Optimization of Acoustic Echo and Noise Reduction in Non Stationary Environment

Vincy, R. Dr. J. William
Department of Electronics and Communication Engineering, M.A.M College of Engineering, Trichy, India

Article Info
Article history:
Received 19 January 2015
Received in revised form
15 February 2015
Accepted 28 February 2015
Available online 15 March 2015

Keywords
Non Stationary,
Echo Reduction,
RLS,
Time Varying Filters

Abstract

Optimized speech enhancement method combines acoustic echo reduction and noise reduction in a unified framework for non stationary environment. Simultaneous optimization of noise and echo reduction is already done in stationary environment. In most of the times in transmission, signal properties change over time. We need to remove the artifacts of sound in those conditions. Recursive least square method proposed for noise and echo reduction. It gives little amount of mean square error and better results. Normally, partial optimization of acoustic echo reduction and noise reduction does not lead to total optimization. A cascade method of multiple functions causes mutual interference between these functions and degrades eventual speech enhancement performance. Unlike cascade methods, the proposed method combines all functions to optimize eventual speech enhancement performance based on a unified framework, which is also robust against the mutual interference problem. With the proposed method, in addition to time-invariant linear filters, time-varying filters are used to reduce residual acoustic echo signal, and background noise signal which cannot be reduced using time-invariant filters. These time-invariant filters and time-varying filters are also optimized based on a unified likelihood function to avoid the mutual interference problem. Under this, all the parameters are optimized simultaneously based on the expectation-maximization algorithm and calculates a minimum mean squared error estimate of a desired signal. The experimental results show that the proposed method is superior to the cascade methods.

1. Introduction

Speech enhancement is essential in teleconferencing systems. There are many unwanted signals in teleconferencing scene, e.g., acoustic echo signal from a far-end loudspeaker, reverberation of a near end talker and background noise signal. The extraction of desired speech signal from a mixture of signals from multiple simultaneously active talkers and background noise is of interest in many hands-free communication systems, including modern mobile devices, smart homes, and teleconferencing systems.

In some applications, e.g., where automatic speech recognition is required, the goal is to obtain an estimate of the signal from a desired talker while reducing noise and signals from interfering talkers. In other applications, an estimate of each talker's signal is required. In practice, information about the number and location of the different talkers, or the presence and type of background noise is unavailable and the estimation is based solely on the microphone signals.

There are many speech enhancement methods that reduce these unwanted signals independently. A cascade connection of acoustic echo and noise reduction is proposed. However, when the cascade connection is used, a mutual interference problem occurs. When acoustic echo reduction is done prior to beam forming, echo reduction performance degrades because the background noise signal and near-end speech signals interrupt optimization of the echo reduction filter.

When beam forming is done prior to acoustic echo reduction, acoustic echo signal is partially reduced by beamforming and background noise reduction performance of beamforming degrades. When a cascade connection of dereverberation and acoustic echo reduction is used, a similar mutual interfering problem also occurs. When dereverberation is done prior to acoustic echo reduction, dereverberation partially reduces the acoustic echo signal, and it leads to degradation of dereverberation performance of near-end speech sources.

Mutual interference problems also occur when linear filters and non-linear filters are connected in a cascade manner .When there are residual reverberation and residual acoustic echo signal after linear filtering, non-linear filters are required for reducing unwanted signals to a sufficient level. However, non-linear filters distort the output signal. To prevent all filters from interfering with each other, optimization of two arbitrary filters has been investigated, e.g., simultaneous optimization of noise and echo reduction simultaneous optimization of dereverberation and echo reduction, simultaneous optimization of dereverberation and noise reduction .

The simplest method for reducing multiple undesired signals simultaneously by using a linear filter is a least squares method, which minimizes the ensemble average of residual signal power. With the least squares method the residual signal is assumed to be an independent and identically distributed signal, but the residual signal is actually a time-varying speech signal, and the output signal is over-whitened.

