

Echo Cancellation: A Novel Adaptive Kalman Filter-Based Scheme

Rajamani Doraiswami¹, Lahouari Cheded²

¹Department of Electrical and Computer Engineering, University of New Brunswick,
Fredericton, New Brunswick, Canada

²Systems Eng. Department, King Fahd University of Petroleum & Minerals,
Dhahran 31261, KSA

Abstract - A novel adaptive echo cancellation scheme, using an accurate and reliable two-stage identification scheme and an adaptive Kalman filter (KF), is proposed. The novel scheme estimates a desired waveform from the received signal which is corrupted by an undesired echo and noise. It is assumed that the desired waveform and the echo are uncorrelated with each other, and that the reference waveform is highly correlated with the echo. Further, the reference waveform and echo are related by a rational transfer function and are assumed to be measurable. An accurate and reliable model of the echo is identified from the received signal using a proposed novel two-stage identification scheme. This two-stage scheme is used to identify accurately and reliably the model relating the reference and actual echo waveforms using least squares and model order reduction. It is implemented in the frequency domain using waveform segmentation to ensure efficient computation and model reduction, signal stationarity and real-time system implementation. The identified model of the echo is then embodied into the KF which is a minimum-variance estimator that is robust to noise and disturbances and has a zero-mean, white noise residual. The performance of the KF is monitored continuously and its gain updated adaptively. If the filter's residual fails the whiteness test, the model of the echo is then re-identified and the KF adapted accordingly. The proposed scheme was successfully evaluated on simulated and real recorded speech corrupted by noise and echo. This novel scheme can be extended to areas such as beamforming

Keywords: Echo, Echo cancellation, Speech, Adaptive Kalman Filtering, Two-stage Identification

Nomenclature

y_d	Desired waveform
y_e	Echo waveform
y_r	Reference waveform
y_d	Received waveform
$G(z)$	Transfer function relating y_e to y_r
\hat{x}	Estimate of x
$F(z)G(z) = 1$	

1. Introduction

Echo and noise cancellation finds applications in telecommunication, power line communication for improving the quality of speech. An echo may be either acoustic or telephone line hybrid. The acoustic echo is the result of feedback between the microphone and the speaker affecting services such as the mobile phone, hearing aid, and teleconferencing. The echo may also be the waveform reflected from walls, ceilings etc. Telephone line echoes result from an impedance mismatch at telephone exchange hybrids [1]. A similar scheme may be applied to active noise control [2].

The problem of echo cancellation is still a subject of intense research and development by both researchers and telecommunication industries. The main problem in echo cancellation is to obtain an accurate and reliable estimate of the echo waveform, which is buried in the received waveform, so that the desired waveform may be obtained by subtracting the estimated echo from the received waveform. Generally, a Finite Impulse Response (FIR) model is identified by treating

the reference waveform as the input and the received waveform as the output. A Least Mean Square (LMS)-type recursive scheme is used. Using this model the echo waveform is then estimated.

In this work, echo cancellation is based on segmentation of the waveforms, two stage identification and an adaptive Kalman filter, which embodies the identified model and estimates the echo waveform. The various stage of the proposed scheme are described next.

Segmentation: The received and the reference waveforms are segmented. The segment length is chosen to be sufficiently small so that the segmented waveforms are stationary. For example a speech waveform may be assumed to be stationary over a 20 ms duration, which corresponds to segment length of 320 samples assuming the sampling rate of 16 kHz. The segmentation reduces the model order, alleviates computational complexity.

Two-Stage Identification: A two-stage identification scheme ([3], [4]) is employed to identify the system model when the input output data is corrupted by both the noise and the echo, especially when the noise is coloured with unknown statistics. First a very high order model is identified so that the order is high enough to capture complete dynamics of the signal and noise. Then in the second stage, a reduced order model is derived from the identified high order model from the first stage, using a frequency-weighted least-squares estimator. The two-stage identification is an intelligent filter as it were, which *automatically* extracts the true model the noisy input-output data thereby avoiding a human in the loop to provide design information and to locate every noise artefacts in the data. It is as it were a generalized phase lock loop, which locks on to the dominant spectral components.

Adaptive Kalman Filter: The Kalman filter embodies the identified model of relating the reference and received waveforms. It provides an accurate, reliable and robust estimate of the echo waveform, with its residual being zero-mean and white *if and only if* there is no model mismatch-i.e. the actual system model is not different from the identified one. The performance of the Kalman filter is continuously monitored in that if its residual fails the the whiteness test, the echo model is then re-identified and the Kalman gain accordingly updated.

