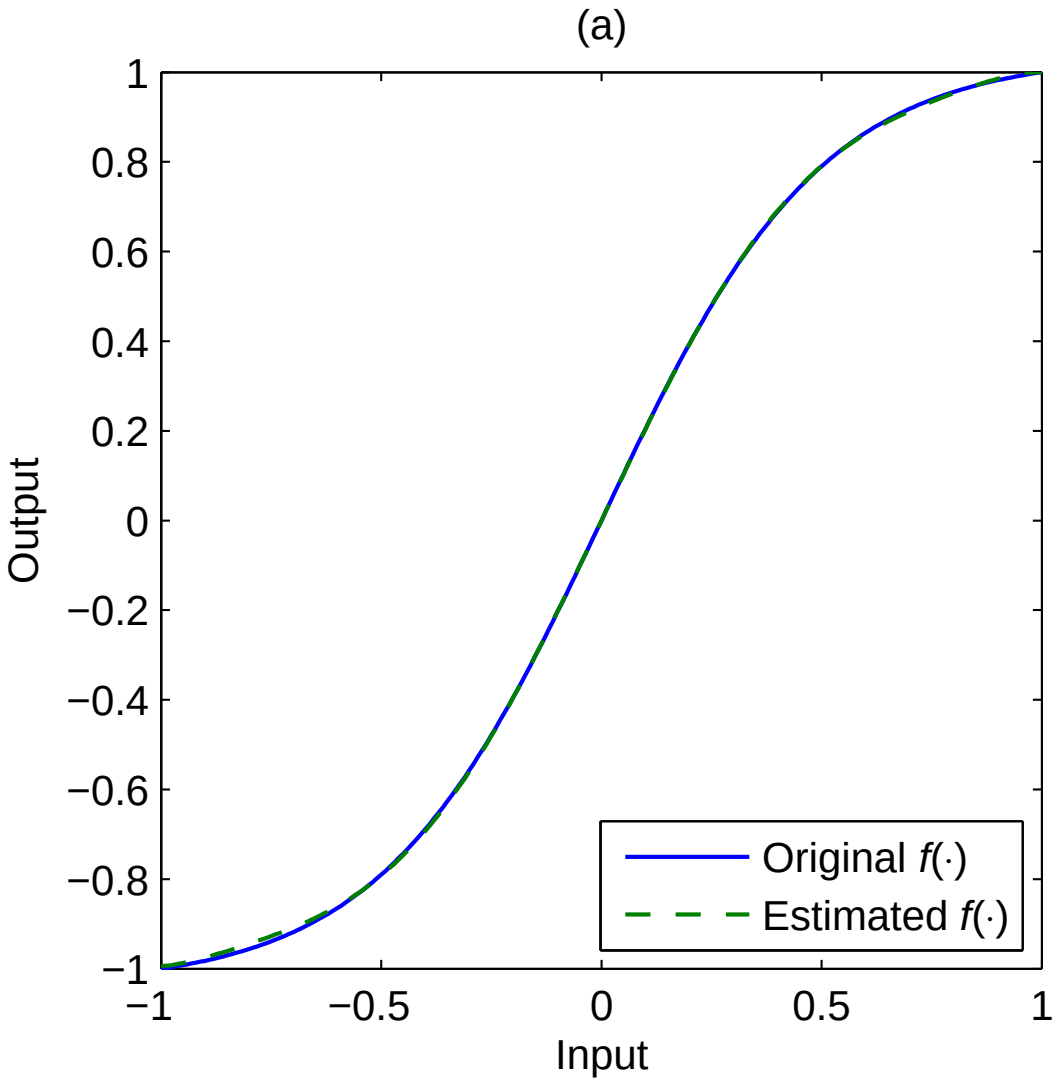
Initialize $\boldsymbol{\alpha}(0)$ and $\boldsymbol{\beta}(0)$.
for $k = 0, 1, ...$ **do**
  All $n \in [kL,\ (k+1)L)$: update $y(n)$ using (145) based on $\boldsymbol{\beta}(k)$.
  update $\boldsymbol{\alpha}(k+1)$ using (153).
  All $n \in [kL,\ (k+1)L)$: update $x(n)$ using (144) based on $\boldsymbol{\alpha}(k+1)$.
  update $\boldsymbol{\beta}(k+1)$ using (153).
end for

3.6.2 Simulations

In the simulations for the nonlinearity identification, the source signal $s(n)$ was generated according to an i.i.d. Gaussian distribution. Both the loudspeaker and microphone nonlinearities obey the hyperbolic tangent function. We approximate the nonlinear functions $f(\cdot)$ and $g^{-1}(\cdot)$ with the polynomial bases and the corresponding orders are $K_f = K_g = 7$. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are initialized such that $x(n) = s(n)$ and $y(n) = r(n)$, respectively.

Figure 19 (a) and (b) show the performance of the nonlinearity identification. Figure 19 (a) shows the loudspeaker nonlinearity $f(\cdot)$ and its estimate $\tilde{f}(\cdot)$; it can be seen that the estimate approximates well the nonlinearity $f(\cdot)$. Figure 19 (b) shows the microphone nonlinearity $g(\cdot)$ and the estimate of its inverse $\widetilde{g}^{-1}(\cdot)$, as well as the concatenated system consisting of $g(\cdot)$ followed by the nonlinear block $\widetilde{g}^{-1}(\cdot)$, i.e., $\widetilde{g}^{-1}(g(\cdot))$, which approximates a linear characteristic.

In Fig. 20, we show the estimate of the objective function in (147) as a function of the number of iterations. It can be seen that $J(\boldsymbol{\alpha}, \boldsymbol{\beta})$ approaches one as the number of iterations



increases. This implies that the two signals $x(n)$ and $y(n)$ are increasingly linearly related, indicating that $\tilde{f}(\cdot)$ and $\widetilde{g}^{-1}(\cdot)$ approach $f(\cdot)$ and $g^{-1}(\cdot)$, respectively, when a sufficient number of samples are available.

In the scenarios of the echo cancellation problem, ERLE [60] is used to measure the performance of the proposed NAEC. The microphone-received signal $r(n)$ was generated under the single-talk scenario with the SNR set at 30 dB. Figure 21 (a) and (b) show the ERLEs for the NAEC with respectively, noise and speech signal as the system input; both demonstrate the effectiveness of the proposed nonlinear echo cancellation algorithm.

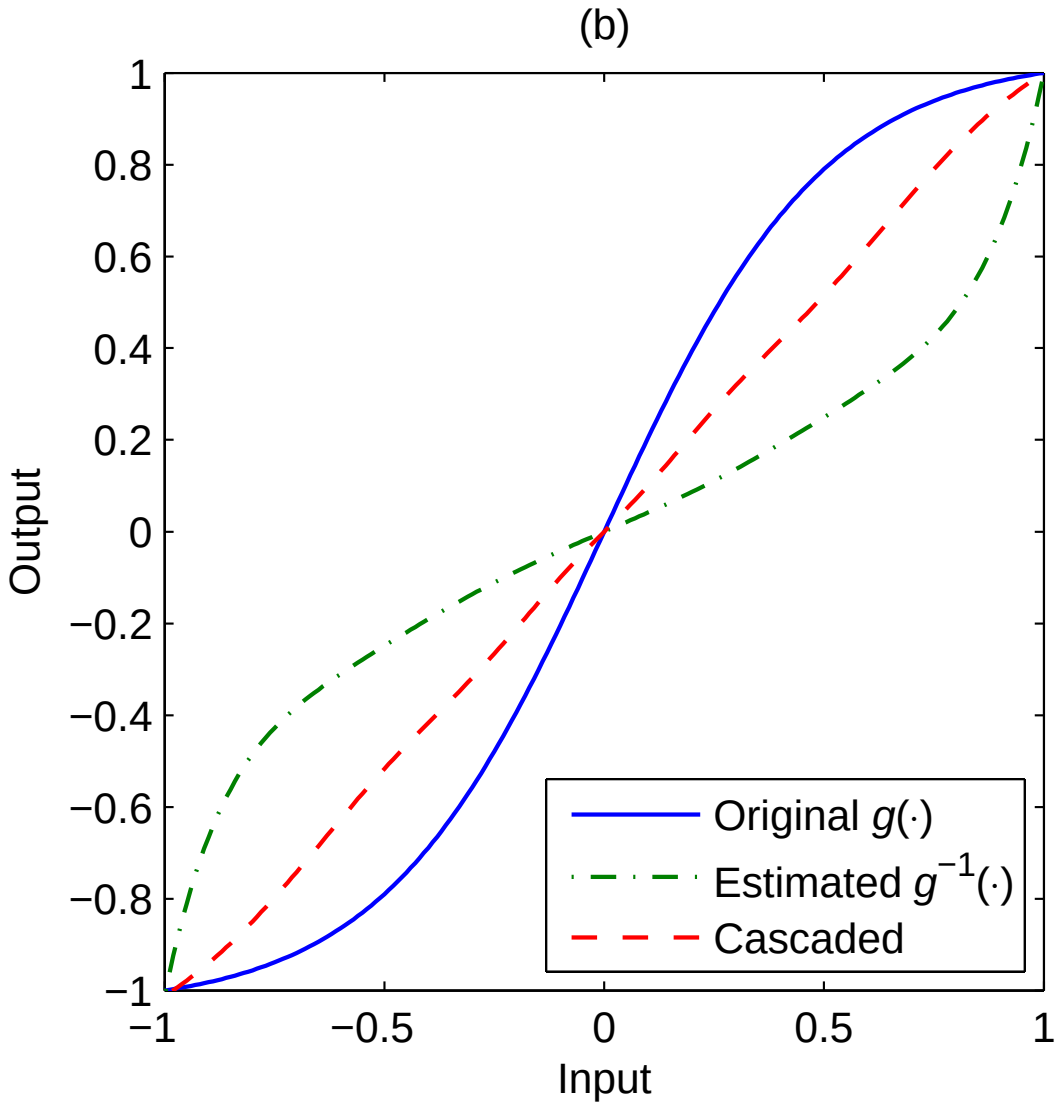


Figure 19: Nonlinearity identification: (a) loudspeaker nonlinearity $f(\cdot)$; (b) inverse of microphone nonlinearity $g^{-1}(\cdot)$.

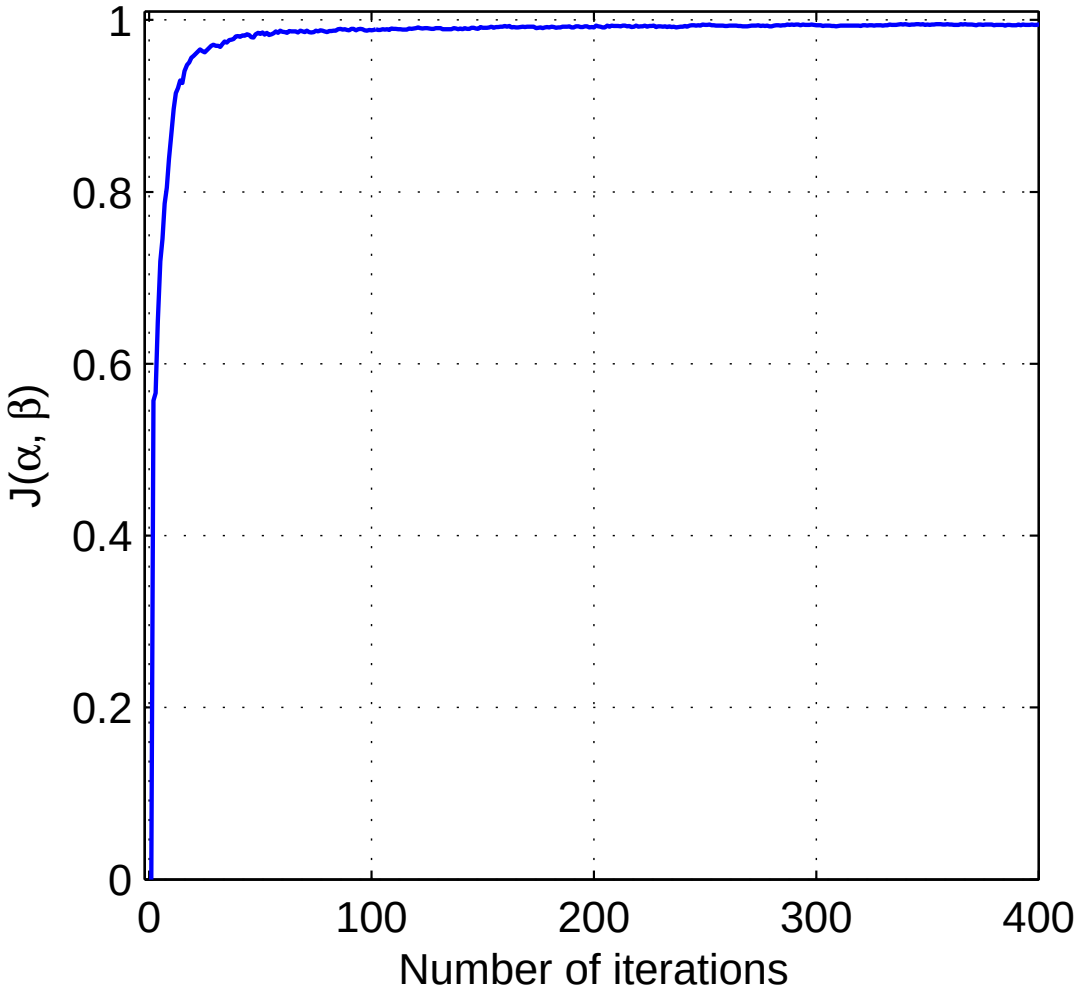


Figure 20: The objective function J approaches one as the number of iterations increases.

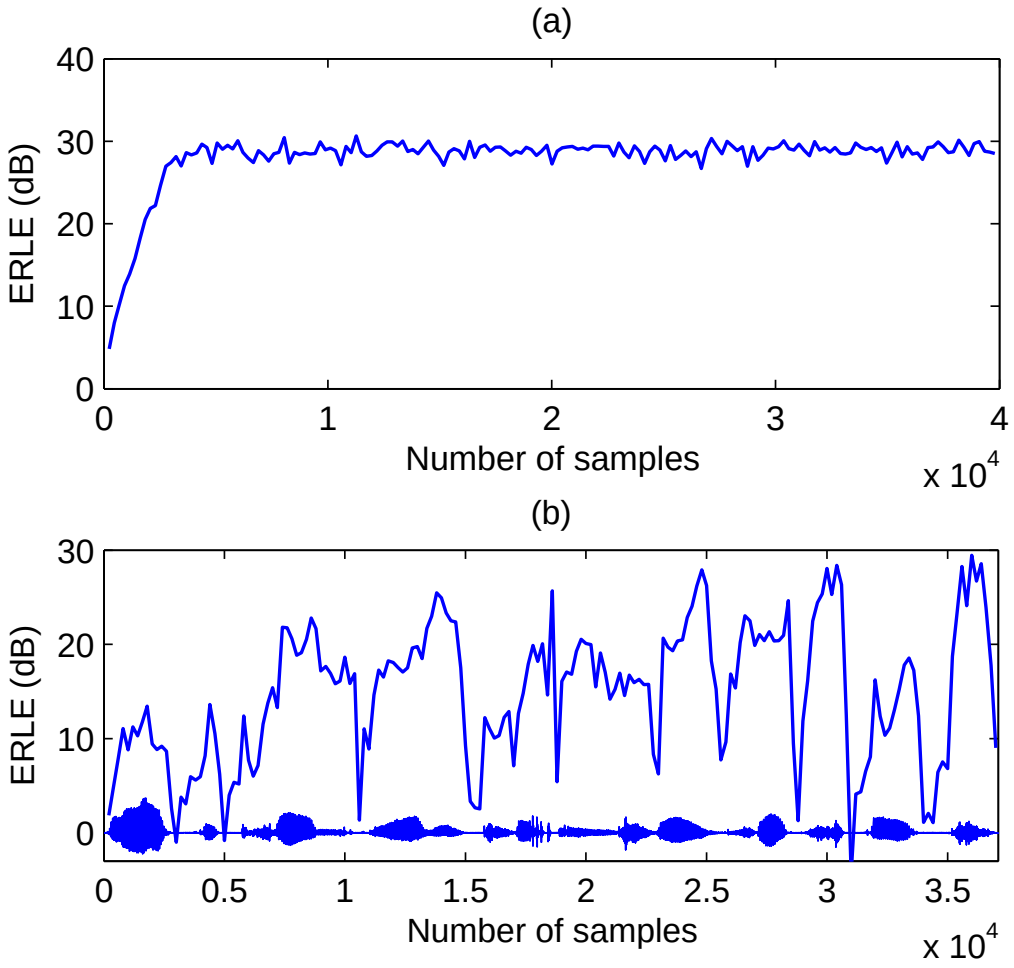


Figure 21: Performance of the nonlinear acoustic echo cancellation: (a) with an i.i.d. Gaussian signal as input; (b) with a speech signal as input.

CHAPTER IV

INTERFERENCE-ROBUST ACOUSTIC ECHO CANCELLATION

In this chapter, we focus on interference-robust echo cancellation algorithms. Double-talk detection and learning-rate adjustment are the control logic design scheme in response to the occurrence so double-talk and ambient noise, respectively. First, we design double-talk detectors (DTDs) by considering nonlinearities in the acoustic echo path. We propose to use mutual information as a decision statistic and show that it can be applied to both monophonic (Section 4.1) and stereophonic (Section 4.2) systems. Later on, we investigate the learning-rate control of the least mean square (LMS) algorithm. Specifically, we investigate a variable step size and variable tap length LMS algorithm under the assumption on an exponential decay envelope of the channel impulse response (Section 4.3).

4.1 DTD Using Mutual Information for Monophonic NAECs

The DTD design is a challenging task since there is no universal rule to discriminate between the echo signal and the near-end speech [11]. Typically, in a DTD, a decision statistic is formulated based on the available signals or signal estimates and compared with a threshold to determine whether a double-talk occurs. As discussed in Section 2.3.1, the cross-correlation-based DTD techniques [110, 29, 7, 28] have been proposed that appear to be successful for AEC applications. However, to the best of our knowledge, DTDs have not been proposed in conjunction with nonlinear AECs. The correlation-based criterion captures only the linear relationship between two random processes [110, 29, 7, 28]. Although [9] derives an optimum log-likelihood ratio test (LRT), the Gaussian assumption of signals does not hold any more when nonlinearity is present. Based on our experience, these schemes do not perform well when the acoustic echo path is nonlinear. Thus, we are motivated to seek a DTD for nonlinear AEC applications. Mutual information (MI) is in many ways the cornerstone of classic information theory, playing central roles in the analysis of both digital and analog communication systems [19]. The primary objective of DTD designs is to detect the

presence of the near-end speech. In this section, we show how the MI is suitable for this task.

4.1.1 Mutual Information (MI) and Its Calculation

We start with the fundamentals of MI. Denote continuous-valued random variables x and y by the pair (x, y). The entropy or uncertainty of the variable x is defined in terms of its probability density function (PDF) $f(x)$:

$$H(x) = -\int_x f(x) \log f(x) \mathrm{d}x. \tag{154}$$

After having observed y, the uncertainty of x is given by the conditional entropy, defined in terms of the conditional PDF $f(x|y)$ and the joint PDF $f(x, y)$:

$$H(x|y) = -\int_x \int_y f(x, y) \log f(x|y) \mathrm{d}y \mathrm{d}x. \tag{155}$$

The MI between x and y is defined as [19]

$$\begin{aligned} I(x; y) &= H(x) - H(x|y) \\ &= -\int_x f(x) \log f(x) \mathrm{d}x + \int_x \int_y f(x, y) \log f(x|y) \mathrm{d}y \mathrm{d}x, \end{aligned} \tag{156}$$

and measures the reduction in the uncertainty of x due to the knowledge of y.