Corresponding Author,
E-mail address: vincysweety.11@gmail.com

Volume 3, Issue 1 (2015) 221-224 ISSN 2347 - 3258
International Journal of Advance Research and Innovation

Single channel noise reduction methods, e.g., spectral subtraction and MMSESTSA- based noise cancellers reduce stationary background noise signal under the assumption that the speech signals are non-stationary sources. Multichannel beamforming methods are often utilized for background noise reduction and separation of multiple speech signals. Least square algorithms based on adaptive filters are commonly used for acoustic echo reduction methods. For speech dereverberation, MINT based dereverberation methods are commonly used. The multi-channel input signals are processed using MVDR beamformer, which aims to suppress sound sources not arriving from the DOA of the target speaker. The noise coherence matrix in the MVDR beam former is estimated from noise-only periods, determined using VAD and the DOA of the target speaker is estimated using the MUSIC algorithm. The beam former output is then processed using a single channel speech enhancement scheme, which aims at jointly suppressing the remaining noise and reverberation and relies on estimates of the power spectral densities (PSDs) of the noise and of the reverberation.

We propose a speech enhancement method that combines speech dereverberation, noise reduction, and acoustic echo reduction by using a time-varying probabilistic model of a speech source in a unified framework. Contrary to the cascade methods, which use multiple cost functions with indirect reflection of the eventual speech enhancement performance, the proposed method uses a unified likelihood function, which directly reflects the eventual channel enhancement performance in non stationary environment

2. Problem Statement

2.1 Input Signal Model

We address a speech enhancement problem with multichannel microphones when there is a direct path of speech sources, reverberation, background noise signal, and acoustic echo signal. In this study, dereverberation and acoustic echo reduction and background noise reduction were done at the time-frequency domain. The microphone input signal in the time domain is converted into the microphone input signal in the time-frequency domain by using short-term Fourier transform. The microphone input signal of each channel,

$$x(l,k)= \sum^{D-1} H_l(l',k)s(l-l',k)+\sum^{L_{h-1}} g_l(l',k)d(l-l',k)+ f(x)\frac{f(x)}{\gamma(x,\tau)} + \frac{D^{1/2}}{\tau^{1/4}\gamma(x,\tau)} w(s) \quad \text{.... (1)}$$

The first term is the summation of the direct path and early reflections of the speech sources, the second term is acoustic echo signal, and the third term is color noise signal. In this paper, extraction of the first term from the microphone input signal is set as the goal of speech enhancement.

2.2 Autoregressive Model

To extract the first term in (1), the original source signal s(l,k) should be estimated. Time varying autoregressive model of y(l,k),

$$y(l,k) \approx \sum^{L_w-1} H_l(l',k)s(l-l',k) + \sum_{l'=D} W_l(l',k)\, y(l-l',k) \quad (2)$$

Where $W_l(l',k)$ is a time varying autoregressive coefficient

2.3 Probabilistic Models

The proposed method enables speech enhancement by using a probabilistic approach. The log-likelihood function of the microphone input signal is defined at each frame-frequency point under the assumption that the far-end speech signal is given in advance as,

$$L(X_{L,k}|D_{L,k};\theta_k) = \sum_{l=1}^{L_T} \log p(x(l,k)|D_{L,k},X_{l-1,k};\theta_k) \quad (3)$$

3. Color noise and echo

3.1 Colored Noise

Color denotes power spectrum characteristics. It don't have constant power spectrum.

Echo: Acoustic echo results from a feedback path set up between the speaker and the microphone in a mobile phone, hands-free phone, teleconference or hearing aid system. Acoustic echo is usually reflected from a multitude of different surfaces, such as walls, ceilings and floors, and travels through different paths.

Gain of an acoustic feedback loop depends on the frequency responses of the electrical and the acoustic signal paths. The undesirable effects of the electrical sections on the acoustic feedback can be reduced by designing systems that have a flat frequency response.