2. Problem Formulation

Echo cancellation concept: A widely popular approach to echo cancellation [5] is represented by a block diagram shown in Fig. 1, where y_d , y_e , y_r and y are the desired, echo, reference and received waveforms. The received waveform (y) is the sum of the desired (y_d) and echo waveforms (y_e). The reference (y_r) and the echo waveforms (y_e) are related by $y_r(z) = F(z)y_e(z)$ where $F(z)$ is an unknown transfer function. Let $G(z)$ be the rational transfer function that relates y_e and y_r , that is $G(z) = y_e(z)/y_r(z)$, and $\hat{G}(z)$ be the echo model identified by treating y_r and y are the input and output, respectively, that is $\hat{G}(z) = y(z)/y_r(z)$. Assuming the ideal case where the identified model is accurate, and y_r and y_d are uncorrelated, i.e. $\hat{G}(z) = G(z)$, implying that the estimated echo waveform $\hat{y}_e(z) = y_e(z)$ and $\hat{G}(z)F(z) = 1$,

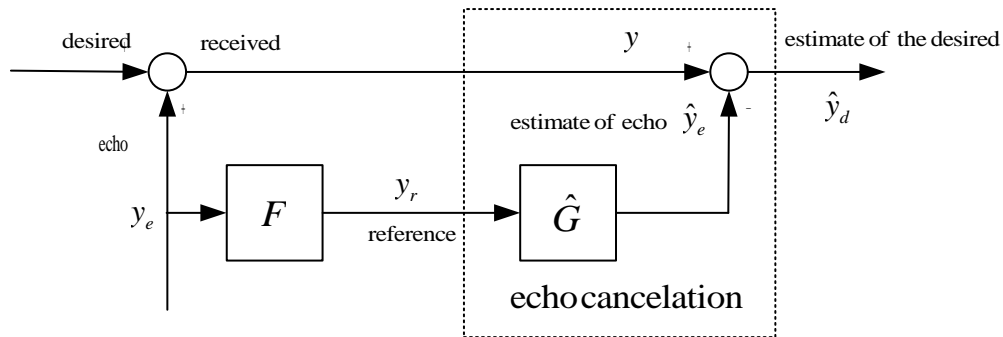


Fig. 1: Block diagram representation of echo cancellation scheme.

3. Echo Cancellation: The Conventional and the Proposed Approaches

There is a lack of an overarching theory that will pull all of the seemingly different echo cancellation techniques into a unified theory that will make all such available theories mere special cases. Echo cancellation is a system identification problem as can be deduced from the block diagram shown in Fig. 2 relating y , y_e , y_d and y_r where only y and y_r are measured, and a map relating y_r and y_e is assumed to be some unknown dynamic model.

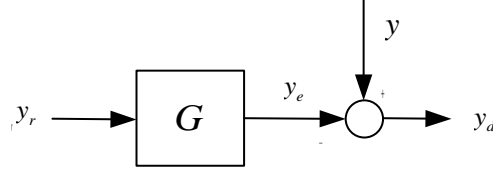


Fig. 2: Block diagram representation relating waveforms and the estimates.

Various echo cancellation schemes have been proposed depending upon the choice of the structure of the dynamic model G .

Conventional approach: The widely-used scheme is based on assuming that G is a moving average (MA) model:

$$y_e(k) = \sum_{i=0}^{n_b} b_i y_r(k-i) + v(k) \quad (1)$$

where $v(k)$ is the equation error.

Proposed approach: The proposed scheme assumes a more general auto-regressive and moving average (ARMA) model given by:

$$y_e(k) = -\sum_{i=0}^{n_a} a_i y_e(k-i) + \sum_{i=1}^{n_b} b_i y_r(k-i) + v(k) \quad (2)$$

Substituting for $y(k) = y_e(k) + y_d(k)$ yields:

$$y(k) = -\sum_{i=1}^{n_a} a_i y_e(k-i) + \sum_{i=1}^{n_b} b_i y_r(k-i) + y_d(k) + v(k) \quad (3)$$

where $\{a_i\}$ with the leading coefficient $a_0 = 1$ are the coefficients of the auto-regressive (AR) part of the ARMA model; $\{b_i\}$ are the MA coefficients of the moving average (MA) part of the ARMA model; and n_b and n_a are respectively the numerator and the denominator orders, and $v(k)$ is the equation error resulting from the combined effects of measurement noise, disturbances and modelling errors.