Another view of MI is that it measures the degree to which x and y are not independent. With the identity $f(x, y) = f(x|y)f(y)$, the expression in (156) can be rewritten as

$$I(x; y) = \int_x \int_y f(x, y) \log \frac{f(x, y)}{f(x)f(y)} \mathrm{d}x \mathrm{d}y. \tag{157}$$

When x and y are statistically independent, $f(x, y) = f(x)f(y)$, and thus $I(x; y) = 0$. The value of $I(x; y)$ grows as x and y become more dependent. The more dependent x is on y, the more information one gains about x once y is known, and therefore the less uncertain x is when y is known. Moreover, MI is equivalent to the Kullback-Leibler distance between the joint distribution $f(x, y)$ and the product of the marginal distributions $f(x)$ and $f(y)$. The following properties hold for MI [19].

Property I: $0 \leq I(x; y) \leq \infty$.

Property II: $I(x; y) = 0$ if and only if x and y are statistically independent.

Property III: $I(x;y) = \infty$ if and only if y is a function of x, i.e., $y = g(x)$, where $g(\cdot)$ is invertible.

Property IV: $I(x;y) = I(u;v)$ if the transformation $(x,y) \to (u,v)$ has the form $u = g(x)$, $v = p(y)$ with $g(\cdot)$ and $p(\cdot)$ being one-to-one mappings.

In order to calculate the MI between x and y, we need to estimate the joint distribution $f(x,y)$. Histogram and Kernel methods are widely used to estimate MI but entail high computational complexity [82]. To reduce the complexity in these methods, we adopt a recent estimator that estimates entropy from the average distance to the k-nearest neighbors [58]. Consider a set of N input-output pairs $z_i = (x_i, y_i),\ i = 1, ..., N$, and the maximum norm [58]

$$\|z - z'\|_\infty = \max\{|x - x'|, |y - y'|\}, \tag{158}$$

for a fixed positive integer k, we find $z_{k(i)} = (x_{k(i)}, y_{k(i)})$ as the k-th nearest neighbor of z_i according to the maximum norm. Define the following distances

$$\begin{aligned} \epsilon_i/2 &= \|z_i - z_{k(i)}\|_\infty, \\ \epsilon_i^x/2 = |x_i - x_{k(i)}|,\ \epsilon_i^y/2 &= |y_i - y_{k(i)}|. \end{aligned} \tag{159}$$

$\epsilon_i/2$ is the distance from z_i to its k-th neighbor. $\epsilon_i^x/2$ and $\epsilon_i^y/2$ are the distances between the same points projected onto x and y subspaces. Let n_i^x and n_i^y be the numbers of sample points that satisfy

$$|x_i - x_j| \le \epsilon_i^x/2 \quad \text{and} \quad |y_i - y_j| \le \epsilon_i^y/2. \tag{160}$$

The estimator of the MI between x and y is then obtained:

$$\hat{I}(x;y) = \psi(k) - \frac{1}{k} - \frac{1}{N}\sum_{i=1}^{N}[\psi(n_i^x) + \psi(n_i^y)] + \psi(N), \tag{161}$$

where $\psi(\cdot)$ is the Digamma function.

4.1.2 A Test Statistic Based on MI

Consider an AEC system with a DTD as illustrated in Fig. 4. The signal $u(n)$ is the output of the far-end speech $x(n)$ at the near-end loudspeaker, causing an echo signal $c(n)$ at

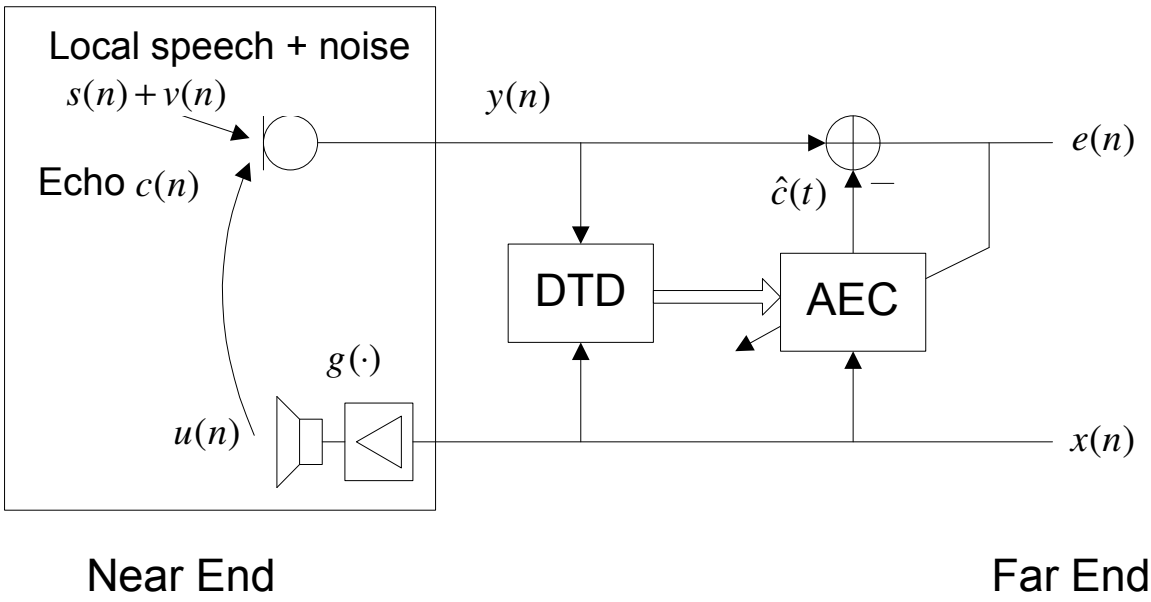


Figure 22: Block diagram of a voice communication system with an AEC and a DTD.

the microphone. The microphone-received signal $y(n)$ is composed of the echo signal $c(n)$, background noise $v(n)$ and near-end speech $s(n)$ (if any). A nonlinearity $g(\cdot)$ is included in the echo path, which may be caused by a non-ideal power amplifier (PA) or loudspeaker. For simplicity, we assume that the PA/loudspeaker is a memoryless nonlinear system. The room impulse response is modeled by a finite impulse response (FIR) filter $h(n)$ and is tracked by the AEC.

Denote the output of the nonlinear block by

$$u(n) = g(x(n)). \tag{162}$$

Define vectors

$$\boldsymbol{u}(n) = [u(n), u(n-1), ..., u(n-L+1)]^T, \tag{163}$$

$$\boldsymbol{h}(n) = [h_0(n), h_1(n), ..., h_{L-1}(n)]^T, \tag{164}$$

where L is the length of the acoustic echo path. Thus, the microphone-received signal can be written as

$$y(n) = \boldsymbol{h}^T(n)\boldsymbol{u}(n) + s(n), \tag{165}$$

and a noise-free scenario ($v(n) = 0$) is considered here. The DTD employs two signals $x(n)$ and $y(n)$ to make the decision on whether the near-end speech $s(n)$ is present or not. The

idea here is to form a statistic ξ and compare it with a preset threshold T. Once double-talk is declared, the AEC filter adaptation is disabled.

To design the decision statistic ξ, we first consider the linear case ($u = g(x) = \alpha x$, α is a non-zero constant). If $s(n)$ is absent then $y(n) = h(n) * g(x(n))$, i.e., $y(n)$ fully depends on $x(n)$. Hence, to determine whether the signal $s(n)$ is present or not is equivalent to measuring the degree of dependency between $x(n)$ and $y(n)$. From the Property III of MI, $I(x; y)$ achieves the maximum when x and y are fully dependent. If we treat sequences $x(n)$ and $y(n)$ as the realizations of random variables x and y, respectively, the presence of the near-end speech reduces $I(x; y)$. Therefore, we propose to use as the decision statistic, the MI between x and y [89]

$$\xi = I(x; y), \tag{166}$$

and formulate a binary hypothesis test:

$$H_0 : \text{ if } s(n) = 0 \text{ (double talk is absent)}, \ \xi \geq T, \tag{167}$$

$$H_1 : \text{ if } s(n) \neq 0 \text{ (double talk is present)}, \ \xi < T. \tag{168}$$

When the echo path is nonlinear ($u = g(x)$ is an invertible nonlinear mapping of x), we know from the property IV of MI that the MI between x and y is the same as that between u and y:

$$I(x; y) = I(g(x); y) = I(u; y). \tag{169}$$

Therefore, the MI-based DTD in (187) and (188) still works in the presence of the memoryless nonlinearity, provided that the nonlinear mapping is one-to-one. It is worth pointing out that Shannon's mutual information is a classical measure of statistical dependence between random variables no matter whether the relationship between two random variables is linear or nonlinear. This makes MI a DTD decision statistic that is robust to nonlinearities.

4.1.3 Performance Evaluation

As discussed in Section 2.3.1, we use receiver operating characteristic (ROC) curve to evaluate the performance of an DTD. Since the statistical distribution of ξ is unknown, P_D and P_{FA} are obtained by numerical methods. Detection or false alarm is counted only during

the active portions of the far-end speech because the effect of a pause or inactive portions of the far-end speech on the filter update is minimal. We define an indicator ν_x to reflect the activity of far-end speech

$$\nu_x = \begin{cases} 1, & \text{far-end speech is active,} \\ 0, & \text{far-end speech is inactive.} \end{cases} \tag{170}$$

Similarly, an indicator ν_s is adopted for the near-end speech $s(n)$, since silence in $s(n)$ usually does not cause AEC filters to diverge. We define the DTD output as a function of the threshold T

$$\phi_T = \begin{cases} 1, & \text{if } \xi < T, \\ 0, & \text{if } \xi \geq T. \end{cases} \tag{171}$$

Before measuring P_D, the threshold T is predetermined to meet the given P_{FA}. First, the decision statistic ξ is calculated with $s(n) = 0$ (i.e., the near-end speech is absent) as a function of the threshold. The probability of false alarm at each threshold point T is estimated as

$$P_{FA}(T) = \frac{\sum_i \phi_T(i)\nu_x(i)}{\sum_i \nu_x(i)}, \tag{172}$$

where the index i indicates the i^{th} DTD decision-making situation, which is actually the data block number in the proposed method, since MI is calculated in a block-by-block fashion. Then, the threshold T is determined to achieve the given P_{FA}. Once the threshold is determined, the near-end speech is applied, and the detection procedure runs again. The probability of detection is calculated as

$$P_D(T) = \frac{\sum_i \phi_T(i)\nu_x(i)\nu_s(i)}{\sum_i \nu_x(i)\nu_s(i)}. \tag{173}$$

Note that the detection probability can be affected by several factors, including

1. Signal-to-noise ratio (SNR), defined as

$$\text{SNR(dB)} = 10\log_{10}\frac{E\left[x^2(n)\right]}{E\left[v^2(n)\right]}. \tag{174}$$

   Recall that $x(n)$ is the far-end speech, and $v(n)$ is the background noise picked up by the microphone.

2. Far-to-near ratio (FNR), defined as the ratio between the far-end speech power level to the near-end speech power level

$$\text{FNR(dB)} = 10\log_{10}\frac{E\left[x^2(n)\right]}{E\left[s^2(n)\right]}. \tag{175}$$

3. Channel gain $\|\boldsymbol{h}\|_2$, where $\|\cdot\|_2$ denotes the l_2 norm.

In order to thoroughly quantify detection performance, exhaustive simulations (in terms of spatial, voice frequency and so on) of DTD methods are required. However, for practical systems it is not necessary to simulate all possible situations to evaluate DTD algorithms. For ROC implementations in the simulation section, we make reasonable assumptions and confine the situation to representative cases.

4.1.4 Simulations

In the simulations, the nonlinearity of the loudspeaker was modeled by a sigmoid function

$$g(x) = \frac{2}{1+e^{-2x}} - 1. \tag{176}$$

The IMAGE method [2] was used to generate a room impulse response of length 256 with a sampling rate 8 kHz. We normalized the room impulse response such that $\|\boldsymbol{h}\|_2 = 1$ in order to remove the dependence on the overall room response level. Both far-end and near-end signals consist of real speech, the level of which were adjusted so that FNR = 0 dB. For comparison purposes, we also implemented the method in [29].

In the first experiment, we applied DTDs to the linear AEC case ($g(x) = x$). The simulation was carried out under a noise-free condition. The results from both the proposed DTD algorithm and the method of [29] are shown in Fig. 23. Figure 23 (a) shows the far-end speech $x(n)$ with 40,000 samples. Figure 23 (b) shows the near-end speech $s(n)$ which starts at the $15,000^{th}$ sample and ends at the $30,000^{th}$ sample. Figure 23 (c) shows the microphone-received signal $y(n)$. The DTD determines during which period the near-end speech is present based on $x(n)$ and $y(n)$. Figure 23 (d) and (e) show the detection statistics $\xi_c = \int |\gamma_{xy}(f)|^2 \mathrm{d}f$ of [29] and the proposed $\xi_m = I(x;y)$, respectively. The DTD decisions are marked as circles on the top representing the "no-double-talk and AEC on" and circles

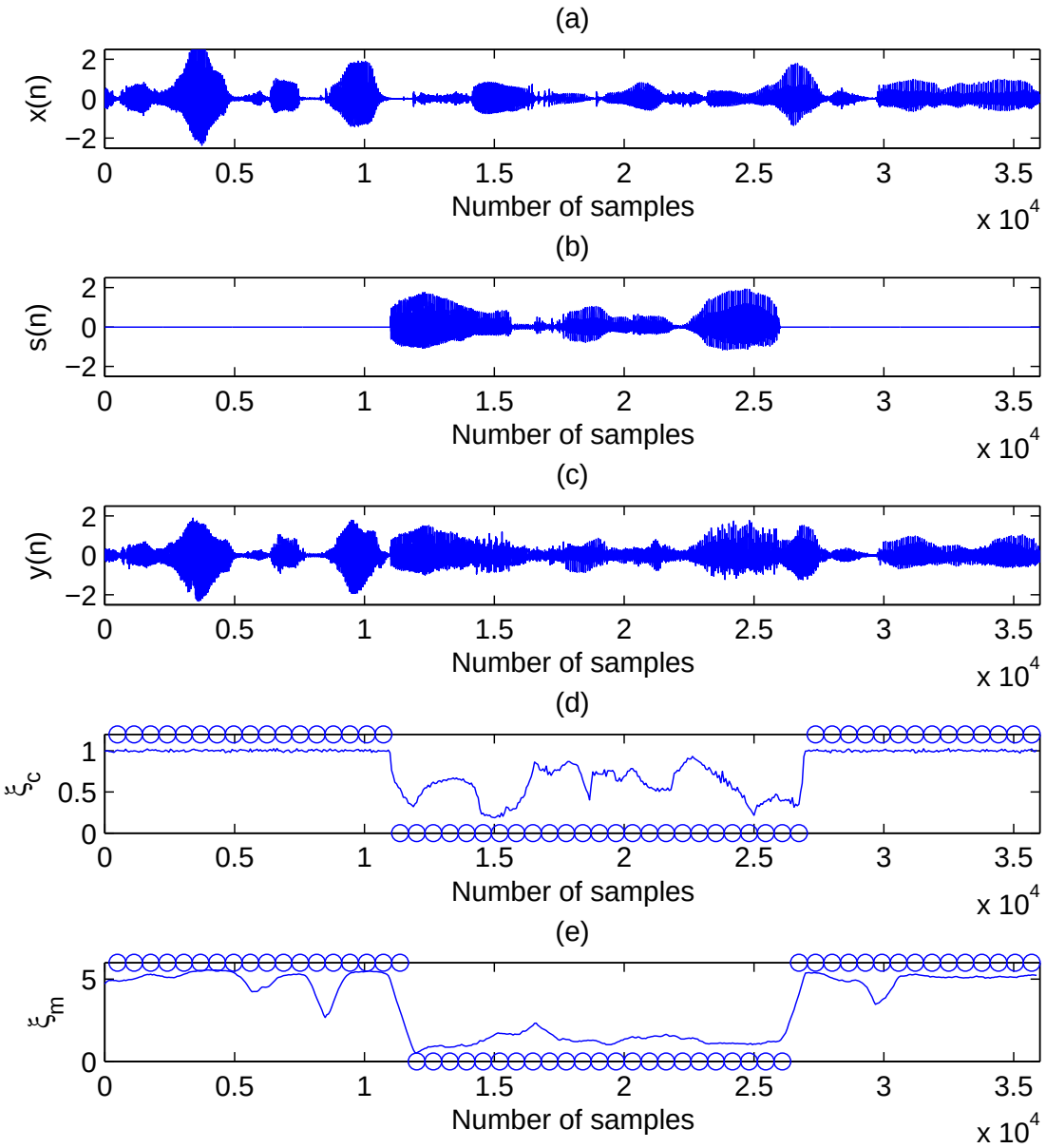


Figure 23: DTD in the linear case: (a) far-end speech $x(n)$, (b) near-end speech $s(n)$, (c) microphone-received signal $y(n)$, (d) ξ_c and the DTD decision, (e) ξ_m and the DTD decision. Circles: DTD decision, top = no double-talk, bottom = double-talk.

on the bottom as "double-talk present and AEC off". The thresholds used in this example were $T_c = 0.92$ and $T_m = 2.5$ for these two methods. We observe that both methods

achieved good double-talk detection performance.