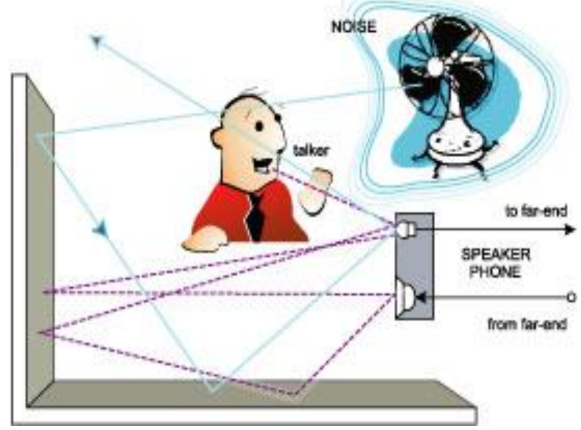

Fig: 1. Acoustic echo

4. Noise Removal

4.1 Color Noise Removal

In this paper, color noise is removed by RLS algorithm. RLS algorithm considers all the available parameters. It should be optimum with respect to all the available data in certain sense. It minimizes the cost function,

$$\in[n] = \sum_{k=0}^{n} \lambda^{n-k} e^2(k) \quad (4)$$

Recent data is given by more weightage. For stationary, $\lambda = 1$. For non stationary, $\lambda = 0.99$ minimize

$$\in [n] = \sum_{k=0}^{n} \lambda^{n-k} \left(x[k]-y'[k]h[n]\right)^2 \text{with respect to } h[n] \quad (5)$$

4.2 Acoustic Echo Removal

Echo is a phenomenon where a delayed and distorted version of an original sound or electrical signal is reflected back to the source. If a reflected wave arrives after a very short time of direct sound, it is considered as a spectral distortion or reverberation. However, when the leading edge of the reflected wave arrives a few tens of milliseconds after the direct sound, it is heard as a distinct echo.

A significant problem in communications is the generation of echoes. The echoes arise for a number of reasons, with the primary reason being an impedance mismatch. The impedance mismatch occurs when the two-wire network meets the four-wire network, this interface is known as the hybrid. This impedance mismatch causes some of the signal energy to be returned to the source as an echo.

4.3 Matlab Implementation of Acoustic Echo Cancellation (AEC)

In acoustic echo cancellation, a measured microphone signal d(n) contains two signals: - the near-end speech signal v(n) - the far-end echoed speech signal dhat(n) The goal is to remove the far-end echoed speech signal from the microphone signal so that only the near-end speech signal is transmitted The Room Impulse Response we describe the acoustics of the loudspeaker-to-microphone signal path where the speakerphone is located. We can use a long finite impulse response filter to describe these characteristics.Far end and near end signals are uploaded from matlab