Assuming the desired speech waveform y_d to be stationary over a frame (of approximately 20 ms.), a linear prediction coding (LPC) all-pole model becomes:

$$y_d(k) = -\sum_{i=0}^{n_a} a_{di} y_d(k-i) + u_d(k) \quad (4)$$

where $\{a_{di}\}$ are LPC coefficients, $u_d(k)$ is the excitation, which is assumed to be a zero-mean white noise process. The choice of the white noise excitation to cover both the voiced and the unvoiced speech gives an acceptable accuracy [6].

Kalman Filter-based identification

Let (A_e, B_e, C_e) be the state-space model relating the input $y_r(k)$ and the output $y(k)$ given by **Error! Reference source not found.**; (A_d, C_d) be the signal model of $y_d(k)$ given by **Error! Reference source not found.** An augmented state space-model is derived, as follows, by appending the 2 state-space models (A_e, B_e, C_e) and (A_d, C_d) to each other:

$$\begin{aligned} x(k+1) &= Ax(k) + Br(k) + Ew(k) \\ y(k) &= Cx(k) + v(k) \end{aligned} \quad (5)$$

where $A = \begin{bmatrix} A_e & 0 \\ 0 & A_d \end{bmatrix}$; $B = \begin{bmatrix} B_e & 0 \\ 0 & B_d \end{bmatrix}$; $C = [C_e \quad C_d]$; $x(k)$ is the state, $r(k) = [y_r(k) \quad u_d(k)]^T$ is the input, $w(k)$ and $v(k)$ are the disturbances and the measurement noise. The echo and the desired waveform are:

$$\begin{aligned} y_e(k) &= C_e x(k) + v_e(k) \\ y_d(k) &= C_d x(k) + v_d(k) \end{aligned} \quad (6)$$

where $v_e(k)$ and $v_d(k)$ satisfy $v(k) = v_e(k) + v_d(k)$

Assumptions:

- It is assumed that $y_d(k)$ is uncorrelated with $y_e(k)$ - that is there is no spectral overlap between these 2 waveforms.
- The reference waveform $y_r(k)$ is highly correlated with $y_e(k)$ and least correlated with $y_d(k)$, i.e. the spectra of $y_r(k)$ and $y_e(k)$ are coincident while there is no (or as small as possible) spectral overlap between $y_r(k)$ and $y_d(k)$.
- The disturbance $w(k)$ and $v(k)$ are both assumed to be zero-mean white-noise processes

Derivation of the residual of the Kalman filter of the system

The Kalman filter model of the system **Error! Reference source not found.** becomes ([7], [8] and [9]):

$$\begin{aligned} \hat{x}(k+1) &= A\hat{x}(k) + Br(k) + K(y(k) - C\hat{x}(k)) \\ \hat{y}(k) &= C\hat{x}(k); \quad \hat{y}_e(k) = C_e\hat{x}(k); \quad \hat{y}_d(k) = C_d\hat{x}(k) \end{aligned} \quad (7)$$

where $\hat{x}(k)$, $\hat{y}(k)$, $\hat{y}_e(k)$ and $\hat{y}_d(k)$ are respectively the estimate of $x(k)$, $y(k)$, $y_e(k)$ and $y_d(k)$; K is the Kalman gain.

An expression for the residual of the Kalman filter (Doraiswami and Cheded, 2012) is given by:

$$e(z) = \frac{D(z)}{F_0(z)} y(z) - \frac{N(z)}{F_0(z)} r(z) \quad (8)$$

where $N(z)$ and $D(z)$ are the numerator and denominator of the system transfer function relating $y(z)$, and $y_r(z)$, and $F_0(z) = |zI - A + KC|$ is the Kalman polynomial.

Proposed Reliable and Accurate System Identification:

The key 2 properties the Kalman filter, namely, that its residual $e(k)$ is a) zero-mean, white-noise process and b) has minimum variance compared to that of any other estimator, are exploited in developing the proposed identification. The coefficients of $N(z)$ and $D(z)$ of the system model are estimated from minimizing the sum of the squares of the residual, i.e.:

$$\min_{N,D} \left(\sum_{k=1}^N e^2(k) \right) \tag{9}$$

In order to avoid dealing with a nonlinear relation governing residual and the coefficients of the Kalman polynomial $F_0(z)$ the rational polynomials $\frac{D(z)}{F_0(z)}$ and $\frac{N(z)}{F_0(z)}$ are approximated by high-order polynomials [4] and the coefficients of $N(z)$ and $D(z)$ are estimated using the proposed two-stage identification scheme. Using these estimated coefficients, the state-space model Kalman filter (A, B, C) is then obtained.