For the second example, we considered the nonlinearity $g(x)$ as in (176) in the acoustic

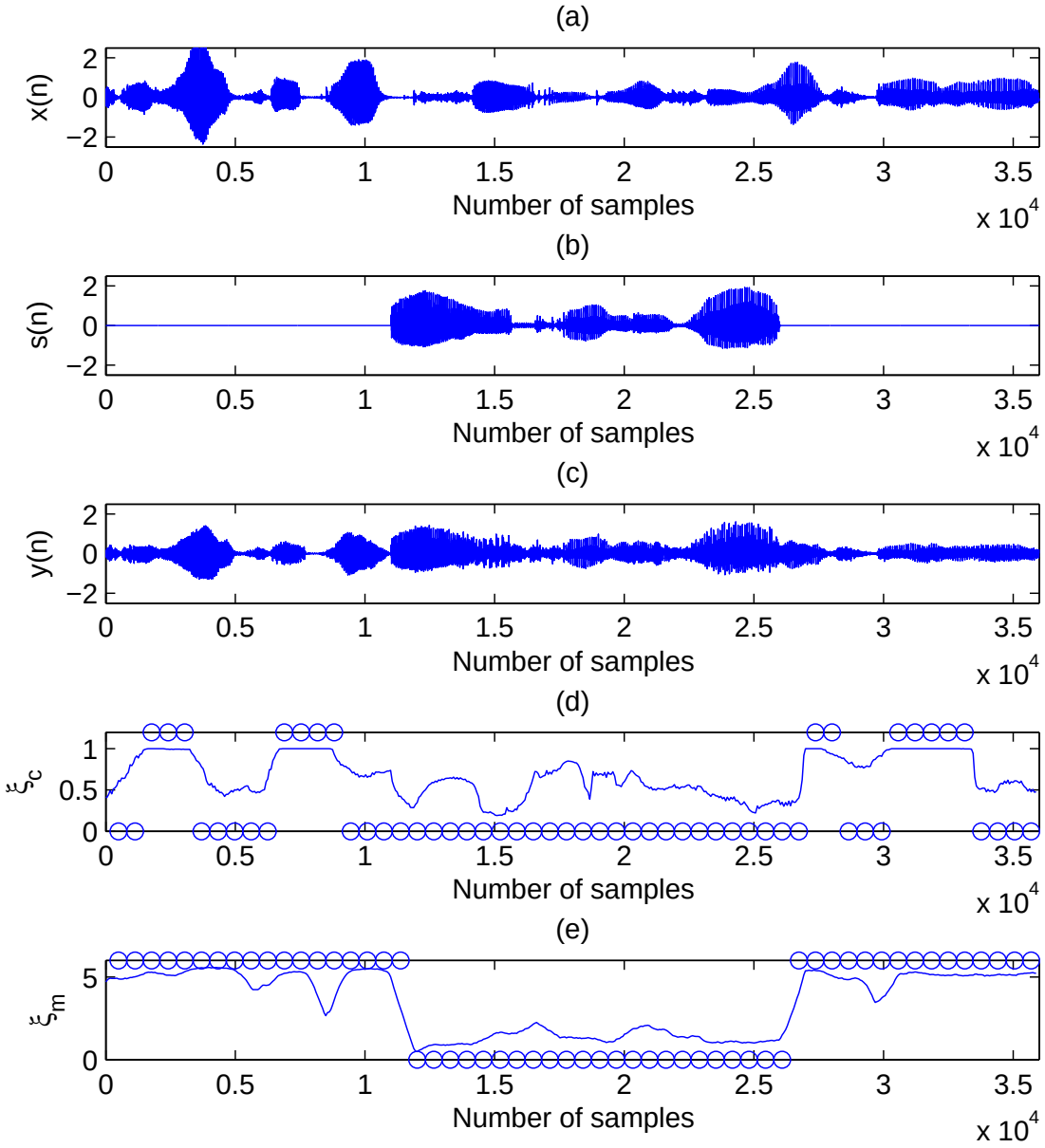
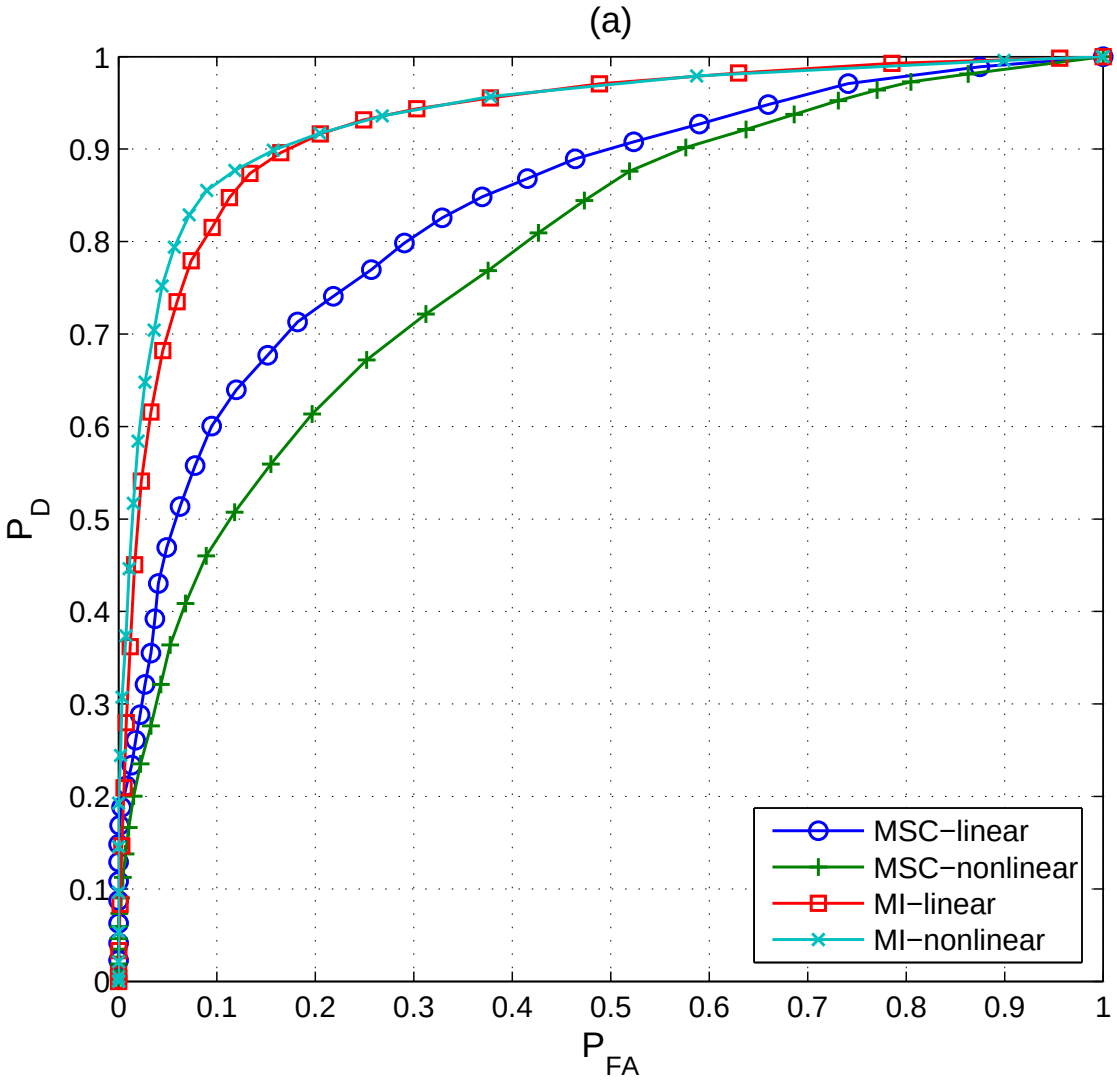


Figure 24: DTD in the nonlinear case: (a) far-end speech $x(n)$, (b) near-end speech $s(n)$, (c) microphone-received signal $y(n)$, (d) ξ_c and the DTD decision, (e) ξ_m and the DTD decision. Circles: DTD decision, top = no double-talk, bottom = double-talk.

echo path. We used the same configurations as in the linear case to perform double-talk detection. The results are shown in Fig. 24. It can be seen that the performance of the coherence-based method of [29] degraded a lot: with the same threshold as in the linear case, the probability of false alarm increased. The proposed MI-based method performed similarly as in the linear case. This example illustrates the robustness of our proposed method in the presence of nonlinearity.

In the final experiment, we obtained ROC curves under different SNRs. The far-end speech was 5 seconds long or 40000 samples at 8 kHz sampling rate. For the near-end, four different speech segments (two males, two females) were chosen, each about 1.875 seconds long. In order to achieve better statistical significance, the calculations were averaged over 16 different conditions: four 1.875-second near-end speech samples located at different positions within the 5-second far-end speech. The ROC curves are shown in Fig. 28 with the SNR of 30 dB and 10 dB, respectively. It can be seen that under both the high-SNR (30 dB) and low-SNR (10 dB) cases, our proposed method outperformed the coherence-based method of [29] in terms of achieving a higher P_D for a given P_{FA}. On the other hand, the performance of the coherence-based method was degraded in the presence of nonlinearity, whereas the proposed MI-based method produced almost the same performance with or without nonlinearity, re-affirming the robustness of the proposed method. Recall that MI is a measure of statistical dependence between random variables, whereas the coherence function only measures the linearity between them. Zero MI always implies statistical independence, but the coherence function can be zero for highly dependent non-Gaussian data. Thus, MI is more powerful than the coherence function for measuring independence.


(a)
P_D
P_FA
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
MSC−linear
MSC−nonlinear
MI−linear
MI−nonlinear

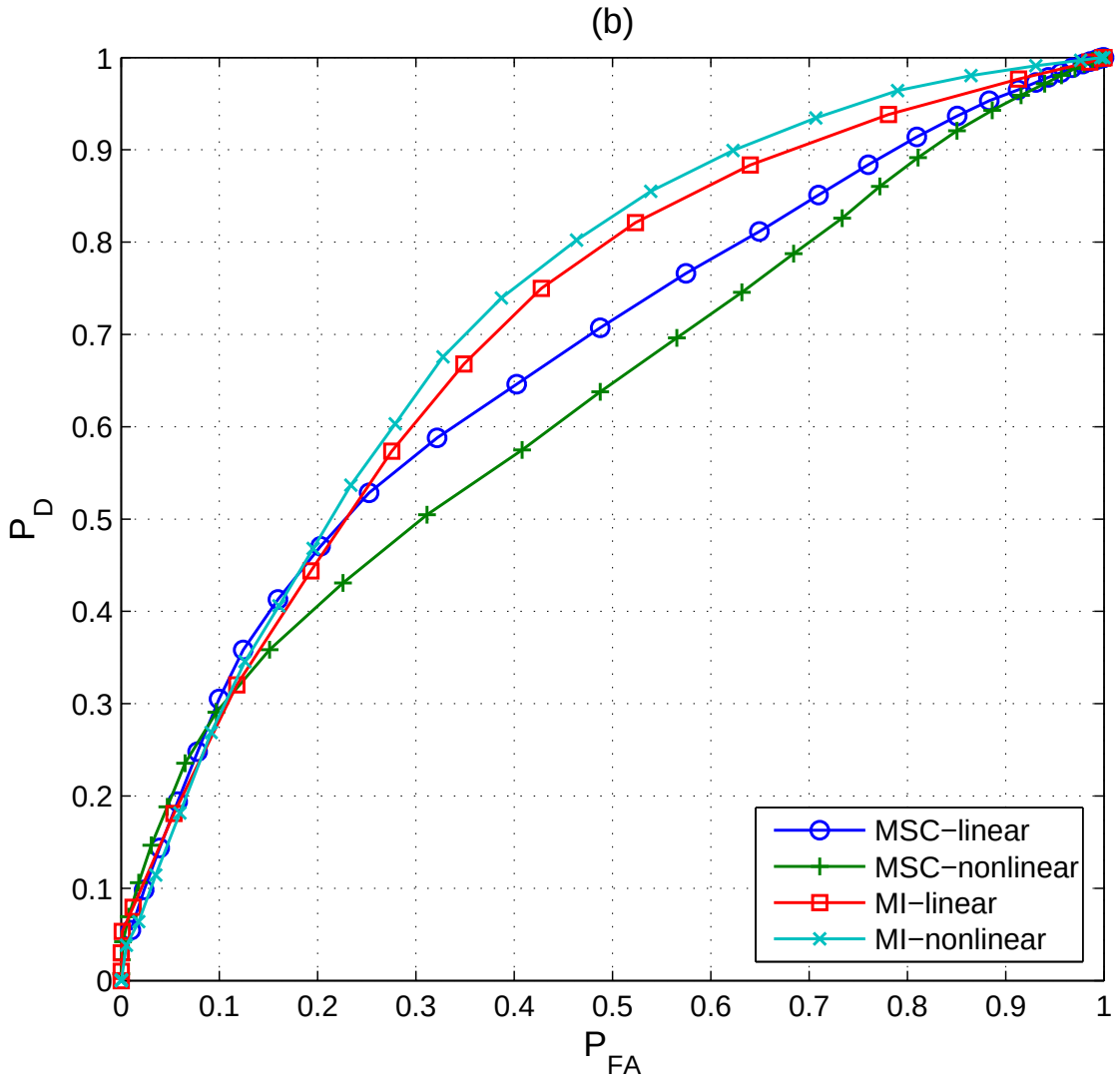


Figure 25: ROCs: (a) SNR=30dB; (b) SNR=10dB.

4.2 DTD Using Generalized Mutual Information for Stereophonic NAECs

The need to improve the quality of service in voice communication systems has led to the exploration of stereo systems, which deploy two microphones and two loudspeakers at both communication ends. Stereophonic acoustic echo cancellation is indispensable for stereo systems and it can be viewed as a straightforward generalization of the single-channel acoustic echo cancellation principle [97]. Similar to signal-channel (monophonic) systems, DTDs play a very important role in stereophonic systems. Existing DTDs for two-channel acoustic echo cancellation mainly utilize normalized cross-correlation vector (NCCV)-based methods, which is first proposed in [7]. The decision statistic is formed based on the NCCV between the far-end signal vector and microphone-received signal. To reduce computational complexity and simplify implementation, a frequency-domain NCCV scheme is proposed in [27]. However, the stereo auto-correlation function can be very ill-conditioned due to inter-channel correlations. [55] proposes a method to decrease the influence of inter-channel correlations on the reliability of DTDs by applying a weight function for the NCCV estimation. However, in the presence of nonlinear acoustic echo path, the NCCV-based method is not expected to perform well. In this section, we design a DTD based on the generalized mutual information (GMI) between the input signal vector and output signal of a stereo acoustic echo path.

4.2.1 DTD in Stereophonic Acoustic Echo Cancellation

Figure 26 shows a block diagram of a stereo nonlinear acoustic echo cancellation system. For simplicity, the echo canceller structure for only one microphone is shown in the receiving room on the left. The other microphone structure, not shown in Fig. 26, has an identical echo canceller structure for different receiving room echo paths. We consider applications where the loudspeaker/PA exhibits nonlinear characteristics; such might be the case when the performance of analog components are sacrificed for price advantage. Let the nonlinearities in two echo paths be represented by mappings $d_1(\cdot)$ and $d_2(\cdot)$. We denote the signals picked up by the microphones in the transmission room by $x_1(n)$ and $x_2(n)$, and denote the echo signal in the receiving room by $c(n)$. At the receiving side, two room impulse responses are

denoted by $h_1(n)$ and $h_2(n)$. Then, the microphone-received signal $y(n)$ is composed of the nonlinear acoustic echo $c(n)$, background noise $v(n)$, and near-end speech $s(n)$ (if any):

$$y(n) = c(n) + s(n) + v(n), \tag{177}$$

where the echo $c(n)$ is expressed as

$$c(n) = h_1(n) * d_1(x_1(n)) + h_2(n) * d_2(x_2(n)). \tag{178}$$

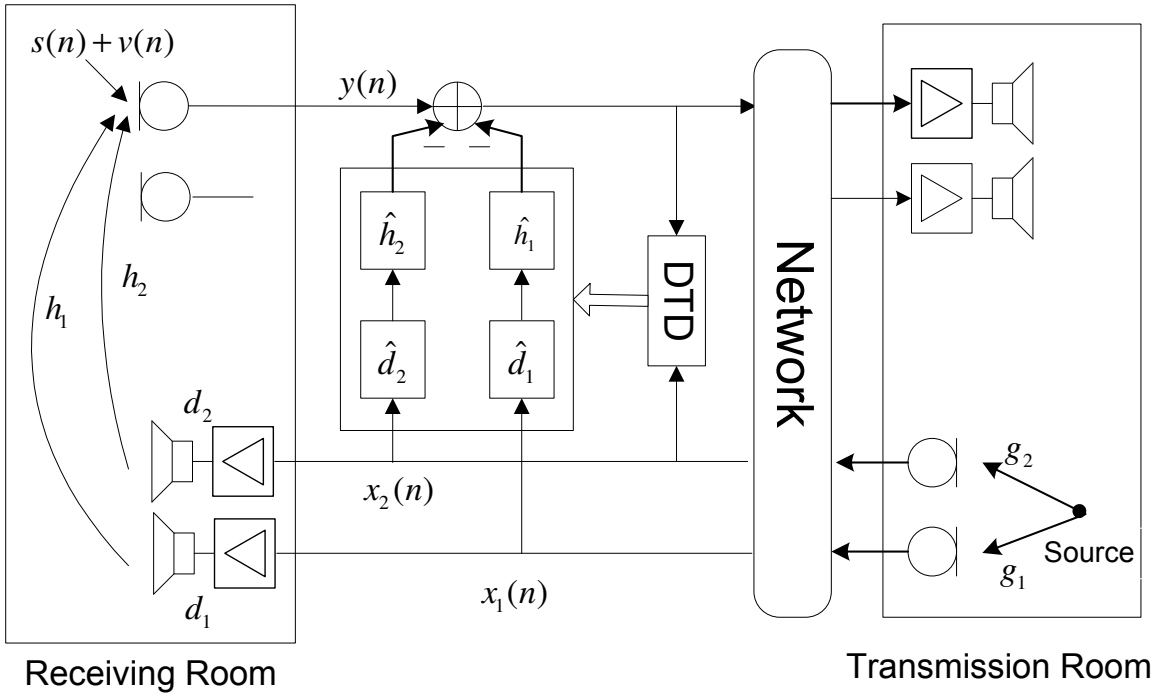


Figure 26: Block diagram of stereo nonlinear acoustic echo cancellation.

From Fig. 26, the nonlinear acoustic echo canceller (NAEC) tries to model this unknown system by a pair of cascaded nonlinear adaptive filters $(\hat{d}_1, \hat{h}_1)$ and $(\hat{d}_2, \hat{h}_2)$. However, the occurrence of the near-end speech $s(n)$ may cause the divergence of the NAEC filters. Thus, a DTD is essential for the NAEC to work. Similar to the rationale in [89], the NCCV-based method [27] is not expected to perform well in the nonlinear scenario, since the NCCV between vector $(x_1(n), x_2(n))$ and $y(n)$ cannot capture the nonlinear relationship between them. Consider $y(n)$ as the n^{th} realization of the random variable y, similarly for $x_1(n)$ and $x_2(n)$. From (177), y is nonlinearly related with vector (x_1, x_2). A straightforward

extension from [89] is that the decision statistic can be formed by

$$\xi = I_1(x_1, x_2; y), \tag{179}$$

where $I_1(x_1, x_2; y)$ denotes the MI between vector (x_1, x_2) and y. However, the high-dimensional computation of MI is difficult and unpractical using traditional methods, such as histograms and kernel methods [55]. Moveover, another problem is that this MI is not well normalized. Indeed, we can say that ξ is maximized when $s(n) = 0$. However, we do not know the value of ξ in general. The amount of MI depends a great deal on the statistics of signals and echo paths. As a result, the best value of the threshold will vary a lot from one experiment to another. So there is no "natural" threshold level associated with the variable ξ. This leads us to GMI.

4.2.2 Generalized Mutual Information (GMI) and Its Calculation

Let $(\boldsymbol{x}, y) = (x_1, ..., x_D, y) \in R^{D+1}$ be a $(D+1)$-dimensional real-valued random vector where each $x_i,\ i = 1, ..., D$, y, $\boldsymbol{x}$, and $(\boldsymbol{x}, y)$ are continuous valued and with probability density function $p_{x_i}(x_i)$, $p_y(y)$, $p_{\boldsymbol{x}}(\boldsymbol{x})$, and $p_{\boldsymbol{x},y}(\boldsymbol{x}, y)$, respectively. The GMI between vector $\boldsymbol{x}$ and y is defined based on the Rényi entropy [79]

$$I_2(\boldsymbol{x}, y) = \log_2 \int\limits_{R^{D+1}} \frac{p^2_{\boldsymbol{x},y}(\boldsymbol{x}, y)}{p_{x_1}(x_1) \cdots p_{x_D}(x_D) p_y(y)} \mathrm{d}\boldsymbol{x}\mathrm{d}y - \log_2 \int\limits_{R^D} \frac{p^2_{\boldsymbol{x}}(\boldsymbol{x})}{p_{x_1}(x_1) \cdots p_{x_D}(x_D)} \mathrm{d}\boldsymbol{x}. \tag{180}$$

GMI measures the degree to which $\boldsymbol{x}$ and y are dependent and the value of $I_2(\boldsymbol{x}; y)$ grows as $\boldsymbol{x}$ and y become more dependent. The following properties hold for GMI.

