Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 70 (2015) 65–75

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Robust acoustic echo cancellation using Kalman filter in double talk scenario

Ayush Jain, Sanchit Goel, Karan Nathwani, Rajesh M. Hegde *

Department of Electrical Engineering, Indian Institute of Technology, Kanpur, India

Received 10 October 2014; received in revised form 19 February 2015; accepted 20 March 2015
Available online 31 March 2015

Abstract

In this work, a novel Kalman filtering framework is developed for joint acoustic echo and noise cancellation in a double talk scenario. The efficiency of echo cancellation algorithms is reduced when signals other than the echoed far end signal are present, since the echo path cannot be modelled accurately in such cases. A double talk detector is also used in conjunction with an acoustic echo canceller to handle such a double talk scenario. The method presented in this work is able to model both the near-end speech signal and background noise, which makes lt robust in double talk scenarios. Apart from jointly cancelling echo and noise, another advantage of this framework is that it does not require a double talk detector. Additionally an expectation maximisation based algorithm is also proposed in this work to estimate linear prediction coefficients of the near end signal. Extensive performance evaluation over the NOIZEUS corpus demonstrates that the proposed framework performs reasonably better than other speech enhancement methods in terms of misalignment of the estimated echo path and perceptual quality of the reconstructed near-end speech signal.

Keywords: Acoustic echo cancellation; Expectation maximisation; Double talk; Kalman filter

1. Introduction

The problem of speech enhancement in a traditional microphone-loudspeaker system is due to the output of the loudspeaker being fed back into the microphone, along with the background noise. The primary objective of this work is to suppress the echoed loudspeaker output and noise from the microphone input. The joint echo and noise cancellation problem have been studied independently while using Kalman filter in literature, and solutions have been proposed for each of them separately. However, this problem of joint echo and noise cancellation has hitherto received little attention and is the primary objective of this paper.

Modelling the speech signal as a stochastic autoregressive (AR) model embedded in additive white Gaussian noise has been suggested by Lim and Oppenheim (1978). In this work, this has been used as the basic model for speech enhancement.

The Kalman filter for noise reduction was first proposed in Paliwal and Basu (1987). It provides an unbiased and linear minimum mean squared error (MMSE) estimate of the clean speech signal. The enhanced speech is estimated on a sample-by-sample basis in a recursive manner. Significant noise cancellation was reported in the case when Linear Prediction Coefficients (LPCs) of the clean speech signal were provided. Poor parameter estimates from noisy speech in practical scenarios significantly degrade the performance. Iterative and sequential Kalman filters in Gannot et al. (1998) offer a solution to this parameter estimation problem. The convergence is not, however, guaranteed in this solution. The enhanced signal may additionally suffer from some distortion. Recently, in So et al. (2010) and So and Paliwal (2011), Kalman filters have

* Corresponding author. Tel.: +91 512 2596248.
E-mail address: nathwani@iitk.ac.in (K. Nathwani).

http://dx.doi.org/10.1016/j.specom.2015.03.002
0167-6393/

been used in the modulation domain for noise cancellation tasks. Upper bound performance of this filter was empirically shown to exceed its time domain variant. This has been attributed to its ability to predict long-term correlation information (like pitch harmonics). However, applying a Kalman filter in each frequency bin in the STFT domain increases the computational complexity.

The unknown LPCs can be estimated by an Expectation Maximisation (EM) algorithm (Dempster et al., 1977). It is generally applied to the parameter estimation problems in which the data for evaluating the parameters are missing or incomplete. When there is a many-to-one mapping from an underlying distribution to the distribution governing the observation, it is known to produce the maximum-likelihood (ML) parameter estimates. The EM algorithm has widespread applications in digital image and speech processing (Zhao and Kleijn, 2007; Hao et al., 2009; Gannot et al., 1998).