Bayesian hypotheses testing: A Bayesian hypothesis testing scheme is employed to verify whether the performance of the Kalman filter is normal. If it is not normal, the echo model is re-identified and the Kalman gain updated [10]

Salient Features of the Proposed Scheme

- The estimation of the echo waveform is reliable and highly accurate in the face of the noise and the disturbances corrupting received and the reference waveforms.
- The quality of the estimated speech is better than that obtained with conventional techniques
- The proposed scheme is successfully evaluated on a number of simulated and actual acoustic waveforms
- In order to reduce the computational burden without sacrificing the accuracy, the following scheme is employed:
 - A complex task involving large computational burden and high memory requirement is divided into a number of smaller tasks requiring lesser computations and memory. The waveforms, which are typically large, are segmented into a number of smaller data samples.
 - identification. The segment length is chosen to be a power of 2 so as to use the efficient fast fourier transform of waveforms and subsequently the auto and cross correlations
 - The Bayesian testing scheme is used to monitor the performance of the proposed scheme. Only if it is found unacceptable is the model is re-identified. Otherwise, the burdensome task of identification is avoided.

4. Evaluation On Audio Data

The proposed scheme is evaluated on audio recordings of a speech of an individual that is corrupted by some background speech. Recordings were provided by [11].

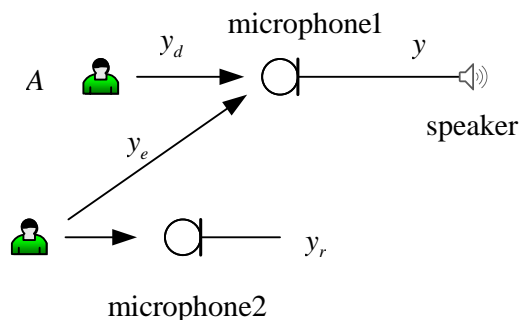


Fig. 3: Audio waveforms from persons A and B.

Figure 3 shows audio signals, desired waveform y_d from person A, echo waveform y_e from person B, received waveform $y = y_e + y_d$ captured by the microphone 1 and played on the speaker, and reference waveform y_r from microphone 2.

The audio signals were a) of duration 4.8 sec., b) digitized at a sampling rate $f_s = 16000$, the total number of samples $N = 76800$. To reduce the computational burden, and to ensure stationary, the waveforms were segmented. The number of segments was 300, and the length of each segment 256 (a power of 2 so that fast Fourier transform could efficiently be computed), the duration of the segment is 16ms.

Figure 4 shows the segmented waveforms of the echo and the desired. The subfigures A and B show respectively the echo x_e and its estimate \hat{x}_e ; the desired waveform \hat{x}_d and its estimate \hat{x}_d . Figure 5 shows the waveforms of the echo and the desired. The subfigures A and B show respectively the echo y_e and the desired waveforms \hat{y}_d .

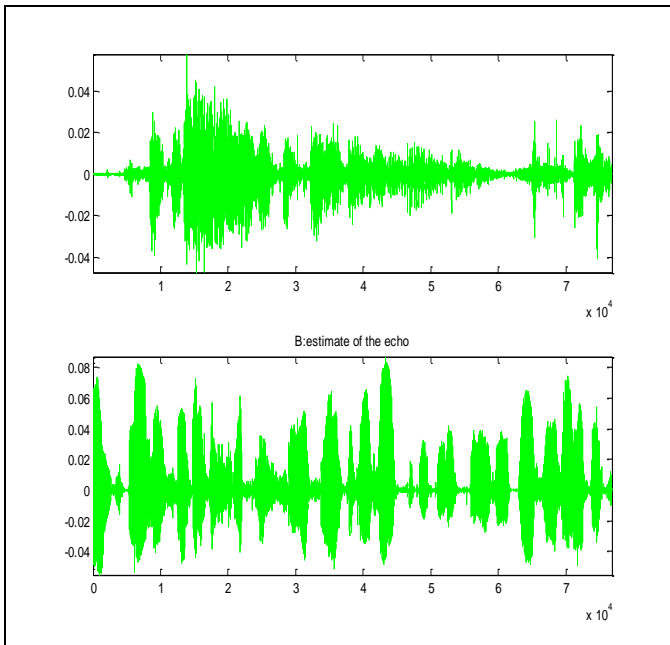


Fig. 4: Segmented waveforms:the echo and the desired and their estimates.