Property I: $0 \leq I_2(\boldsymbol{x}; y) \leq \infty$.

Property II: $I_2(\boldsymbol{x}; y) = 0$ if and only if $\boldsymbol{x}$ and y are statistically independent.

Property III: $I_2(\boldsymbol{x}; y) = I_2(\boldsymbol{u}; v)$ for any one-to-one mapping $f_i : u_i = f_i(x_i),\ i = 1, ..., D$ and $g : v = g(y)$.

In stead of directly estimating probability density functions, [76] proposed an algorithm to estimate $I_2(\boldsymbol{x}; y)$ bypassing the estimation of distributions. Assuming N realizations of random vector $\{\boldsymbol{x}(n), y(n)\}_{n=1}^N$, the algorithm in [76] works in two steps:

Step 1: Transform each sequence to the series of relative rank number $x_i(n) \to r_{x_i}(n),\ i =$

$1, ..., D,\ n = 1, ..., N$, where

$$r_{x_i}(n) = \frac{|\{x_i(m) < x_i(n), 1 \leq m \leq N\}|}{N}, \tag{181}$$

and $|\cdot|$ denotes the cardinality of a set. The same transform applies to $y(n)$.

Step 2: Construct a $(D+1)$-dimensional vector

$$\boldsymbol{r}^{(D+1)}(n) = (r_{x_1}(n), ..., r_{x_D}(n), r_y(n)), \tag{182}$$

and determine the number of pairs with distance less than $\varepsilon/2$

$$C_{D+1} = \frac{\left|\{(i,j) : \|\boldsymbol{r}^{(D+1)}(i) - \boldsymbol{r}^{(D+1)}(j)\|_\infty < \varepsilon/2\}\right|}{N_{total}}, \tag{183}$$

where $1/N \ll \varepsilon/2 \ll 1$, $1 \leq i < j \leq N$, $\|\cdot\|_\infty$ denotes the infinity norm, and the total number of pairs is given by $N_{total} = N(N-1)/2$. Calculate similarly C_D but with the D-dimensional vector

$$\boldsymbol{r}^{(D)}(n) = (r_{x_1}(n), ..., r_{x_D}(n)). \tag{184}$$

Then, the estimation of $I_2(\boldsymbol{x}, y)$ is be obtained as

$$\hat{I}_2(\boldsymbol{x}, y) = -\log_2 \varepsilon - \log_2 C_D + \log_2 C_{D+1}. \tag{185}$$

In addition, it has been shown in [76] that the estimator is consistent for $N \to \infty$ and $\varepsilon \to 0$ and $\hat{I}_2(\boldsymbol{x}, y) \leq -\log_2 \varepsilon$. The upperbound of the estimate makes it a desirable feature for DTD designs.

4.2.3 A Test Statistic Based on GMI

Based on (177), if $s(n)$ is absent then $y(n) = h_1(n) * d_1(x_1(n)) + h_2(n) * d_2(x_2(n))$, i.e., $y(n)$ fully depends on $x_1(n)$ and $x_2(n)$ (a noise-free scenario is considered here $v(n) = 0$). Hence, to determine whether the near-end signal $s(n)$ is present or not is equivalent to measuring the degree of dependency between $(x_1(n), x_2(n))$ and $y(n)$. Therefore, we propose to use as the decision statistic, the normalized GMI between (x_1, x_2) and y [85]:

$$\xi = -\frac{I_2(x_1, x_2; y)}{\log_2 \varepsilon}, \tag{186}$$

and formulate a binary hypothesis test

$$H_0: \text{ if } s(n) = 0 \text{ (double talk is absent)}, \ \xi \geq T, \tag{187}$$

$$H_1: \text{ if } s(n) \neq 0 \text{ (double talk is present)}, \ \xi < T. \tag{188}$$

It is pointed out that the advantage of using GMI in (186) instead of MI in (179) is twofold: (i) the estimation of GMI without direct calculation of distribution relieves the computational complexity; (ii) the normalized representation of GMI leads to a function running between 0 and 1, which facilitates the threshold selection.

4.2.4 Simulations

In the receiving room, the nonlinearity of each loudspeaker was modeled by a sigmoid function

$$d_i(x) = \frac{2}{1 + e^{-\alpha_i x}} - 1, \tag{189}$$

with $\alpha_1 = 2$, $\alpha_2 = 2.5$ for each channel, respectively. The receiving and transmission room impulse responses $h_i (i = 1, 2)$ and $g_i (i = 1, 2)$ were both generated using the IMAGE method with lengths 256 and 50, respectively [2]. For comparison purposes, we also implemented the NCCV-based method in [27].

The first experiment was carried out under a noise-free condition. The results from both the proposed DTD algorithm and the method of [27] are shown in Fig. 27. Figure 27 (a) and (b) show the microphone-received signal $x_1(n)$ and $x_2(n)$ in the transmission room. Figure 27 (c) and (d) show the near-end speech $s(n)$ and the microphone-received signal $y(n)$ in the receiving room, respectively. The DTD determines during which period the near-end speech is present based on signals $x_1(n)$, $x_2(n)$, and $y(n)$. Figure 27 (e) and (f) show the detection statistics ξ_c of [27] and the proposed $\xi_m = I_2(x_1, x_2; y)$, respectively. The decisions made by the DTDs are marked as circles on the top representing the "double-talk absent and AEC on" and circles on the bottom representing "double-talk present and AEC off". The thresholds used in this example were $T_c = 0.91$ and $T_m = 0.52$ for these two methods. We observe that the proposed method achieves better performance than the method of [27], since based on the threshold without miss detection, the false alarm occurred several times in the method of [27].

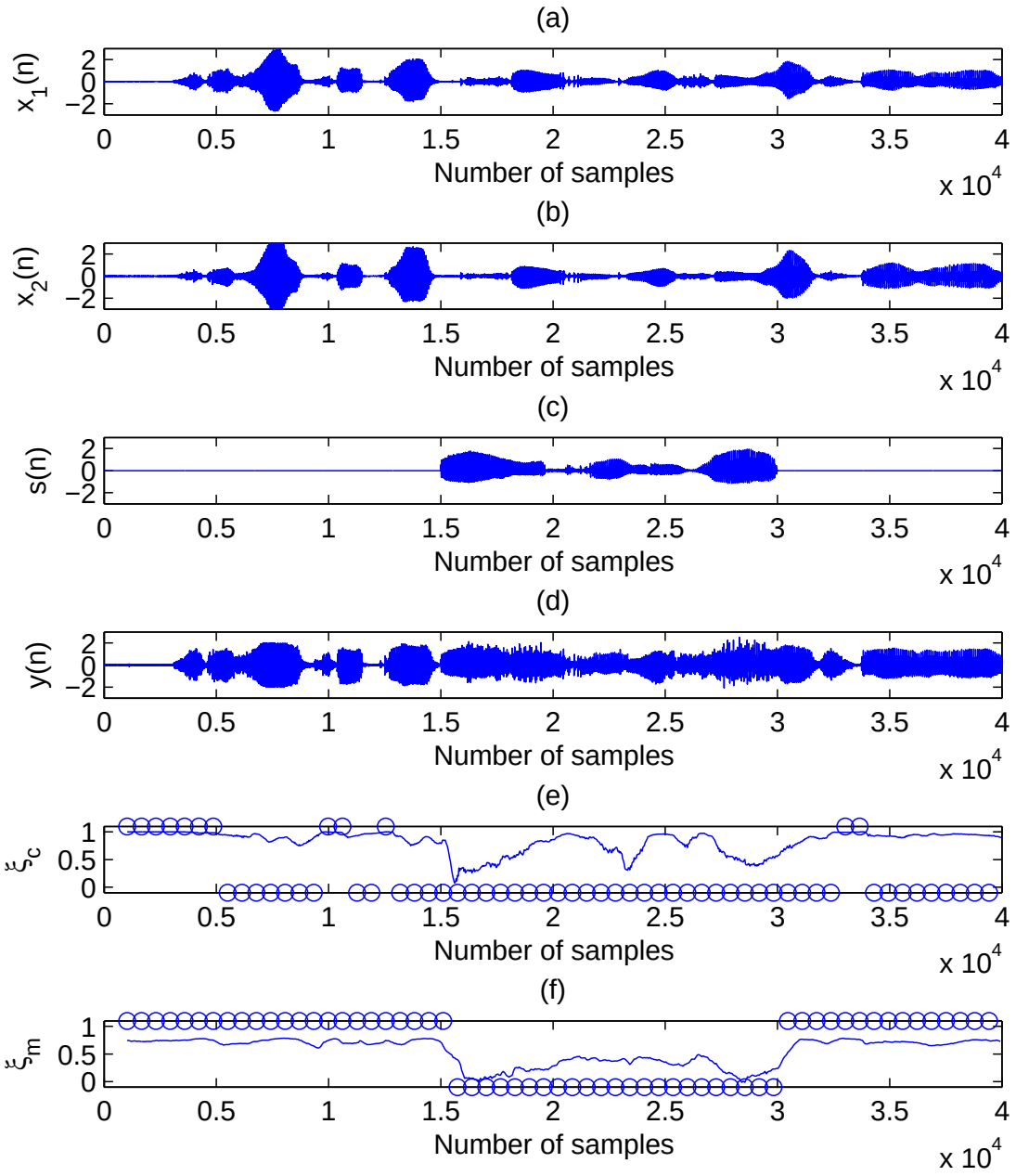


Figure 27: DTD performance: (a) far-end speech in the 1st channel $x_1(n)$, (b) far-end speech in the 2nd channel $x_2(n)$, (c) near-end speech $s(n)$, (d) microphone-received signal $y(n)$, (e) ξ_c and the DTD decision, (f) ξ_m and the DTD decision. Circles: DTD decision, top = no double-talk, bottom = double-talk.

In the second experiment, receiver operating characteristic (ROC) curves were obtained with signal-to-noise ratio (SNR) of 30dB for both channels. In order to achieve better statistical significance, the calculations were averaged over 16 different conditions similar to [89]. The ROC curves are shown in Fig. 28. It can be seen that our proposed method outperformed the NCCV-based method of [27] in terms of achieving a higher P_D for a given P_{FA}.

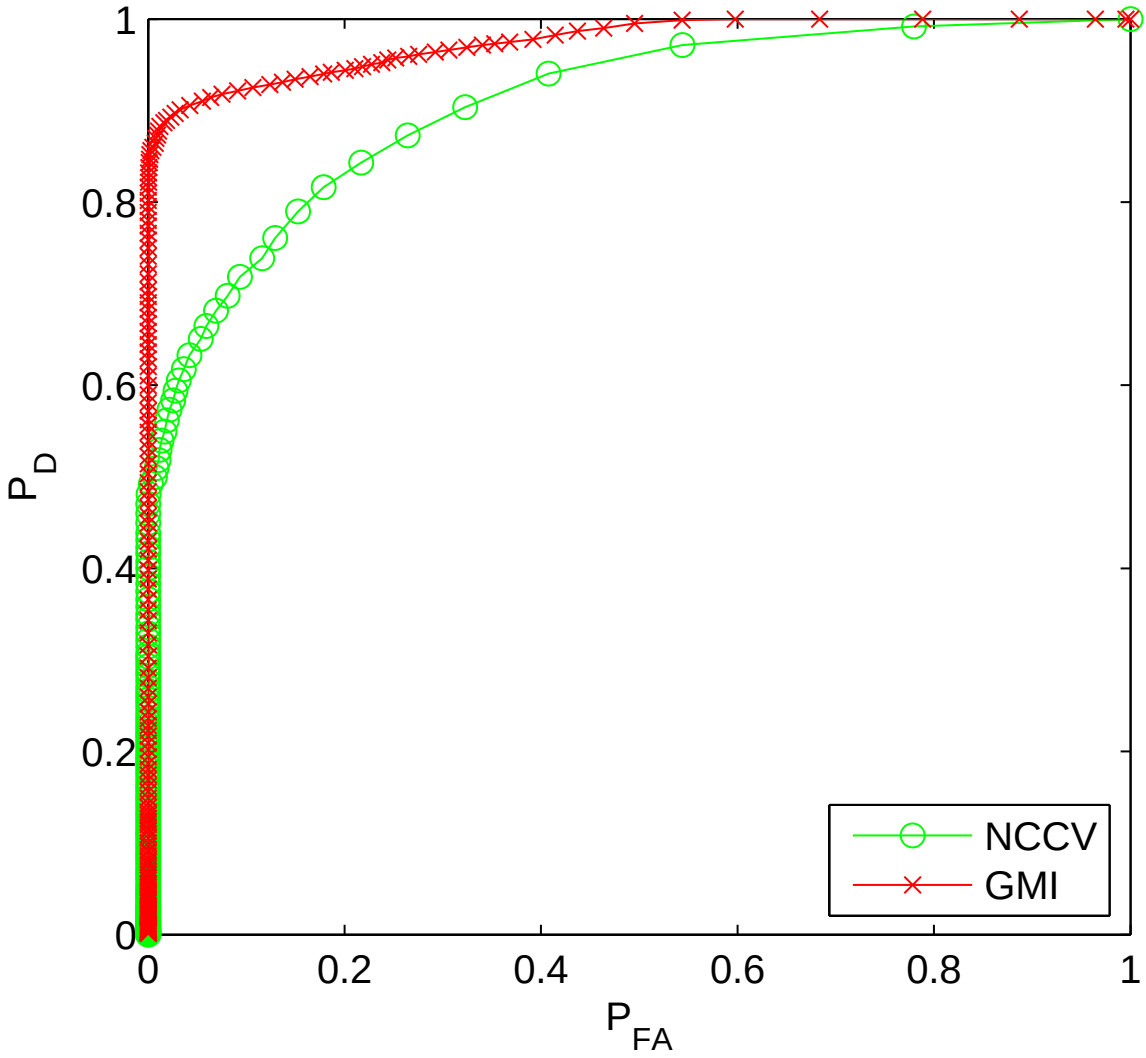


Figure 28: ROC with SNR of 30dB.

4.3 Variable Step Size (VSS) and Variable Tap Length (VTL) LMS

In this section, we focus on the learning-rate control for the LMS algorithm. Among various types of adaptive algorithms, the LMS algorithm is well known and widely used for its simplicity and robustness [81]. As discussed in Section 2.3.2, the performance of the LMS algorithm, in terms of convergence rate, misadjustment, mean-square error (MSE), and computational cost, is believed to be governed by the step size of the adaptive filter [95, 75]. On the other hand, recent research results indicate that the filter tap-length is factor to affect the LMS performance [68].

Usually, to describe an unknown linear time-invariant system accurately, a sufficiently large filter tap length is needed, since the MSE is likely to increase if the tap length is undermodeled [37, 68]. However, the computational cost is proportional to the tap length. Moreover, an increase in filter length can slow down the convergence rate dramatically due to the step-size restrictions [37, 109]. Thus, a variable tap-length algorithm, which finds the appropriate tap-length for each iteration, is necessary to achieve both small MSE and fast convergence. Existing variable tap-length algorithms such as [36, 34] are sensitive to the parameter selection, i.e., different parameters result in different performance, according to the discussion in [34].

Recently, the impulse response envelope is suggested to be one essential factor that determines the convergence rate of a deficient-length filter [37, 111]. In many applications such as acoustic echo cancellation, the unknown impulse response follows an exponential decay envelope. For this kind of systems, a theoretically optimal variable tap-length sequence is introduced in [37]. However, this algorithm entails large computational complexity as a result of trying to solve Lambert's W-function. To reduce the complexity, an adaptive solution for the optimal tap length is proposed in [111], which ensures a well-behaved transient tap-length convergence. However, to the best of our knowledge, variable tap-length algorithms have not been proposed in conjunction with a variable step size. It is well known that with the stability conditions, the efficient step-size control trade-offs fast convergence rate and tracking ability with filter misadjustment. Thus, we are motivated to develop a low complexity algorithm with both a variable tap length and step size.

4.3.1 Convergence Analysis of Deficient-Length LMS Filter

Consider an unknown length N exponential decay impulse response $\boldsymbol{c}_N = [c_0, c_1, ..., c_{N-1}]^T$ modeled by

$$c_i = e^{-i\tau} r(i), \quad i = 0, 1, ..., N-1, \tag{190}$$

where the decay rate τ is a known positive constant and $r(i)$ is a zero-mean i.i.d. Gaussian random process with variance σ_r^2. The observed signal is a linear convolution of the transmitted signal and the impulse response:

$$d(n) = \boldsymbol{x}_N^T(n)\boldsymbol{c}_N + v(n), \tag{191}$$

where $\boldsymbol{c}_N = [c_0, c_2, ..., c_{N-1}]^T$ is the channel response and $\boldsymbol{x}_N(n) = [x(n), x(n-1), ..., x(n-N+1)]^T$ is the input vector and $v(n)$ is the additive noise. Here the problem we are considering is to estimate $\{c_i\}$ given $d(n)$ and $x(n)$ using an LMS algorithm with variable tap length and step size.

In the variable tap-length and variable step-size LMS algorithm, both the tap-length and step-size are time-varying rather than fixed. We denote by $M(n)$ and $\mu(n)$, respectively, the integer tap-length and step-size for the coefficients updated at the n^{th} iteration, and assume that $M(n) \leq N$. With the LMS criterion, the filter coefficients are updated by [37]

$$\boldsymbol{w}_{M(n+1)} = \begin{bmatrix} \boldsymbol{w}_{M(n)}(n) \\ \underline{0}_{M(n+1)-M(n)} \end{bmatrix} + \mu(n+1)e(n)\boldsymbol{x}_{M(n+1)}(n+1), \tag{192}$$

where $e(n)$ is the estimated error defined as

$$e(n) = d(n) - \boldsymbol{x}_{M(n)}^T(n)\boldsymbol{w}_{M(n)}, \tag{193}$$

$\boldsymbol{x}_{M(n)}(n) = [x(n), x(n-1)..., x(n-M(n)+1)]^T$ and $\boldsymbol{w}_{M(n)} = \left[w_1(n), w_2(n)..., w_{M(n)}(n)\right]^T$ are the $M(n)$-tap adaptive filter input vector and the coefficients vector, respectively, and $\underline{0}$ denotes a vector with all-zero entries. In the following, we introduce how to update the tap-length $M(n)$ and step-size $\mu(n)$ at each iteration.