A detailed study on the use of Kalman filter for echo cancellation has recently been carried out in Paleologu et al. (2013) and Benesty et al. (2005), where an adaptive Kalman filter has been used to estimate the echo path. However, the assumption of Gaussian nature of near-end signals restricts its application to single-talk scenarios. We use a more general model for the near-end speech and additionally consider additive noise. Some previous work on the problem of single-talk echo and noise cancellation assume multi-channel input (Reuven et al., 2007; Buchner et al., 2003). However in this work, we have considered the single-channel input. This is because, we have considered single channel echo and noise cancellation scenario where single input is present. In literature, many authors had utilised multi microphone information for joint echo and noise cancellation which is easily achievable when compared to information present in single microphone case.

The problem of echo cancellation has been extensively studied by the research community. Due to their performance, various adaptive echo cancellers (AEC) with finite impulse responses (FIR) have been investigated. The least mean squares (LMS), normalised least mean squares (NLMS) and recursive least squares (RLS) have been popular owing to their simplicity and predictable behaviour (Haykin, 2000). Related work in this context has been done in the context of double talk detection (Jung et al., 2005). When signals from both near-end and far-end speakers co-exist, the AEC give erroneous results. A Double Talk Detector (DTD) allows the adaptive filter to update its coefficients only during single-talk periods and freezes adaptation during double-talk periods to avoid unwanted divergence. This creates a problem in scenarios where there is continuous double talk. The adaptive filter does not get sufficient samples to adapt its estimate of the echo path. During the time required by the DTD to detect double talk, the estimate of the RIR often diverges. This is because a few undetected large amplitude samples perturb the echo path estimate considerably (Gansler et al., 2000). Additionally, the outputs of Double Talk Detectors can be erroneous, specially in the presence of background noise. This error will then propagate to the succeeding AEC technique thereby affecting performance. In another work (Enzner and Vary, 2006), a frequency domain Kalman filter has been proposed for echo cancellation.

In our previous work (Tanan et al., 2014), we have presented acoustic echo and noise cancellation in a generalised sidelobe canceller framework. The primary contribution of Tanan et al. (2014) is the development of multichannel adaptive Kalman filter (MCAKF) in a modified generalised sidelobe canceller (MGSC) framework.

In this work, a novel Kalman filtering framework has been developed for cancelling acoustic echo and noise in a double talk scenario. Apart from handling noise in addition to echo, another advantage of this framework is that it does not rely on a DTD. This can be desirable in many cases, as noted above. The general Kalman filtering framework for echo cancellation requires the knowledge of linear prediction coefficients of the near-end speech signal. An Expectation Maximisation based recursive algorithm is also proposed in this work to estimate these parameters.The difference between the proposed and Enzner and Vary (2006) that we develop the Kalman filter in time-domain and additionally consider robustness to noise. As we show in our experiments Section, typical echo cancellation techniques do not perform well in the presence of noise, even after pre-processing to remove noise.

Practical applications of this work can be in devices where deterministic loudspeaker signals are being fed into the microphone along with near-end speech. This may be in the form of a user giving commands to a television (near-end speech), with the output of the television also being recorded into the microphone.

The rest of the paper is organised as follows. In Section 2, the echo and noise cancellation model is described. In Section 3, a Kalman filtering framework is developed for solving the problem of joint acoustic echo and noise cancellation. Following this, an Expectation Maximisation based algorithm to simultaneously estimate the unknown LP parameters is detailed in Section 4. Finally, in Section 5, we present a detailed performance evaluation of the proposed algorithms.

2. The model for joint acoustic echo and noise cancellation in double talk scenario

The model for joint acoustic echo and noise cancellation in double-talk scenario is shown in Fig. 1. Here, a speech signal from the far-end $f(n)$ is being broadcast by a loudspeaker. A microphone is present to record a local near-end speech signal $x(n)$ and its output is being transmitted to the loudspeaker. The presence of an echo path between the loudspeaker and the microphone corrupts the microphone signal with an echoed signal $e(n)$, in addition to background noise $v(n)$. The variance of this additive noise is $\sigma_v^2 = E[v^2(n)]$, where $E[.]$ is the mathematical expectation.