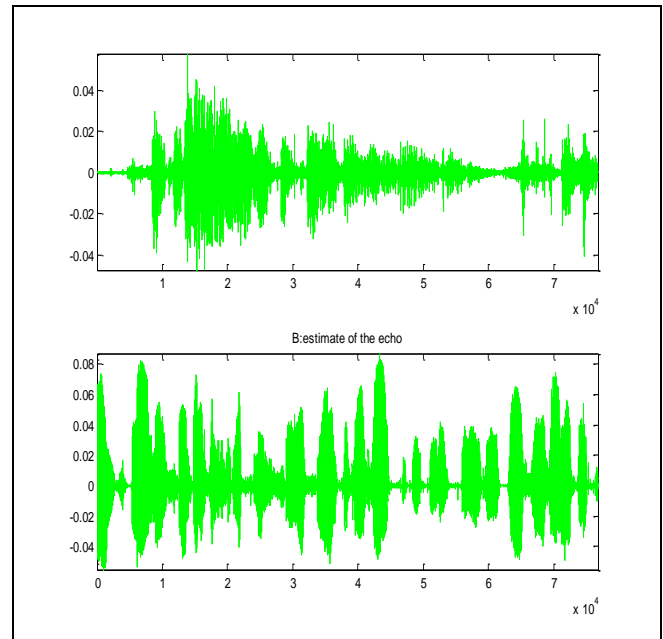


Fig. 5: The echo and the desired waveforms.

Comment: It can be deduced from Fig. 4 that the estimates of the echo and the desired waveforms track

5. Conclusions

Direct identification of the Kalman filter model from the received noisy waveform and the reference echo waveform distinguishes the proposed approach from other published ones in that it completely avoids the problem of computing the Kalman filter gains when both the noise and disturbances are unknown, which represents real and practical scenarios. The evaluation on both simulated and real recorded waveforms have been successful and are very encouraging to extend this work further.

Acknowledgements

We acknowledge Stephane Dedieu for the help, advice, motivation and technical support on beam forming, echo cancellation and especially for providing the test bed for evaluation of the proposed scheme. The second author acknowledges the support provided by KFUPM, Saudi Arabia.

References

- [1] S. V. Vaseghi, *Multimedia Signal Processing Theory and Applications in Speech, Music and Communications*. John Wiley and Sons, 2007.
- [2] S. M. Kuo and D. R. Morgan. "Active Noise Control: A tutorial Review," in *Proceedings of the IEEE*, vol. 87, no. 6, 1999.
- [3] R. Doraiswami, "A two-stage Identification with Application to Control, Feature Extraction, and Spectral Estimation." in *IEE Proceedings: Control Theory and Applications*, vol. 152, no. 4, 2005, pp. 379-386.
- [4] R. Doraiswami and L. Cheded, "High-Order Least Squares Identification: A new Approach," in *Proceedings of the International Conference of Control, Dynamic Systems and Robotics*, Ottawa, ON, Canada, 2014.
- [5] N. Karimian and P. Geydecki. "Novel Digital Techniques for Echo Cancellation Applied to Speech Signals," in *Proceedings of the Institute of Acoustics*, vol. 35, 2013.
- [6] S. Gannot, D. Burshtein, and E. Weinstein, "Iterative and Sequential Kalman Filter-based Speech Enhancement Algorithms," in *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 4, pp. 373-385, 1998.
- [7] R. Doraiswami and L. Cheded, "A Unified Approach to Detection and Isolation of Parametric Faults Using a Kalman Filter Residuals," *Journal of Franklin Institute*, vol. 350, no. 5, pp. 938-965, 2013.
- [8] R. Doraiswami and L. Cheded. "Kalman Filter for Fault Detection: an Internal Model approach," *IET Control Theory and Applications*, vol. 6, no. 5, pp. 1-11, 2012.
- [9] R. Doraiswami and L. Cheded, "A Unifed Approach to Detection and Isolation of Faults using a Kalman Filter Residual-based Approach," Patent granted by United States, Patent 9002678, 2015.
- [10] R. Doraiswami, C. Diduch, and M. Stevenson. *Identification of Physical Systems: Applicationsto Condition Monitoring, Fault Diagnosis, Soft Sensor and Controller Design*. John Wiley and Sons, 2014.
- [11] S. Dedieu, "Acoustic and Audio Consulting," 2014.