Similar to [37, 111], we partition the impulse response $\boldsymbol{c}_N$ into two parts as

$$\boldsymbol{c}_N \triangleq \begin{bmatrix} \boldsymbol{c}'_{M(n)} \\ \boldsymbol{c}''_{N-M(n)} \end{bmatrix}, \tag{194}$$

where $\boldsymbol{c}'_{M(n)}$ can be viewed as the part modeled by $\boldsymbol{w}_{M(n)}$, and $\boldsymbol{c}''_{N-M(n)}$ is the undermodeled part. Define the estimation errors of partial and total coefficients, respectively, as

$$\boldsymbol{\delta}_{M(n)}(n) = \boldsymbol{w}_{M(n)} - \boldsymbol{c}'_{M(n)}, \tag{195}$$

and

$$\boldsymbol{\delta}_N(n) = \begin{bmatrix} \boldsymbol{w}_{M(n)} \\ \underline{0}_{N-M(n)} \end{bmatrix} - \boldsymbol{c}_N. \tag{196}$$

Combining (191) and (195), we rewrite the signal estimate error in (193) as

$$e(n) = -\boldsymbol{x}_N^T(n)\boldsymbol{\delta}_N(n) + v(n). \tag{197}$$

Substituting (197) into (192), we obtain

$$\boldsymbol{\delta}_N(n+1) = \mathbf{A}(n)\boldsymbol{\delta}_N(n) + \mu(n+1)v(n) \begin{bmatrix} \boldsymbol{x}_{M(n+1)}(n+1) \\ \underline{0}_{N-M(n+1)} \end{bmatrix}, \tag{198}$$

where

$$\mathbf{A}(n) = \boldsymbol{I}_N - \mu(n+1) \begin{bmatrix} \boldsymbol{x}_{M(n+1)}(n+1) \\ \underline{0}_{N-M(n+1)} \end{bmatrix} \boldsymbol{x}_N^T(n), \tag{199}$$

and $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.

To quantitatively evaluate the misadjustment of the filter coefficients, MSD is taken as a figure of merit, which is defined as

$$\Lambda(n) \triangleq \Lambda(M(n), \mu(n)) = E\left[\|\boldsymbol{\delta}_N(n)\|_2^2\right], \tag{200}$$

where $\|\cdot\|_2$ denotes the l_2 norm. Note that at each iteration, MSD depends on both $M(n)$ and $\mu(n)$. Assume that both the signals $x(n)$ and $v(n)$ are i.i.d. zero-mean Gaussian with variances σ_x^2 and σ_v^2, respectively. According to the analysis in [37, 111], we find that

$$\begin{aligned} \Lambda(n+1) =& \beta(n+1)\Lambda(n) + (\eta(n+1) - \beta(n+1)) \\ & E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right] + \gamma(n+1), \end{aligned} \tag{201}$$

where

$$\beta(n+1) = 1 - 2\mu(n+1)\sigma_x^2 + (M(n+1)+2)\mu^2(n+1)\sigma_x^4, \tag{202}$$

$$\eta(n+1) = 1 + M(n+1)\mu^2(n+1)\sigma_x^4, \tag{203}$$

$$\gamma(n+1) = M(n+1)\mu^2(n+1)\sigma_x^2\sigma_v^2. \tag{204}$$

From (201), we know that the convergence is affected by step-size, filter length, and the undermodeled system response $\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2$. When $M(n)$ increases to N, $\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2$ will drop to zero. Thus, we get the MSD update equation of the full-length LMS filter

$$\Lambda(n+1) = \beta(n+1)\Lambda(n) + \gamma(n+1). \tag{205}$$

Note that no information about the system response exists in (205). Therefore, the convergence of the full-length LMS filter will not be affected by the shape of the system response, while that of the deficient-length LMS filter will. This is the primary difference between the full-length filter and deficient-length filter [37].

4.3.2 VSS-VTL LMS Algorithm with Exponential Decay Impulse Response

With (201), we may speculate the existence of an optimal time-variant filter length and step-size, which can result in the fastest convergence. In the following, we propose to a solution of both tap length and step size by minimizing MSD at each iteration [87]. Observing that the MSD is a multi-variate function with respect to tap-length and step-size. We start by trying to solve stationary points for this function. Taking the first-order partial derivative of $\Lambda(n+1)$ with respect to $M(n+1)$ and $\mu(n+1)$, respectively, we obtain

$$\begin{aligned}\frac{\partial \Lambda(n+1)}{\partial M(n+1)} =& \mu^2(n+1)\sigma_x^4\Lambda(n) + \mu^2(n+1)\sigma_x^2\sigma_v^2 \\ &+ 2\mu(n+1)\sigma_x^2\left(1-\mu(n+1)\sigma_x^2\right)\frac{\mathrm{d}E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right]}{\mathrm{d}M(n+1)}, \qquad (206)\\ \frac{\partial \Lambda(n+1)}{\partial \mu(n+1)} =& 2\sigma_x^2\left((M(n+1)+2)\mu(n+1)\sigma_x^2 - 1\right)\Lambda(n) \\ &+ 2\sigma_x^2(1-2\mu(n+1)\sigma_x^2)E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right] \\ &+ 2\mu(n+1)\sigma_x^2\sigma_v^2 M(n+1). \qquad (207)\end{aligned}$$

Based on the impulse pulse model in (190), we obtain

$$E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right] = \frac{e^{-2M(n+1)\tau} - e^{-2N\tau}}{1-e^{-2N\tau}} E\left[\|\boldsymbol{c}_N\|_2^2\right], \tag{208}$$

$$E\left[\|\boldsymbol{c}_N\|_2^2\right] = \frac{1-e^{-2N\tau}}{1-e^{-2\tau}}\sigma_r^2. \tag{209}$$

Substituting (208) and (209) into (206) and setting the first-order partial derivatives $\partial\Lambda(n+1)/\partial M(n+1)$ and $\partial\Lambda(n+1)/\partial\mu(n+1)$ to zero, we obtain

$$M(n+1) = -\frac{1}{2\tau}\ln\frac{\mu(n+1)\left(\sigma_x^2\Lambda(n)+\sigma_v^2\right)\left(1-e^{-2\tau}\right)}{4\tau\left(1-\mu(n+1)\sigma_x^2\right)\sigma_r^2}, \tag{210}$$

$$\mu(n+1) = \frac{1-\frac{E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right]}{\Lambda(n)}}{(M(n+1)+2)\sigma_x^2+\frac{M(n+1)\sigma_v^2}{\Lambda(n)}-\frac{2\sigma_x^2E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right]}{\Lambda(n)}}. \tag{211}$$

Then, at each iteration, a pair of stationary points $M(n+1)$ and $u(n+1)$ can be obtained by jointly solving Eqs. (210) and (211). Based on Eqs. (210) and (211), it is difficult to find closed-form solutions for $M(n+1)$ and $\mu(n+1)$. Moreover, the stationary points from (210) and (211) lead to the global minimum of $\Lambda(n+1)$ only if the MSD is a convex function with respect to the tap-length and step-size. However, the convexity is difficult to be verified due to the complicated Hessian matrix. In the following, we find an approximate solution of $M(n)$ and $\mu(n)$ rather than explicitly solving (210) and (211).

By assuming that $M(n)$ is close to $M(n+1)$, we replace $M(n+1)$ by $M(n)$ in (211)

$$\mu(n+1) = \frac{1-\frac{E\left[\|\boldsymbol{c}''_{N-M(n)}\|_2^2\right]}{\Lambda(n)}}{(M(n)+2)\sigma_x^2+\frac{M(n)\sigma_v^2}{\Lambda(n)}-\frac{2\sigma_x^2E\left[\|\boldsymbol{c}''_{N-M(n)}\|_2^2\right]}{\Lambda(n)}}. \tag{212}$$

Thus, in each iteration $\mu(n+1)$ and $M(n+1)$ are obtained in an alternating manner by using (212) and (210). Next, we show that in this alternating manner, convergence condition is satisfied. Moreover, by removing the dependence in Eqs. (211) and (210) between each other, $\mu(n+1)$ in (212) and $M(n+1)$ in (210) are optimal solutions in terms of minimizing $\Lambda(n+1)$ given the other.

Combining (194), (196), and (200), we obtain

$$\Lambda(n) = E\left[\|\boldsymbol{\delta}_{M(n)}(n)\|_2^2\right] + E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right]. \tag{213}$$

Substituting (213) into (211), it is then straightforward to verify that $u(n+1)$ ensures the convergence of (201) according to the condition in [37]

$$0 < \mu(n+1) < \frac{2}{(M(n+1)+2)\sigma_x^2}. \tag{214}$$

Moreover, if there is no background noise $\left(\sigma_v^2 = 0\right)$ and the filter tap length is perfectly modeled $\left(\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2 = 0\right)$, the step size in (211) simplifies to

$$\mu(n+1) = \frac{1}{(M(n+1)+2)\sigma_x^2}, \tag{215}$$

which is consistent with the step-size that achieves the optimum convergence rate and adjustment in [46].

To analyze the behavior of $\mu(n+1)$ in (212), we take the second-order partial derivative of (201) with respect to $\mu(n+1)$:

$$\begin{aligned}\frac{\partial^2 \Lambda(n+1)}{\partial \mu^2(n+1)} =& 2M(n+1)\sigma_x^2\left(\sigma_x^2\Lambda(n)+\sigma_v^2\right) \\ &+ 4\sigma_x^4\left(\Lambda(n) - E\left[\|\boldsymbol{c}''_{N-M(n+1)}\|_2^2\right]\right).\end{aligned} \tag{216}$$

Based on (213), we know that

$$\frac{\partial^2 \Lambda(n+1)}{\partial \mu^2(n+1)} > 0, \tag{217}$$

which indicates that for any given tap length, MSD is a convex function in the step size parameter. Therefore, with tap-length $M(n)$, the step-size in (212) minimizes the MSD at the $(n+1)^{st}$ iteration. Similarly, the second-order partial derivative of (201) with respect to $M(n+1)$ leads to

$$\frac{\partial^2 \Lambda(n+1)}{\partial M^2(n+1)} = \frac{8\mu(n)\tau^2\sigma_x^2(1-\mu(n)\sigma_x^2)e^{-2M(n+1)\tau}\sigma_r^2}{1-e^{-2\tau}}. \tag{218}$$

For any step size that guarantees convergence (see (214)), it is straightforward to show

$$\frac{\partial^2 \Lambda(n+1)}{\partial M^2(n+1)} > 0, \tag{219}$$

which indicates that with a given step size, MSD is also a convex function in the tap length parameter. Therefore, with the step size $\mu(n+1)$, the tap length in (210) achieves the minimum MSD. So far, an optimal solution for the step size and tap length at each iteration is described by (212) and (210). However, the estimates of $\mu(n)$ and $M(n)$ still depend on $\Lambda(n)$. Next, we show how to estimate $\Lambda(n)$ in the $(n+1)^{st}$ iteration.

Based on the independence assumption between the filter input signal and the filter coefficients, the MSE of the LMS filter is expressed as (see also [81, 111])

$$E\left[e^2(n)\right] = \sigma_x^2\Lambda(n) + \sigma_v^2. \tag{220}$$

Combining (220), (212), and (210), the tap length and step size are obtained as follows:

$$\mu(n+1) = \frac{E\left[e^2(n)\right] - \sigma_v^2 - \sigma_x^2 E\left[\|\boldsymbol{c}''_{N-M(n)}\|_2^2\right]}{\sigma_x^2\Big(M(n)E\left[e^2(n)\right] - 2\sigma_x^2 E\left[\|\boldsymbol{c}''_{N-M(n)}\|_2^2\right]\Big)}, \tag{221}$$

$$M(n+1) = -\frac{1}{2\tau}\ln\frac{\mu(n+1)\left(1-e^{-2\tau}\right)E\left[e^2(n)\right]}{4\tau\left(1-\mu(n+1)\sigma_x^2\right)\sigma_r^2}. \tag{222}$$

In practice, the statistical average $E\left[e^2(n)\right]$ can be estimated recursively by its time average:

$$\overline{e^2(n)} = \rho\overline{e^2(n-1)} + (1-\rho)e^2(n), \tag{223}$$

where $0 < \rho < 1$ is the forgetting factor. Moreover, since the tap-length of a filter must be an integer, we choose to only keep the integer part of $M(n+1)$ after its update by Eq. (222). Finally, the entire adaptive algorithm is described sequentially by (193), (208), (221), (222), and (192), which is summarized in Table 9.

Table 9: Variable step-size and tap-length LMS algorithm.

Step 1: Compute estimate error $e(n) = d(n) - \boldsymbol{x}^T_{M(n)}(n)\boldsymbol{w}_{M(n)}(n)$.

Step 2: Compute the undermodeled channel gain $G_c = \frac{e^{-2M(n+1)\tau} - e^{-2N\tau}}{1-e^{-2\tau}}\sigma_r^2$.

Step 3: Estimate error power $\overline{e^2(n)} = \rho\overline{e^2(n-1)} + (1-\rho)e^2(n)$.

Step 4: Update step-size $\mu(n+1) = \frac{\overline{e^2(n)} - \sigma_v^2 - \sigma_x^2 G_c}{\sigma_x^2\left(M(n)\overline{e^2(n)} - 2\sigma_x^2 G_c\right)}$.

Step 5: Update tap-length $M(n+1) = -\frac{1}{2\tau}\ln\frac{\mu(n+1)\left(1-e^{-2\tau}\right)\overline{e^2(n)}}{4\tau(1-\mu(n+1)\sigma_x^2)\sigma_r^2}$.

Step 6: Update coefficients $\boldsymbol{w}_{M(n+1)}(n) = \begin{bmatrix} \boldsymbol{w}_{M(n)}(n) \\ \boldsymbol{0}_{M(n+1)-M(n)} \end{bmatrix} + \mu(n+1)e(n)\boldsymbol{x}_{M(n+1)}$.

4.3.3 Simulations

In this section, the performance of the proposed method is assessed via computer simulations. For comparison purposes, we also implemented the fixed tap-length LMS algorithm and the variable tap-length LMS algorithm in [111]. The setup of all the simulations is

similar to that in [111]: The impulse response was generated according to (190), which was a white Gaussian noise sequence with zero-mean and variance σ_r^2 of 0.01 weighted by an exponential decay profile. The impulse response length was $N = 1024$, and the envelope decay rate τ was 0.005. One realization of the unknown response is shown in Fig. 29. The filter input was a zero-mean i.i.d. Gaussian process with variance $\sigma_x^2 = 1$. The noise was another white Gaussian process with zero mean and variance σ_v^2 of 0.01. All the following results were obtained by averaging over 100 Monte Carlo trials.

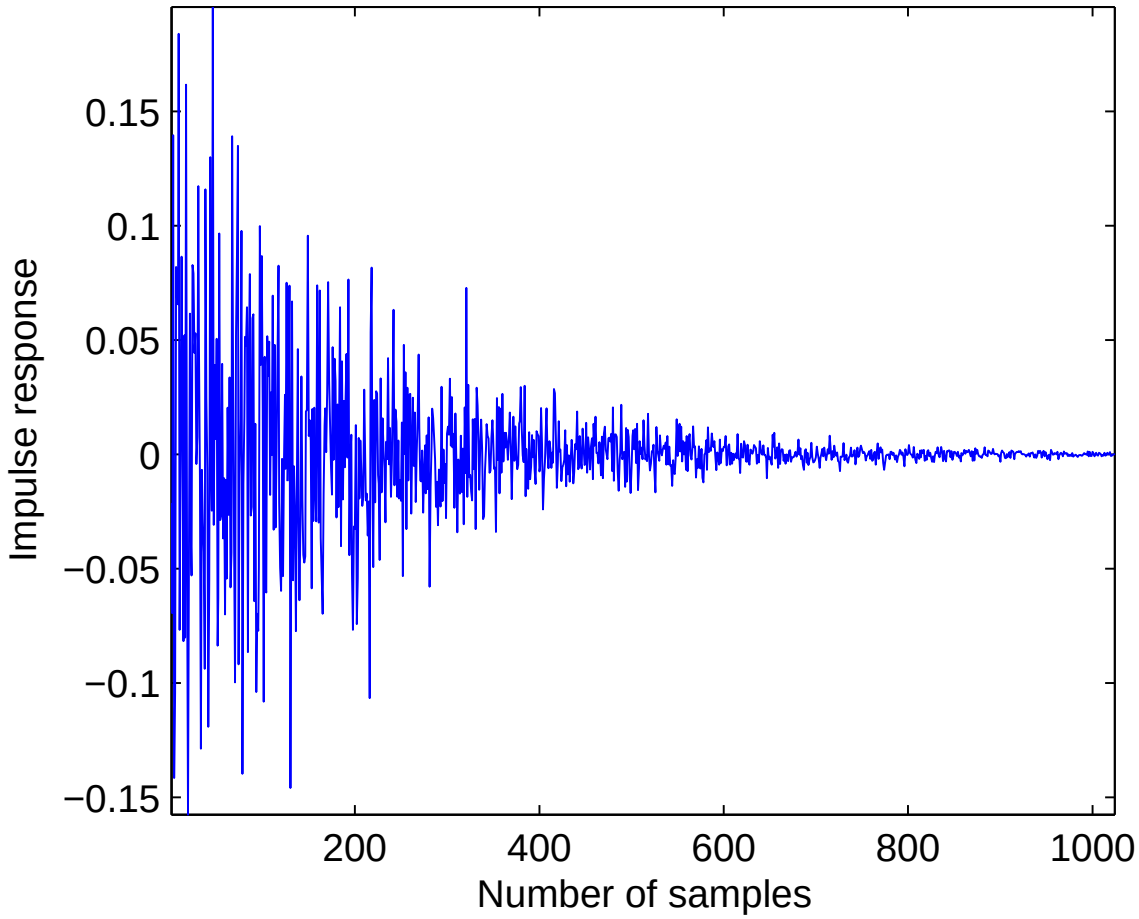
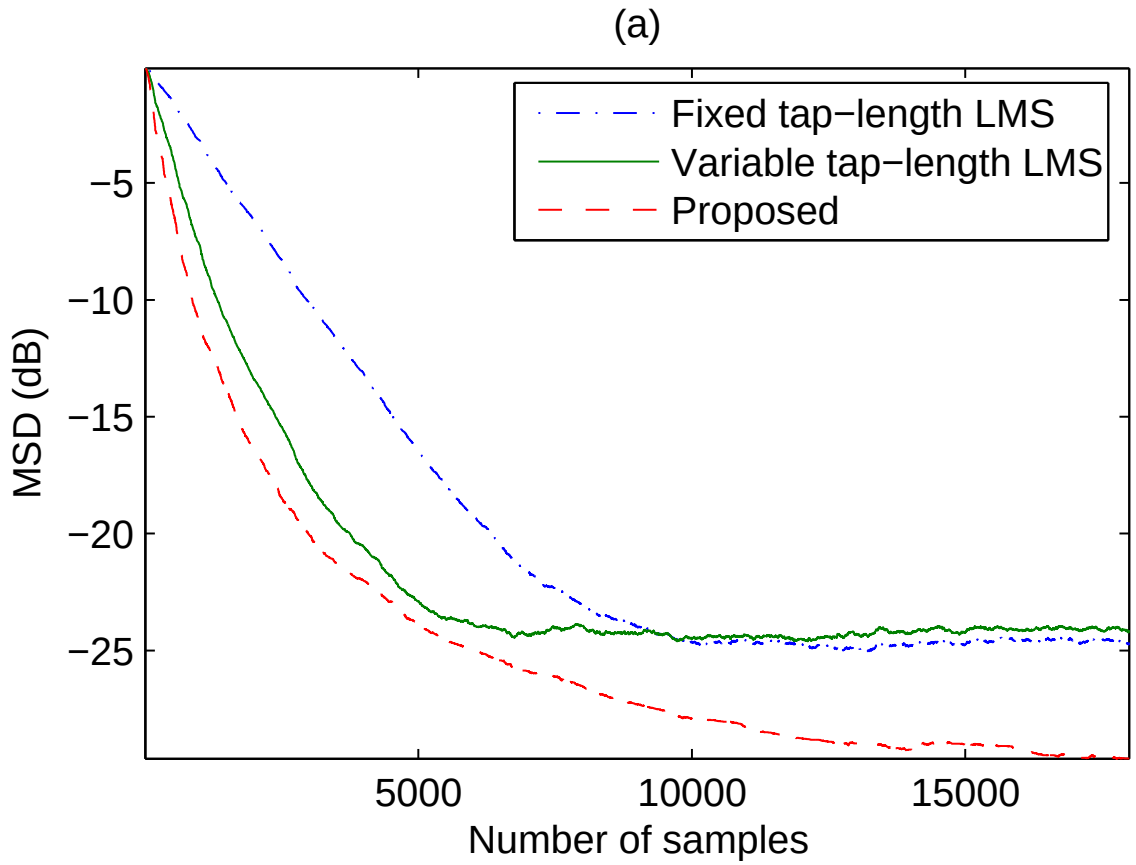


Figure 29: One realization of impulse response.