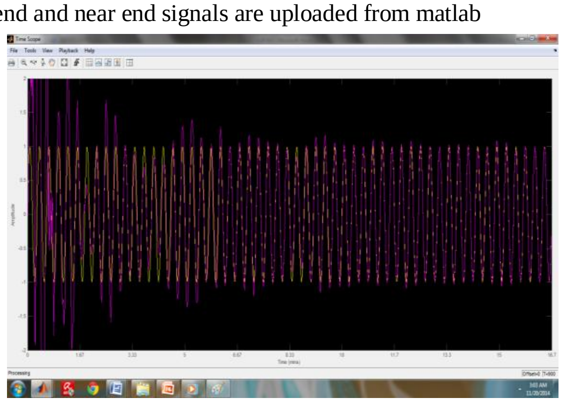

Fig: 2. Time Scope of Composite Signal

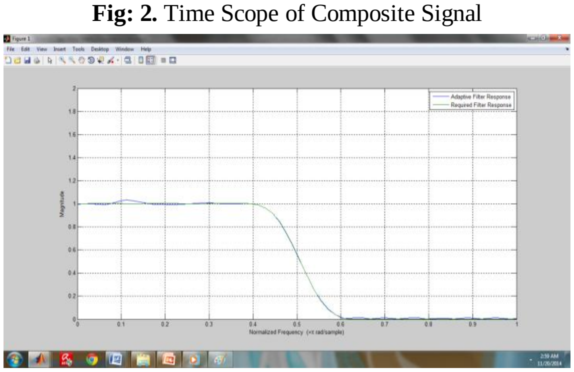

Fig: 3. Filter Response of RLS Adaptive Noise Canceller

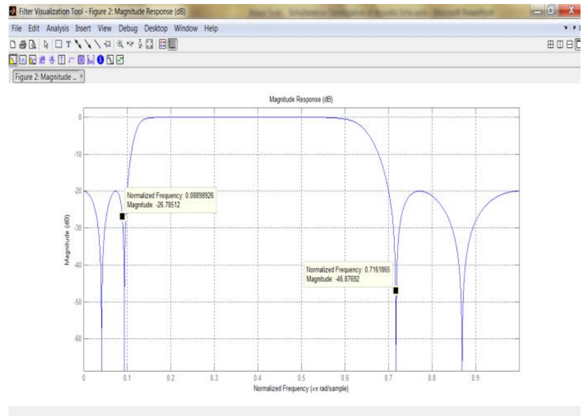

Fig: 4. Filter Response of Adaptive Frequency Domain Filter

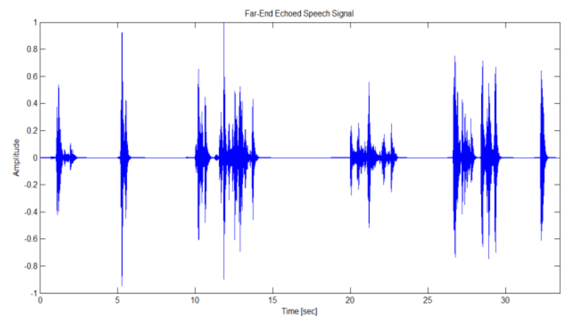

Fig: 5. Far end echoed speech signal

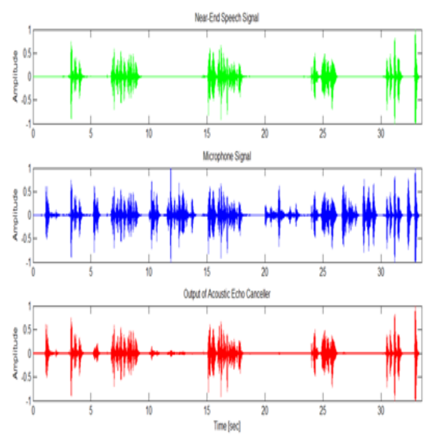

Fig: 6. Acoustic Echo canceller Output

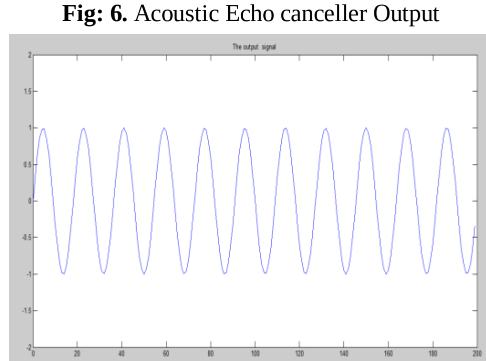

Fig: 7. Output Signal

References

[1] M Yukawa, R C de Lamare, R Sampaio-Neto, Efficient acoustic echo cancellation with reduced-rank adaptive filtering based on selective decimation and adaptive interpolation. IEEE Trans. Audio, Speech, Lang. Process 16(4), 2008, 696–710

[2] B. Milner, Applying noise compensation methods to robustly predict acoustic speech features from MFCC vectors in noise, Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference

[3] M. Popescu, A. Mahnot, Acoustic fall detection using one-class classifiers, in: 31st Annual International Conference of the IEEE EMBS, Minneapolis, MN, USA, 2009

[4] S Wu, X Qiu, M Wu, Stereo acoustic echo cancellation employing frequency-domain preprocessing and adaptive filter. IEEE Trans. Audio, Speech, Lang. Process, 19(3), 2011, 614–623

[5] M. Togami, Y. Kawaguchi, Speech enhancement combined with dereverberation and acoustic echo reduction for time varying systems, in Proc. Statist. Signal Process. Workshop, 2012, 357–360.

[6] R Nath, Adaptive echo cancellation based on a multipath model of acoustic channel. Circuits, Syst. Signal Process., Springer US, 32(4), 2013, 1673–1698