First, we evaluated the convergence performance of the proposed method. MSD and MSE curves with respect to the number of iterations are depicted with different types of LMS algorithms. The step size for the fixed tap-length LMS algorithm is set to 1/1026,

which corresponds to the optimal step-size in (215). For both the algorithm in [111] and the proposed method, the initial tap-length $M(0)$ was chosen as 20 and the forgetting factor in (220) was 0.99. The MSDs are shown in Fig. 30(a). It is seen that the algorithm in [111] converged faster than the fixed tap-length LMS algorithm due to the variable step size, and both exhibited similar steady-state MSDs. The proposed method further improved the convergence rate and achieved lower steady-state MSD, due to the fact that MSD is minimized in terms of both the tap length and step size at each iteration. The MSEs are shown in Fig. 30(b), which also validates the advantages of the proposed method in terms of both convergence rate and steady-state performance. We point out that the consistency between the MSD and MSE results is in agreement with the theoretical analysis in (220). Both of them are presented here since different applications may focus on different criteria. For instance, MSD is more suitable in channel estimation, whereas MSE is preferable in echo cancellation applications.



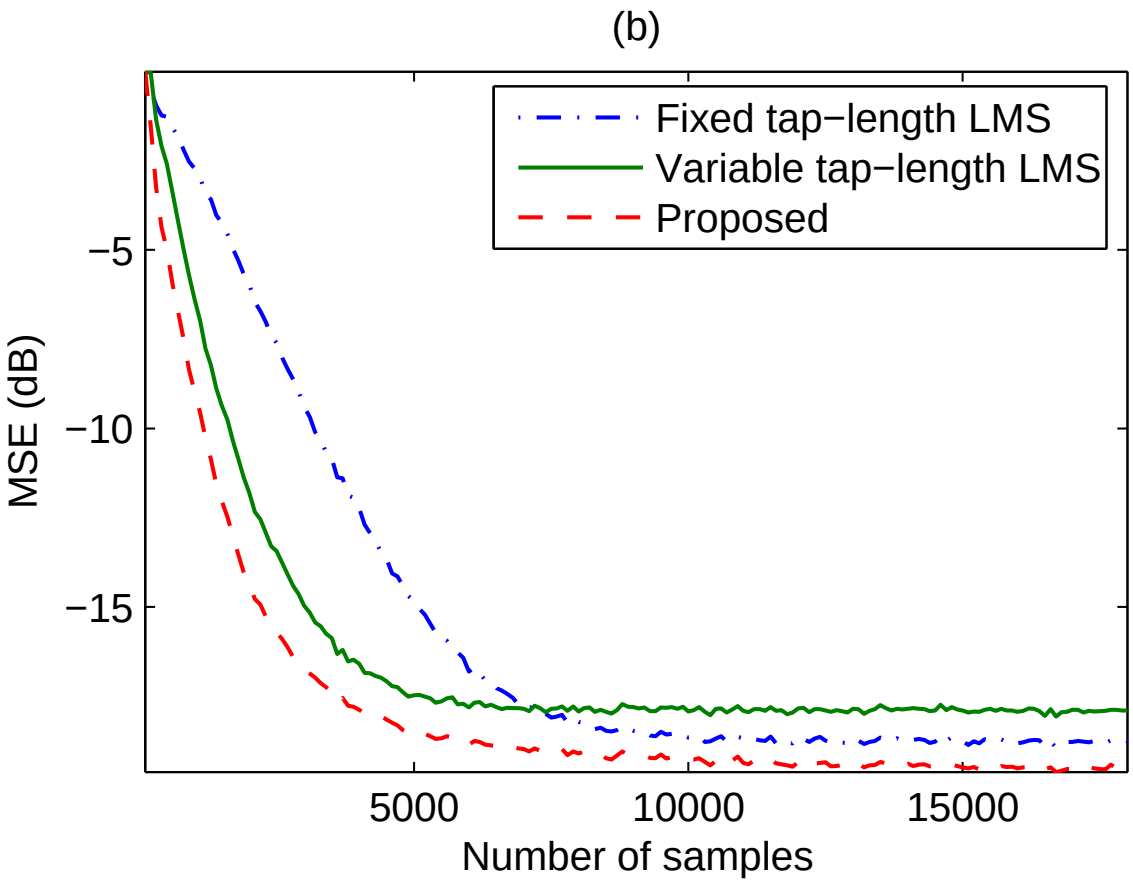


Figure 30: Comparison of convergence performance with different LMS algorithms: (a) MSD; (b) MSE.

The values of tap length and step size of the proposed method and the method in [111] are shown in Fig. 31. The step sizes are shown in log scale for demonstration purposes. For the method in [111], it is seen that the tap length saturates at around 800. Similar variability is observed for the step size, since the step size simply follows $\mu(n) = \mu'/(M(n-1)+\delta)\sigma_x^2$, with the parameters δ and μ' being set to 5 and 0.5, respectively. Comparatively, the step size in the proposed method saturates at a smaller value, which provides finer coefficients update. Therefore, the proposed method achieves better performance (see Fig. 30).

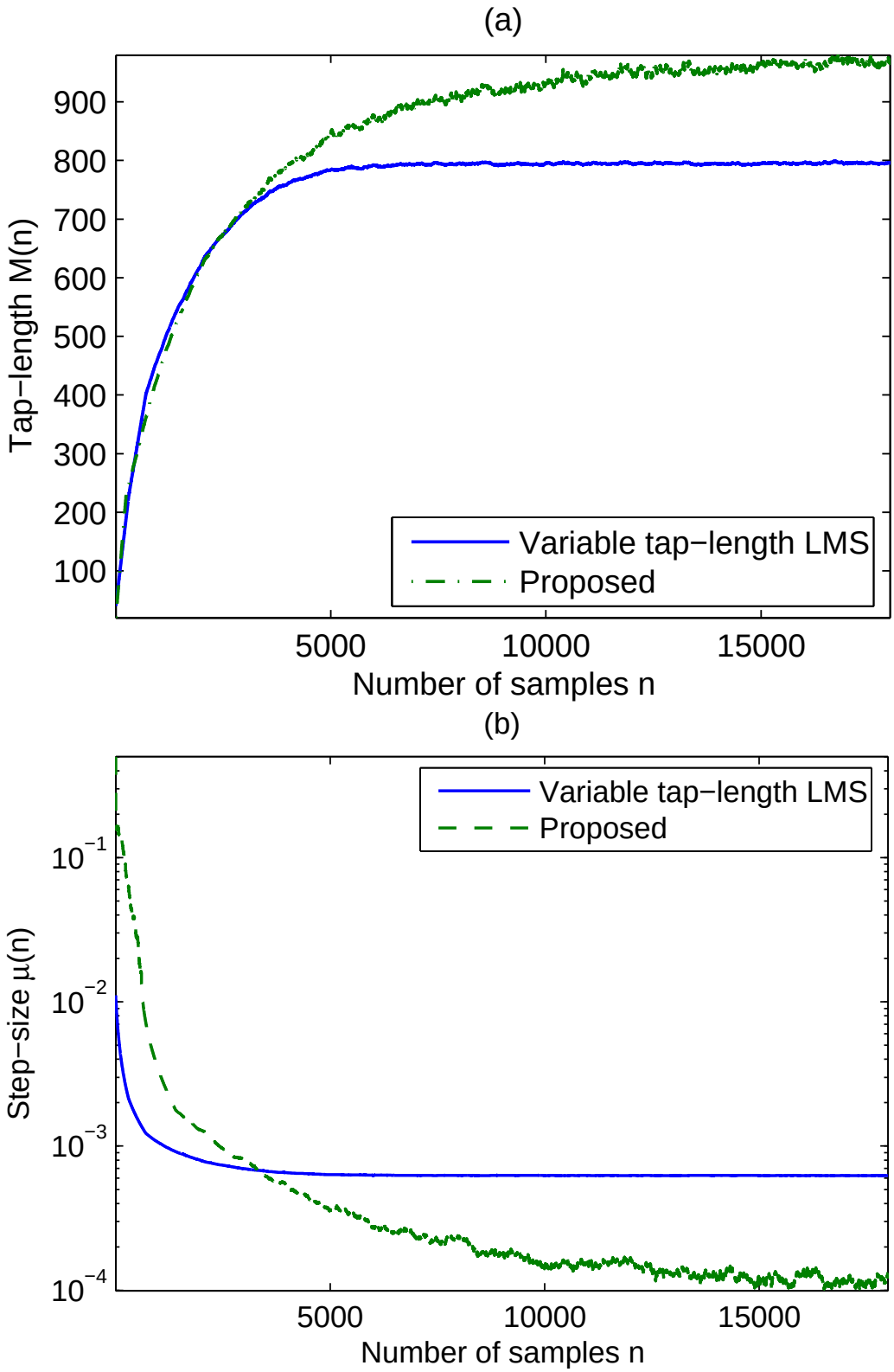


Figure 31: Comparison of tap-length and step-size with different LMS algorithms: (a) tap-length $M(n)$; (b) step-size $\mu(n)$.

Finally, we evaluate the performance of the proposed method with respect to the initial value. The steady-state MSDs with different initial tap-length are shown in Table 10. It can be observed that with a wide range of the initial tap-length, the MSD converges to the value that achieves an effective modeling of the significant energy within the impulse response. Thus, we claim that the proposed algorithm is robust to the selection of initial tap-length.

Table 10: Steady-state MSDs and MSEs with different initial tap-length.

$M(0)$	20	50	100	200	500	800
MSD(∞) (dB)	-29.4	-29.7	28.6	-29.3	-30.1	-29.4
MSE(∞) (dB)	19.6	19.7	19.2	19.5	19.9	19.7

CHAPTER V

CONCLUSIONS

This dissertation aimed to design AECs in the presence of nonlinearity in the echo path and the controller logic working associatively with AECs. Specifically, to remove the nonlinear acoustic echo, we investigated three different structures: predistortion linearization, cascade structure, and post-processing technique. The loudspeaker linearization improved both the far-end and near-end talkers' experience. In cascade structure, a pseudo-MSC function-based method was proposed to identify the nonlinear acoustic echo path. This method decoupled the identification of the nonlinear part from linear part in a Hammerstein system, and thus guaranteed system stability. The post-procssing method employed a residual echo suppressor to enhance the convergence rate of filter adaptation, which also combined the echo cancellation with noise reduction. Focusing on the issues of convergence rate and computational complexity, we also proposed other methods to combat nonlinear echo in the system, such as cascade NAECs with a shortening filter and AECs in the presence of multiple nonlinearities. For the control logic design, we investigated the DTD design and learning-rate control. To detect double-talk, a mutual information-based criterion was introduced to construct a DTD decision statistic, which is applicable to both the linear and nonlinear scenarios. Furthermore, a generalized mutual information-based statistic was suggested to extend the DTD design into stereophonic systems, which facilitates threshold selection and reduces complexity. To adjust the learning rate, we proposed a variable step size and tap length LMS algorithm for the systems with the impulse response exhibiting an exponential decay envelope. The proposed method achieved both better steady-state performance and faster convergence rate. Throughout this research, computer simulations and real speech data experiments were conducted to demonstrate the performance of the proposed algorithms.

5.1 Contributions

We summarize below primary contributions of this dissertation:

- Proposed a predistortion linearization structure for nonlinear acoustic echo cancellation to improve the near-end's listening experience.
- Proposed a peudo-MSC function-based NAEC design to guarantee system stability and improve filter convergence rate.
- Proposed an NRES without explicitly estimating the power spectral density of residual echo signal to improve the convergence rate.
- Proposed an efficient NAEC design by incorporating a shortening filter to improve convergence rate and reduce computational complexity.
- Proposed a Hammerstein-Wiener model-based NAEC to handle multiple nonlinearities in acoustic echo cancellation systems.
- Proposed a mutual information-based DTD design to enhance robustness to system nonlinearity.
- Proposed a generalized information-based DTD for NAECs in stereophonic acoustic echo cancellation systems.
- Proposed a variable step size and tap length LMS algorithm to improve steady-state performance and convergence rate.

5.2 Suggestions for Future Research

The following is a list of interesting research topics that can be pursued as extensions of this dissertation:

- Design the NAEC or NRES by incorporating the psychoacoustic concepts and models.
- Develop the NAEC or NRES for multi-channel acoustic echo cancellation systems.

- Develop AECs and DTDs for general nonlinearities, such as nonlinearity with memory effects.
- Develop DTDs based on statistical models.

APPENDIX A

CONVERGENCE ANALYSIS OF NLMS-BASED NAEC WITH SHORTENING FILER

The convergence of adaptive Hammerstein system has been well studied in [49, 50]. However, the introduction of shortening filter makes the convergence analysis more challenging. Based on the following assumptions, we discuss the convergence behavior of the residual echo power for NLMS adaptation. Denote the optimal solutions of filters $\boldsymbol{h}(n)$, $\boldsymbol{w}(n)$, and $u(\cdot;\boldsymbol{\theta}(n))$ by $\boldsymbol{h}$, $\boldsymbol{w}$, and $\boldsymbol{\theta}$, respectively.

- The nonlinearity of loudspeaker and linear room impulse response are time invariant.
- There is no mismatch between the nonlinear model and the loudspeaker nonlinearity.
- The optimal solution $\boldsymbol{h}$ approximates the first L_h taps of $\tilde{\boldsymbol{g}}$.
- There is no noise in the microphone received signal, i.e., $v(n) = 0$.
- Double talk situation does not exist before the filters converge.
- The input signal $s(n)$ is wide-sense stationary.

Define

$$\tilde{\mathbf{S}}(n) = [\mathbf{S}(n)\ \mathbf{S}_{\Delta}(n)]\,, \quad \hat{\tilde{\boldsymbol{g}}}(n) = \left[\hat{\boldsymbol{g}}^T(n)\ \hat{\boldsymbol{g}}_{\Delta}^T(n)\right]^T, \tag{224}$$

where

$$\begin{aligned}
\hat{\boldsymbol{g}}(n) &= [\hat{g}_0(n), \hat{g}_1(n), ..., \hat{g}_{L_h-1}(n)]^T\,, \\
\hat{\boldsymbol{g}}_{\Delta}(n) &= [\hat{g}_{L_h}(n), \hat{g}_{L_h+1}(n), ..., \hat{g}_{L_h+\Delta-1}(n)]^T\,, \\
\mathbf{S}_{\Delta}(n) &= [\boldsymbol{b}(n-L_h), \boldsymbol{b}(n-L_h-1), ..., \boldsymbol{b}(n-L_h-\Delta+1)]\,.
\end{aligned}$$

The reference signal $\hat{d}(n)$ and the estimated signal $\hat{z}(n)$ can be expressed as

$$\hat{d}(n) = \boldsymbol{\theta}^T \tilde{\mathbf{S}}(n)\hat{\tilde{\boldsymbol{g}}}(n) = \boldsymbol{\theta}^T \mathbf{S}(n)\hat{\boldsymbol{g}}(n) + \boldsymbol{\theta}^T \mathbf{S}_\Delta(n)\hat{\boldsymbol{g}}_\Delta(n), \tag{225}$$

$$\hat{z}(n) = \hat{\boldsymbol{\theta}}^T(n)\mathbf{S}(n)\hat{\boldsymbol{h}}(n). \tag{226}$$

Then, the error signal can be written as

$$e(n) = \hat{d}(n) - \hat{z}(n) = \boldsymbol{\theta}^T \mathbf{S}(n)\hat{\boldsymbol{g}}(n) + \boldsymbol{\theta}^T \mathbf{S}_\Delta(n)\hat{\boldsymbol{g}}_\Delta(n) - \hat{\boldsymbol{\theta}}^T(n)\mathbf{S}(n)\hat{\boldsymbol{h}}(n). \tag{227}$$

Define

$$e_\theta(n) \triangleq \boldsymbol{\theta}^T \mathbf{S}(n)\boldsymbol{h} - \hat{\boldsymbol{\theta}}^T(n)\mathbf{S}(n)\boldsymbol{h} = \boldsymbol{\epsilon}_\theta^T(n)\boldsymbol{u}(n), \tag{228}$$

where $\boldsymbol{\epsilon}_\theta(n)$ is the nonlinear coefficients error $\boldsymbol{\epsilon}_\theta(n) = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(n)$. Note that $e_\theta(n)$ can be interpreted as the estimation error produced by the nonlinear AEC filter under the assumption of perfect linear coefficients, i.e., $\hat{\boldsymbol{h}}(n) = \boldsymbol{h}$ and $\hat{\boldsymbol{w}}(n) = \boldsymbol{w}$. Similarly, define the tracking errors caused by imperfect $\boldsymbol{h}$ or $\boldsymbol{w}$, respectively

$$e_h(n) = \boldsymbol{\epsilon}_h^T(n)\boldsymbol{x}(n), \tag{229}$$

$$e_w(n) = -\boldsymbol{\epsilon}_w^T(n)\boldsymbol{y}(n), \tag{230}$$

where $\boldsymbol{\epsilon}_h(n)$ and $\boldsymbol{\epsilon}_w(n)$ are errors of the coefficients of AEC filter $\boldsymbol{h}(n)$ and shortening filter $\boldsymbol{w}(n)$, respectively

$$\boldsymbol{\epsilon}_h(n) = \boldsymbol{h} - \hat{\boldsymbol{h}}(n), \tag{231}$$

$$\boldsymbol{\epsilon}_w(n) = \boldsymbol{w} - \hat{\boldsymbol{w}}(n). \tag{232}$$

Note that the desired $\boldsymbol{w}$ should give $\hat{\boldsymbol{g}}_\Delta \approx \boldsymbol{0}$. Thus, the estimation error of $\boldsymbol{w}$ leads to an imperfect $\tilde{\boldsymbol{g}}(n)$. Therefore, an alternative way to express the error signal due to the inaccurate estimate of $\boldsymbol{w}$ is given by

$$e_w(n) = -\boldsymbol{\theta}^T \mathbf{S}(n)\boldsymbol{\epsilon}_g(n) - \boldsymbol{\theta}^T \mathbf{S}_\Delta(n)\boldsymbol{\epsilon}_{g_\Delta}(n), \tag{233}$$

where $\boldsymbol{\epsilon}_g = \boldsymbol{g} - \hat{\boldsymbol{g}}(n) \approx \boldsymbol{h} - \hat{\boldsymbol{g}}(n)$, and $\boldsymbol{\epsilon}_{g_\Delta} = \boldsymbol{g}_\Delta - \hat{\boldsymbol{g}}_\Delta(n)$.

Finally, the error signal in (227) can be approximated by the first order terms

$$e(n) \approx e_\theta(n) + e_h(n) + e_w(n) + \boldsymbol{\theta}^T \mathbf{S}_\Delta(n)\boldsymbol{g}_\Delta, \tag{234}$$

where the second order terms are neglected. It is pointed out that this approximation ignores the interactions between update of different parameters. This is suitable when each coefficients vector gets more and more converged. We will use some simulation results to show that the solution found from the proposed method is not far from the "perfect" one. Based on the assumption that $\boldsymbol{h} \approx \boldsymbol{g}$ and (136) we obtain

$$e_{res}(n) \approx \boldsymbol{x}_{\Delta}^{T}(n)\boldsymbol{g}_{\Delta} = \boldsymbol{\theta}^{T}\mathbf{S}_{\Delta}(n)\boldsymbol{g}_{\Delta}, \tag{235}$$

where $\boldsymbol{x}_{\Delta}(n) = [x(n-L_h), ..., x(n-L_h-\Delta+1)]^{T}$. Combining (234) and (235), we obtain

$$e(n) \approx e_{\theta}(n) + e_{h}(n) + e_{w}(n) + e_{res}(n). \tag{236}$$

Hence, the residual error is decoupled into four terms, where the first three terms are the errors caused by the estimation errors of the individual unknown coefficients, and the last one is due to the imperfect shortening. Note that, during the update of each filter coefficients, the quadratic form of (114) makes it difficult to analyze the convergence. However, the decoupling of the residual error allows us to analyze the algorithm's performance, because each decoupled error term is generated by the coefficients estimation error of only one filter.

In the following, we discuss the convergence characteristic of the algorithm. Based on (236), the MSE can be expressed as

$$\begin{aligned} J(n) &= E\left[e^{2}(n)\right] \\ &\approx E\left[e_{w}^{2}(n)\right] + E\left[e_{h}^{2}(n)\right] + E\left[e_{\theta}^{2}(n)\right] + E\left[e_{res}^{2}(n)\right] \\ &+2E\left[e_{w}(n)e_{h}(n)\right]+2E\left[e_{w}(n)e_{\theta}(n)\right]+2E\left[e_{h}(n)e_{\theta}(n)\right] \\ &+2E\left[e_{w}(n)e_{res}(n)\right]. \end{aligned} \tag{237}$$

Consider only $\boldsymbol{w}$ as an unknown parameter, which is updated by (119). Combining (232) and (119), we obtain

$$\begin{aligned} \boldsymbol{\epsilon}_{w}(n+1) &= \boldsymbol{w} - \hat{\boldsymbol{w}}(n+1) \\ &= \left[\mathbf{I}_{L_w} - \frac{\mu_w}{\|\boldsymbol{y}(n)\|_{2}^{2}}\boldsymbol{y}(n)\boldsymbol{y}^{T}(n)\right]\boldsymbol{\epsilon}_{w}(n), \end{aligned} \tag{238}$$

where $e(n) = e_w(n) = -\boldsymbol{\epsilon}_w^T(n)\boldsymbol{y}(n)$. According to the direct-averaging method [42, p. 259], the solution of the difference equation (238), operating under the assumption of a small step-size, is close to the solution of the following stochastic difference equation

$$\boldsymbol{\epsilon}_w(n+1) = E\left[\mathbf{I}_{L_w} - \frac{\mu_w}{\|\boldsymbol{y}(n)\|_2^2}\boldsymbol{y}(n)\boldsymbol{y}^T(n)\right]\boldsymbol{\epsilon}_w(n). \tag{239}$$

Assuming that the size of $\boldsymbol{y}(n)$ is long enough, we obtain $E\left[\frac{\boldsymbol{y}(n)\boldsymbol{y}^T(n)}{\|\boldsymbol{y}(n)\|_2^2}\right] \approx \frac{E\left[\boldsymbol{y}(n)\boldsymbol{y}^T(n)\right]}{L_w E[y^2(m)]}$. Denote $\sigma_y^2 = L_w E\left[y^2(m)\right]$ and $\mathbf{R}_y = E\left[\boldsymbol{y}(n)\boldsymbol{y}^T(n)\right]$. By applying the eigenvalue decomposition on $\mathbf{R}_y$, we obtain

$$\mathbf{R}_y = \mathbf{Q}_y\mathbf{\Lambda}_y\mathbf{Q}_y^T, \tag{240}$$

where $\mathbf{Q}_y$ is a unitary matrix and $\mathbf{\Lambda}_y$ is a diagonal matrix consisting of the eigenvalues $\lambda_y^{(i)},\ i = 1, 2, ..., L_w$. Define $\bar{\boldsymbol{\epsilon}}_w(n) = \mathbf{Q}_y^T\boldsymbol{\epsilon}_w(n)$. We rewrite (239) as

$$\bar{\boldsymbol{\epsilon}}_w(n+1) = \left(\mathbf{I} - \frac{\mu_w}{\sigma_y^2}\mathbf{\Lambda_y}\right)\bar{\boldsymbol{\epsilon}}_w(n). \tag{241}$$

For the ith entry of $\bar{\boldsymbol{\epsilon}}_w(n)$ we obtain

$$\bar{\epsilon}_w^{(i)}(n) = \left(1 - \frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n \bar{\epsilon}_w^{(i)}(0). \tag{242}$$

Since $\boldsymbol{\epsilon}_w(n)$ is independent of $\boldsymbol{y}(n)$, we may replace the stochastic product $\boldsymbol{y}(n)\boldsymbol{y}^T(n)$ by its expected value and hence write

$$E\left[e_w^2(n)\right] = E\left[\boldsymbol{\epsilon}_w^T(n)\boldsymbol{y}(n)\boldsymbol{y}^T(n)\boldsymbol{\epsilon}_w(n)\right] = E\left[\boldsymbol{\epsilon}_w^T(n)\mathbf{R}_y\boldsymbol{\epsilon}_w(n)\right]. \tag{243}$$

Using (241) and (242), we may express $E\left[e_w^2(n)\right]$ in (243) as

$$\begin{aligned} E\left[e_w^2(n)\right] &= E\left[\bar{\boldsymbol{\epsilon}}_w^T(n)\mathbf{\Lambda}_y\bar{\boldsymbol{\epsilon}}_w(n)\right] = \sum_{i=1}^{L_w}\lambda_y^{(i)}E\left[\left|\bar{\epsilon}_w^{(i)}(n)\right|^2\right] \\ &= \sum_{i=1}^{L_w}\lambda_y^{(i)}\left(1 - \frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^{2n}\left(\bar{\epsilon}_w^{(i)}(0)\right)^2. \end{aligned} \tag{244}$$

Similarly, the MSEs of the estimates of $\boldsymbol{h}$ and $\boldsymbol{\theta}$ are obtained respectively

$$E\left[e_h^2(n)\right] = \sum_{i=1}^{L_h}\lambda_x^{(i)}\left(1 - \frac{\mu_h}{\sigma_x^2}\lambda_x^{(i)}\right)^{2n}\left(\bar{\epsilon}_h^{(i)}(0)\right)^2, \tag{245}$$

$$E\left[e_\theta^2(n)\right] = \sum_{i=1}^{K}\lambda_u^{(i)}\left(1 - \frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(i)}\right)^{2n}\left(\bar{\epsilon}_\theta^{(i)}(0)\right)^2, \tag{246}$$

where $\lambda_x^{(i)}$ and $\lambda_u^{(i)}$ are the ith eigenvalues of the auto-correlation matrices $\mathbf{R}_x = E\left[\boldsymbol{x}(n)\boldsymbol{x}^T(n)\right]$ and $\mathbf{R}_u = E\left[\boldsymbol{u}(n)\boldsymbol{u}^T(n)\right]$, respectively. σ_x^2, σ_u^2, $\bar{\epsilon}_h^{(i)}$, and $\bar{\epsilon}_\theta^{(i)}$ are defined in a similar way as σ_y^2 and $\bar{\epsilon}_w^{(i)}$.

For the cross terms in (237), we assume that different coefficients are independent on each other. Following the similar procedure, we obtain

$$E[e_w(n)e_h(n)] = -\sum_{i=1}^{L_w}\sum_{j=1}^{L_h} R_{yx}(i,j)\bar{\epsilon}_w^{(i)}(0)\bar{\epsilon}_h^{(j)}(0)\cdot\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n\left(1-\frac{\mu_h}{\sigma_x^2}\lambda_x^{(j)}\right)^n, \tag{247}$$

$$E[e_w(n)e_\theta(n)] = -\sum_{i=1}^{L_w}\sum_{j=1}^{K} R_{yu}(i,j)\bar{\epsilon}_w^{(i)}(0)\bar{\epsilon}_\theta^{(j)}(0)\cdot\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n\left(1-\frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(j)}\right)^n, \tag{248}$$

$$E[e_h(n)e_\theta(n)] = \sum_{i=1}^{L_h}\sum_{j=1}^{K} R_{xu}(i,j)\bar{\epsilon}_h^{(i)}(0)\bar{\epsilon}_\theta^{(j)}(0)\cdot\left(1-\frac{\mu_h}{\sigma_x^2}\lambda_x^{(i)}\right)^n\left(1-\frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(j)}\right)^n, \tag{249}$$

$$E\left[e_w(n)e_{res}(n)\right] = -\sum_{i=1}^{L_w} r_{yx}(i)\bar{\epsilon}_w^{(i)}(0)\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n, \tag{250}$$

where $R_{yx}(i,j)$, $R_{yu}(i,j)$, and $R_{xu}(i,j)$ are, respectively, the (i,j)th entries of matrices $\mathbf{R}_{yx} = \mathbf{Q}_w^T E\left[\boldsymbol{y}(n)\boldsymbol{x}^T(n)\right]\mathbf{Q}_h$, $\mathbf{R}_{yu} = \mathbf{Q}_w^T E\left[\boldsymbol{y}(n)\boldsymbol{u}^T(n)\right]\mathbf{Q}_\theta$, and $\mathbf{R}_{xu} = \mathbf{Q}_h^T E\left[\boldsymbol{x}(n)\boldsymbol{u}^T(n)\right]\mathbf{Q}_\theta$; $r_{yx}(i)$ is the ith entry of the vector $\boldsymbol{r}_{yx} = \mathbf{Q}_w^T E\left[\boldsymbol{y}(n)\boldsymbol{x}_\Delta^T(n)\right]\boldsymbol{g}_\Delta$. Therefore, the MSE in (237) can be written in eq. (251) [cf. (244) - (250)].

$$\begin{aligned}
J(n) \approx \lambda_{\min} &+ \sum_{i=1}^{L_w}\lambda_y^{(i)}\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^{2n}\left(\bar{\epsilon}_w^{(i)}(0)\right)^2 \\
&+ \sum_{i=1}^{L_h}\lambda_x^{(i)}\left(1-\frac{\mu_h}{\sigma_x^2}\lambda_x^{(i)}\right)^{2n}\left(\bar{\epsilon}_h^{(i)}(0)\right)^2 \\
&+ \sum_{i=1}^{K}\lambda_u^{(i)}\left(1-\frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(i)}\right)^{2n}\left(\bar{\epsilon}_\theta^{(i)}(0)\right)^2 \\
&- \sum_{i=1}^{L_w}\sum_{j=1}^{L_h} R_1(i,j)\bar{\epsilon}_w^{(i)}(0)\bar{\epsilon}_h^{(j)}(0)\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n\left(1-\frac{\mu_h}{\sigma_x^2}\lambda_x^{(j)}\right)^n \\
&- \sum_{i=1}^{L_w}\sum_{j=1}^{K} R_{yu}(i,j)\bar{\epsilon}_w^{(i)}(0)\bar{\epsilon}_\theta^{(j)}(0)\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n\left(1-\frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(j)}\right)^n \\
&+ \sum_{i=1}^{L_h}\sum_{j=1}^{K} R_{xu}(i,j)\bar{\epsilon}_h^{(i)}(0)\bar{\epsilon}_\theta^{(j)}(0)\left(1-\frac{\mu_h}{\sigma_x^2}\lambda_x^{(i)}\right)^n\left(1-\frac{\mu_\theta}{\sigma_u^2}\lambda_u^{(j)}\right)^n \\
&- \sum_{i=1}^{L_w} r_{yx}(i)\bar{\epsilon}_w^{(i)}(0)\left(1-\frac{\mu_w}{\sigma_y^2}\lambda_y^{(i)}\right)^n.
\end{aligned} \tag{251}$$

According to (251) we know that the convergence rate depends on the eigenvalues $\lambda_y^{(i)},\ i = 1, 2, .., L_w,\ \lambda_x^{(i)},\ i = 1, 2, .., L_h$, and $\lambda_u^{(i)},\ i = 1, 2, .., K$, and the smallest eigenvalue dominates the convergence rate. The step sizes should be small enough to guarantee the convergence, and the selection of step size depends on the statistical properties of signals. Different from other AECs, we notice that as time goes on, there is a residual echo power $\lambda_{\min}$ which is due to the imperfect shortening filter.

APPENDIX B

ITU-T G.167 AEC TEST PROCEDURE

B.1 Weighted terminal coupling lossCsingle talk (TCLwst)

Step 1: All the AEC functional units are initially reset and then enabled.

Step 2: A signal is applied at R_{in} for a sufficient time (to be defined, under study) so that the different functional units (in particular the acoustic echo canceller) reach their steady states. No other speech signal than the acoustic return from the loudspeaker(s) is applied to the microphone(s).

Step 3: Make an electrical measurement of the signal at S_{out}. The value TCLwst is the difference (in dB) between the signal level before the enabling of the AEC and the signal level at this step in the test.

B.2 Weighted terminal coupling loss C double talk (TCLwdt)

Step 1: The AEC is firstly operated as in the test of TCLwst (steps 1 and 2).

Step 2: After the echo loss has attained TCLwst, an acoustic signal simulating the local user's speech is applied at the Sin point for 2 seconds.

Step 3: The processing unit is frozen, and then the simulated local speech is removed.

Step 4: Make an electrical measurement of the signal at S_{out}. The value TCLwdt is the difference (in dB) between the signal level before the enabling of the AEC and the signal level at this step in the test.

B.3 Terminal coupling loss during echo path variation (TCLwpv)

Step 1: The AEC is firstly operated as in the test of TCLwst (steps 1 and 2).

Step 2: After TCLwst has attained its recommended value, a simulated or real echo path variation is applied for 5 seconds (means to produce echo path variations are under

study).

Step 3: At the end of the echo path variation, the processing unit is frozen, and the signal level at S_{out} is measured. The value TCLwpv is the difference (in dB) between the signal level before the enabling of the AEC and the signal level at this step in the test.

B.4 Initial convergence time (Tic)

Step 1: All the AEC functional units are initially reset and then enabled.

Step 2: A signal is applied at R_{in} and a timer is started.

Step 3: After 1 second, the processing unit is frozen.

Step 4: Make an electrical measurement of the signal at S_{out}. The time interval specified in step 3 is called Tic.

B.5 Recovery time after double talk (Trdt)

Step 1: The AEC is firstly operated as in the test of TCLwst (steps 1 and 2).

Step 2: After TCLwst has attained its recommended value, the signal applied at R_{in} is cut off and a signal simulating the local user's speech is applied at the S_{in} point for 2 seconds.

Step 3: The received signal is again applied at R_{in}, and after 2 seconds the signal simulating the local user's speech is cut off; then a timer is started.

Step 4: After 1 second the timer is stopped and the processing unit is frozen.

Step 5: The electric signal level at S_{out} is measured. The time interval specified in step 4 is called Trdt.

B.6 Recovery time after echo path variation (Trpv)

Step 1: The AEC is firstly operated as in the test of TCLwst (steps 1 and 2).

Step 2: After TCLwst has attained its recommended value, a simulated or real echo path variation is applied during 5 seconds (means to produce echo path variations are under study).

Step 3: At the end of the echo path variation a timer is started.

Step 4: After 1 second, the processing unit is frozen and the signal level at S_{out} is measured. The time interval specified in this step of the test is called Trpv.

REFERENCES

[1] Al-Dhahir, N. and Cioffi, J., "Efficiently computed reduced-parameter input-aided MMSE equalizersfor ML detection: a unified approach," *IEEE Trans. on Information Theory*, vol. 42, pp. 903–915, May 1996.

[2] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *Journal of Acoustic Society America*, vol. 65, pp. 943–950, 1979.

[3] Ang, W.-P. and Farhang-Boroujeny, B., "A new class of gradient adaptive step-size lms algorithms," *IEEE Trans. on Signal Processing*, vol. 49, pp. 805–810, Apr. 2001.

[4] Bai, E. W., "An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems," *Automatica*, vol. 34, pp. 333–338, Mar. 1998.

[5] Bai, E., "A blind approach to the Hammerstein–Wiener model identification," *Automatica*, vol. 38, pp. 967–979, Jun. 2002.

[6] Benesty, J., Gaensler, T., Morgan, D. R., Sondhi, M. M., and Gay, S. L., *Advances in Network and Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2001.

[7] Benesty, J., Morgan, D. R., and Cho, J. H., "A new class of doubletalk detector based on cross-correlation," *IEEE Trans. on Speech and Audio Processing*, vol. 8, pp. 168–172, Mar. 2000.

[8] Benesty, J., Rey, H., Vega, L. R., and Tressens, S., "A nonparametric vss nlms algorithm," *IEEE Signal Processing Letters*, vol. 13, pp. 581–584, Oct. 2006.

[9] Bershad, N. J. and Tourneret, J.-Y., "Echo cancellation-a likelihood ratio test for double-talk versus channel change," *IEEE Trans. on Signal Processing*, vol. 54, pp. 4572–4581, Dec. 2006.

[10] Birkett, A. N. and Goubran, R. A., "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, New York), pp. 103–106, Oct. 1995.

[11] Breining, C., Dreiscitel, P., Hänsler, E., Mader, A., Nitsch, B., Puder, H., Schertler, T., Schmidt, G., and Tilp, J., "Acoustic echo control," *IEEE Signal Processing Magazine*, vol. 16, pp. 42–69, Jul. 1999.

[12] Buchner, H., Benesty, J., Gänsler, T., and Kellermann, W., "Robust extended multidelay filter and double-talk detector for acoustic echo cancellation," *IEEE Trans. on Speech and Audio Processing*, vol. 14, pp. 1633–1644, Sept. 2006.

[13] Cadzow, J. and Solomon Jr, O., "Linear modeling and the coherence function," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 19–28, Jan. 1987.

[14] Carter, G., "Coherence and time delay estimation," *Proceedings of the IEEE*, vol. 75, pp. 236–255, Feb. 1987.

[15] Celka, P., Bershad, N., and Vesin, J., "Stochastic gradient identification of polynomial Wiener systems: analysis and application," *IEEE Trans. on Signal Processing*, vol. 49, pp. 301–313, Feb. 2001.

[16] Cho, J. H., Morgan, D. R., and Benesty, J., "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Trans. on Speech and Audio Processing*, vol. 7, pp. 718–724, Nov. 1999.

[17] Costa, J.-P., Lagrange, A., and Arliaud, A., "Acoustic echo cancellation using nonlinear cascade filters," in *Proc. IEEE ICASSP*, (Hong Kong, China), pp. 389–392, Apr. 2003.

[18] Costa, M., Bermudez, J., and Bershad, N., "Statistical analysis of the lms algorithm with a zero-memory nonlinearity after the adaptive filter," in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing*, (Phoenix, AZ), pp. 1661–1664, Mar. 1999.

[19] Cover, T. and Thomas, J., *Elements of Information Theory.* Wiley New York, 1991.

[20] Dai, H. and Zhu, W. P., "Compensation of loudspeaker nonlinearity in acoustic echo cancellation using raised-cosine function," *IEEE Trans. on Circuits and Systems II: Express Briefs*, vol. 3, pp. 1190–1194, Nov. 2006.

[21] Dentino, M., Mccool, J., and B.Widrow, "Adaptive filtering in frequency domain," *Proceedings of IEEE*, vol. 66, pp. 1658–1659, Dec. 1978.

[22] Doherty, J. F. and Porayath, R., "A robust echo canceler for acoustic environments," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, pp. 389–396, May 1997.

[23] Duttweiler, D. L., "A twelve-channel digital echo canceler," *IEEE Trans. on Communications*, vol. COMM-26, pp. 647–653, May 1978.

[24] Ehrmann, F., Le Bouquin-Jeannès, R., and Faucon, G., "Optimization of a two-sensor noise reduction technique," *IEEE Signal Processing Letters*, vol. 2, pp. 108–110, June 1995.

[25] Eskinat, E., Johnson, S., and Luyben, W., "Use of Hammerstein models in identification of nonlinear systems.," *American Institute of Chemical Engineers Journal*, vol. 37, pp. 255–268, 1991.

[26] Fancourt, C. and Parra, L., "The coherence function in blind source separation of convolutivemixtures of non-stationary signals," in *IEEE Workshop on Neural Networks for Signal Processing*, (Falmouth, MA), pp. 303–312, 2001.

[27] Gänsler, T. and Benesty, J., "A frequency-domain double-talk detector based on a normalized cross-correlation vector," *Signal Processing*, vol. 81, pp. 1783–1787, Aug. 2001.

[28] Gänsler, T. and Benesty, J., "The fast normalized cross-correlation double-talk detector," *Signal Processing*, vol. 86, pp. 1124–1139, June 2006.

[29] Gänsler, T., Hansson, M., Ivarson, C.-J., and Salomonsson, G., "A doubletalk detector based on coherence," *IEEE Trans. on Communications*, vol. 44, pp. 1421–1427, Sept. 1996.

[30] Gao, F. X. Y. and Snelgrove, W. M., "Adaptive linearization of a loudspeaker," in *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing*, (Toronto, Canada), pp. 3589–3592, May 1991.

[31] Goethals, I., Pelckmans, K., Hoegaerts, L., Suykens, J., and De Moor, B., "Subspace intersection identification of Hammerstein-Wiener systems," in *Proc. IEEE Conference on Decision and Control*, (Sevilla, Spain), pp. 7108–7113, 2005.

[32] Golub, G. and Van Loan, C., *Matrix Computations*. Johns Hopkins University Press, 1996.

[33] Gómez, J. and Baeyens, E., "Identification of block-oriented nonlinear systems using orthonormal bases," *Journal of Process Control*, vol. 14, pp. 685–697, Sept. 2004.

[34] Gong, Y. and Cowan, C. F. N., "An LMS style variable tap-length algorithm for structure adaptation," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2400–2407, Jul. 2005.

[35] Greblicki, W., "Nonlinearity estimation in Hammerstein systems based on ordered-observations," *IEEE Trans. on Signal Processing*, vol. 44, pp. 1224–1233, May 1996.

[36] Gu, Y., Tang, K., and Cui, H., "LMS algorithm with gradient descent filter length," *IEEE Signal Processing Letters*, vol. 11, pp. 305–307, Mar. 2004.

[37] Gu, Y., Tang, K., Cui, H., and Du, W., "Convergence analysis of a deficient-length LMS filter and optimal-length sequence to model exponential decay impulse response," *IEEE Signal Processing Letters*, vol. 10, pp. 4–7, Jan. 2003.

[38] Guérin, A., Faucon, G., and Bouquin-Jeannes, R. L., "Nonlinear acoustic echo cancellation based on volterra filters," *IEEE Trans. on Speech and Audio Processing*, vol. 11, pp. 672–683, Nov. 2003.

[39] Gustafsson, S., Martin, R., and Vary, P., "Combined acoustic echo control and noise reduction for hands-free telephony," *Signal Processing*, vol. 64, pp. 21–32, Jan. 1998.

[40] Hagenblad, A., "Aspects of the identification of Wiener models," *Licentiate thesis LIU-TEK-LIC-1999*, vol. 51, Nov. 1999.

[41] Hänsler, E. and Schmidt, G., *Acoustic Echo and Noise Constrol: A Practical Approach*. John Wiley & Sons, New Jersey, 2004.

[42] Haykin, S., *Adaptive Filter Theory*. Prentice Hall, 4 ed., 2002.

[43] Heitkamper, P., "An adaptation control for acoustic echo cancellers," *IEEE Signal Processing Letter*, vol. 4, pp. 170–172, Jun. 1999.

[44] Horn, R. and Johnson, C., *Matrix Analysis*. Melbourne, Australia: Cambridge Univ. Press, 1993.

[45] Hoshuyama, O. and Sugiyama, A., "An acoustic echo suppressor based on a frequency-domain model of highly nonlinear residual echo," in *Proc. IEEE ICASSP*, (Toulouse, France), pp. 269–272, May 2006.

[46] Hsia, T., "Convergence analysis of LMS and NLMS adaptive algorithms," in *Proc. IEEE ICASSP*, vol. 8, (Boston, MA), Apr. 1983.

[47] Jenq, J. and Hsieh, S., "Acoustic echo cancellation using iterative-maximal-lengthcorrelation and double-talk detection," *IEEE Trans. on Speech and Audio Processing*, vol. 9, pp. 932–942, Nov. 2001.

[48] Jeraj, J. and Mathews, V. J., "Stochastic mean-square performance analysis of an adaptive Hammerstein filter," in *Proc. IEEE ICASSP*, vol. 54, pp. 725–728, 2004.

[49] Jeraj, J. and Mathews, V. J., "A stable adaptive Hammerstein filter employing partial orthogonalization of the input signals," in *IEEE Trans. on Signal Processing*, vol. 54, pp. 1412–1420, 2006.

[50] Jeraj, J. and Mathews, V. J., "Stochastic mean-square performance analysis of an adaptive Hammerstein filter," in *IEEE Trans. on Signal Processing*, vol. 54, p. 2168C2177, 2006.

[51] Jia, T., Jia, Y., Li, J., and Hu, Y., "Subband doubletalk detector for acoustic echo cancellation systems," in *Proc. IEEE ICASSP*, (Hong Kong, China), pp. 604–607, Apr. 2003.

[52] Jiang, G. Y. and Hsieh, S. F., "Nonlinear acoustic echo cancellation using orthogonal polynomial," in *Proc. IEEE ICASSP*, (Toulouse, France), pp. 273–276, May 2006.

[53] Jones, A., "Transformed coherence functions for multivariate studies," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 317–319, Apr. 1981.

[54] Kallinger, M., Mertins, A., and Kammeyer, K.-D., "Enhanced doubletalk detection based on pseudo-coherence in stereo," in *Proc. International Workshop on Acoustic Echo Noise Control*, (Eindhoven, The Netherlands), pp. 177–180, Sept. 2005.

[55] Kallinger, M., Mertins, A., and Kammeyer, K., "Enhanced doubletalk detection based on pseudo-coherence in stereo," in *Proc. International Workshop on Acoustic Echo Noise Control*, (Eindhoven, The Netherlands), pp. 177–180, Sept. 2005.

[56] Kellermann, W., "Current topics in adaptive filtering for hands-freeacoustic communication and beyond," *Signal Processing*, vol. 80, pp. 1695–1696, Sept. 2000.

[57] Koike, S., "A class of adaptive step-size control algorithms for adaptivefilters," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1315–1326, 2002.

[58] Kraskov, A., Stögbauer, H., and Grassberger, P., "Estimating mutual information," *Physical Review E*, vol. 69, p. 66138, 2004.

[59] Küch, F. and Kellermann, W., "Partitioned block frequency-domain adaptive second-order volterra filter," *IEEE Trans. on Signal Processing*, vol. 53, pp. 564–575, Feb. 2005.

[60] Küch, F., Mitnacht, A., and Kellermann, W., "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proc. IEEE ICASSP*, (Philadelphia, PA), pp. 105–108, Mar. 2005.

[61] Kuech, F. and Kellermann, W., "Nonlinear residual echo suppression using a power filter model of the acoustic echo path," in *Proc. IEEE ICASSP*, (Honolulu, Hawaii), pp. 73–76, May 2007.

[62] Kyriakakis, C., "Fundamental and technological limitations of immersive audio systems," *Proceedings of IEEE*, vol. 86, pp. 941–951, May 1998.

[63] Le Bouquin, R. and Faucon, G., "Using the coherence function for noise reduction," *IEE Proceedings I: Communications, Speech and Vision*, vol. 139, pp. 276–280, 1992.

[64] Lu, X. and Champagne, B., "Acoustic echo cancellation over a non-linear channel," in *Proc. IWAENC*, vol. 1, (Darmstadt, Germany), pp. 139–142, Sept. 2001.

[65] Mader, A., Puder, H., and Schmidt, G., "Step-size control for echo cancellation filtersan overview," *Signal Processing*, vol. 80, pp. 1697–1719, Sept. 2000.

[66] Mandic, D., "A generalized normalized gradient descent algorithm," *IEEE Signal Processing Letters*, vol. 11, pp. 115–118, Feb. 2004.

[67] Mansour, D. and Gray, A. H., "Unconstrained frequency domain adaptive filter," *IEEE Trans. on Acoustic, Speech, Signal Processing*, vol. ASSP-30, pp. 726–734, Oct. 1982.

[68] Mayyas, K., "Performance analysis of the deficient length LMS adaptive algorithm," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2727–2734, Aug. 2005.

[69] Melsa, P., Younce, R., Rohrs, C., Center, T., and Mishawaka, I., "Impulse response shortening for discrete multitone transceivers," *IEEE Trans. on Communications*, vol. 44, pp. 1662–1672, Dec. 1996.

[70] Myllylä, V., "Residual echo filter for enhanced acoustic echo control," *Signal Processing*, vol. 86, pp. 1193–1205, Jan. 2006.

[71] Nollett, B. S. and Jones, D. L., "Nonlinear echo cancellation for hands-free speakerphones," in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, (Mackinac Island, Michigan), Sept. 1997.

[72] Oku, T. and Sano, A., "Nonlinear blind source separation using coherence function," in *SICE Annual Conference*, vol. 3, (Fukui, Japan), pp. 2555–2560, 2003.

[73] OKUNO, T., FUKUSHIMA, M., and TOHYAMA, M., "Adaptive cross-spectral technique for acoustic echo cancellation," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 82, no. 4, pp. 634–639, 1999.

[74] Otnes, R. and Enochson, L., *Applied Time Series Analysis*. John Wiley & Sons Inc, 1978.

[75] Paleologu, C., Ciochină, S., and Benesty, J., "Variable step-size nlms algorithm for under-modeling acoustic echo cancellation," *IEEE Signal Processing Letters*, vol. 15, pp. 5–8, 2008.

[76] Pompe, B., "Ranking and entropy estimation in nonlinear time series analysis," *Nonlinear Analysis of Physiological Data*, pp. 67–90, 1996.

[77] Rao, Y. N. and Principe, J. C., "An RLS type algorithm for generalized eigendecomposition," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, (Falmouth, MA), pp. 263–272, 2001.

[78] Rec, I., "G. 167: acoustic echo controllers," *ITU-T, WTSC, Helsinki, ITU-WWW*, 1993.

[79] Rényi, A., "On measures of entropy and information," in *Proc. Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 547–561, 1961.

[80] Rombouts, G., van Waterschoot, T., Struyve, K., and Moonen, M., "Acoustic feedback suppression for long acoustic paths using a nonstationary source model," *IEEE Trans. on Signal Processing*, vol. 54, pp. 3426–3434, Sept. 2006.

[81] Sayed, A., *Fundamentals of Adaptive Filtering*. Wiley-IEEE Press, 2003.

[82] Scott, D., *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley-Interscience, 1992.

[83] Sentoni, G. and Altenberg, A., "Nonlinear acoustic echo canceller with dabnet + fir structure," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (Net Paltz, NY), pp. 37–40, Oct. 2005.

[84] Shi, K., Ma, X., and Zhou, G. T., "An efficient acoustic echo cancellation Design for systems with long room impulses and nonlinear loudspeakers," *Signal Processing*, to be published.

[85] Shi, K., Ma, X., and Zhou, G. T., "A double-talk detector based on generalized mutual information for stereophonic acoustic echo cancellation in the presence of nonlinearity," in *Proc. IEEE Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), Oct. 2008.

[86] Shi, K., Ma, X., and Zhou, G. T., "A residual echo suppression technique for systems with nonlinear acoustic echo paths," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, (Las Vegas, NE), pp. 257–260, Apr. 2008.

[87] Shi, K., Ma, X., and Zhou, G. T., "A variable step-size and variable tap-length LMS algorithm with exponential decay impulse response," *IEEE Signal Processing Letters*, submitted, 2008.

[88] Shi, K., Ma, X., and Zhou, G. T., "Adaptive acoustic echo cancellation in the presence of multiple nonlinearities," in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, (Las Vegas, NE), pp. 7108–7113, Apr. 2008.

[89] Shi, K., Ma, X., and Zhou, G. T., "A mutual information based double talk detector for nonlinear systems," in *Proc. IEEE Conference on Information Sciences and Systems*, (Princeton, NJ), Mar. 2008.

[90] Shi, K., Ma, X., and Zhou, G. T., "Acoustic echo cancellation using a pseudo-coherence function in the presence of memoryless nonlinearity," *IEEE Trans. on Circuits and Systems I*, to be published.

[91] Shi, K., Ma, X., and Zhou, G., "A channel shortening approach for nonlinear acoustic echo cancellation," in *IEEE Workshop on Statistical Signal Processing*, (Madison, WI), pp. 351–354, 2007.

[92] Shi, K., Ma, X., and Zhou, G., "Adaptive nonlinearity identification in a Hammerstein system using a pseudo coherence function," in *Proc. IEEE Workshop on Statistical Signal Processing*, (Madison, WI), pp. 745–748, Aug. 2007.

[93] Shi, K., Zhou, G., and Viberg, M., "Hammerstein system linearization with application to nonlinear acoustic echo cancellation," in *IEEE Workshop on Digital Signal Processing Signal Processing Education*, (Grand Teton National Park, WY), pp. 183–186, Sept. 2006.

[94] Shi, K., Zhou, G., and Viberg, M., "Compensation for nonlinearity in a Hammerstein system using the coherence function with application to nonlinear acoustic echo cancellation," *IEEE Trans. on Signal Processing*, vol. 55, pp. 5853–5858, Dec. 2007.

[95] Shin, H.-C., Sayed, A. H., and Song, W.-J., "Variable step-size nlms and affine projection algorithms," *IEEE Signal Processing Letters*, vol. 11, pp. 132–135, Feb. 2004.

[96] Sondhi, M. M. and Berkley, D. A., "Silencing echoes on the telephone network," *Proceedings of the IEEE*, vol. 68, pp. 948–963, Aug. 1980.

[97] Sondhi, M. M., Morgan, D. R., and Hall, J. L., "Stereophonic acoustic echo cancellation-an overview of thefundamental problem," *IEEE Signal Processing Letters*, vol. 2, pp. 148–151, Aug. 1995.

[98] Soo, J. S. and Pang, K. K., "Multidelay block frequency domain adaptive filter," *IEEE Trans. on Acoustic, Speech, Signal Processing*, vol. 38, pp. 373–376, Feb. 1990.

[99] Spriet, A., Proudler, I., Moonen, M., and Wouters, J., "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Trans. on Signal Processing*, vol. 53, pp. 3749–3763, Oct. 2005.

[100] Stenger, A. and Kelermann, W., "Nonlinear acoustic echo cancellation with fast converging memoryless preprocessor," in *Proc. IEEE ICASSP*, (Istanbul, Turkey), pp. 805–808, June 2000.

[101] Stenger, A. and Kellermann, W., "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *IEEE Trans. on Acoustic, Speech, Signal Processing*, vol. 80, pp. 1747–1760, Sept. 2000.

[102] Stenger, A., Kellermann, W., and Rabenstein, R., "Adaptation of acoustic echo cancellers incorporating a memoryless nonlinearity," in *Proc. IEEE Workshop on Acoustic Echo and Noise Control*, (Pocono Manor, PA), pp. 168–171, Sept. 1999.

[103] Stoica, P., "On the convergence of an iterative algorithm used for Hammerstein system identification," *IEEE Trans. on Automatic Control*, vol. 26, pp. 967–969, Aug. 1981.

[104] Sugiyama, A., Berclaz, J., and Sato, M., "Noise-robust double-talk detection based on normalized cross correlation and a noise offset," in *Proc. IEEE ICASSP*, vol. 3, (Philadelphia, PA), pp. 153–156, Mar. 2005.

[105] Valin, J. M., "On adjusting the learning rate in frequency domain echo cancellation with double-talk," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, pp. 1030–1034, Mar. 2007.

[106] Van Trees, H., *Detection, Estimation, and Modulation Theory. Part 1, Detection, Estimation, and Linear Modulation Theory*. New York: Wiley, 2001.

[107] van Waterschoot, T. and Moonen, M., "Double-talk robust acoustic echo cancellation with continuous near-end activity," in *Proc. European Signal Processing Conference*, (Antalya, Turkey), Sept. 2005.

[108] van Waterschoot, T., Rombouts, G., Verhoeve, P., and Moonen, M., "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Processing*, vol. 55, pp. 846–858, Mar. 2007.

[109] Woo, T., "Fast hierarchical least mean square algorithm," *IEEE Signal Processing Letters*, vol. 8, pp. 289–291, Nov. 2001.

[110] Ye, H. and Wu, B.-X., "A new double-talk detection algorithm based on the orthogonality theorem," *IEEE Trans. on Communications*, vol. 39, pp. 1542–1545, Nov. 1991.

[111] Zhang, Y., Chambers, J., Sanei, S., Kendrick, P., and Cox, T., "A new variable tap-length LMS algorithm to model an exponential decay impulse response," *IEEE Signal Processing Letters*, vol. 14, pp. 263–266, Apr. 2007.

[112] Zhu, Y., "Estimation of an N-L-N Hammerstein-Wiener model," *Automatica*, vol. 38, pp. 1607–1614, Sept. 2002.

VITA

Kun Shi was born in Qin Huangdao, Hebei, China. He received the B.S. degree in Telecommunication Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2002, and the M.S. degree in Electrical Engineering from Tsinghua University, Beijing, China, in 2005. Since then, he has been working towards his Ph.D. degree in Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, Georgia, USA.

His general research interests are in the areas of signal processing and communications. Specific current interests include acoustical signal processing, nonlinear signal processing, and adaptive algorithm design and development on digital signal processors.