Acoustic Echo Cancellation in Hands-Free Systems

Pathri Vidya Praveen, IIT Hyderabad

January 4, 2026

1 Introduction

Hands-free communication systems such as smart speakers, conferencing devices, and in-car voice assistants suffer from *acoustic echo*, caused by coupling between the loudspeaker and microphone. The loudspeaker signal propagates through an unknown acoustic path and is picked up by the microphone, degrading speech intelligibility and user experience.

This project implements a **real-time, industry-grade Acoustic Echo Cancellation (AEC) system** using **pure digital signal processing techniques**, without employing any machine learning or neural network models. The system dynamically estimates the acoustic echo path using adaptive filtering, robustly handles double-talk scenarios, and suppresses residual echo artifacts using classical non-linear processing techniques.

2 Problem Formulation

Let

- $x(n)$ denote the far-end (loudspeaker) signal,
- $h(n)$ denote the unknown acoustic echo path,
- $s(n)$ denote near-end speech,
- $d(n)$ denote the microphone signal.

The microphone signal can be modeled as

$$d(n) = (x * h)(n) + s(n) + v(n),$$

where $v(n)$ represents background noise.

The objective of acoustic echo cancellation is to estimate an adaptive filter $\hat{h}(n)$ such that the residual error

$$e(n) = d(n) - (x * \hat{h})(n)$$

minimizes the echo component while preserving near-end speech.

3 System Overview

The proposed AEC system consists of the following major components:

1. Partitioned Block Frequency-Domain Adaptive Filter (PBFDAF)
2. Coherence-based Double-Talk Detection (DTD)
3. Non-Linear Processing (NLP) with gain smoothing
4. Overlap-save FFT-based block processing

All components operate in real time and are designed with portability to embedded hardware platforms in mind.

4 Adaptive Echo Cancellation (PBFDAF)

4.1 Motivation for Frequency-Domain Processing

Acoustic echo paths in real environments often extend over hundreds of milliseconds. Implementing adaptive filtering in the time domain using NLMS becomes computationally expensive for such long impulse responses. Frequency-domain adaptive filtering significantly reduces complexity from $O(N^2)$ to $O(N \log N)$ by leveraging FFT-based convolution.

4.2 Partitioned Block Frequency-Domain Adaptive Filter

The echo path is partitioned into multiple frequency-domain blocks. The estimated echo spectrum is computed as

$$\hat{Y}(k) = \sum_{p=0}^{P-1} W_p(k) X_p(k),$$

where $W_p(k)$ are the adaptive filter partitions and $X_p(k)$ are delayed FFT blocks of the far-end signal.

4.3 NLMS Update Rule

The adaptive filter is updated using a frequency-domain NLMS rule:

$$W_p(k) \leftarrow W_p(k) + \mu \frac{E(k) X_p^*(k)}{\sum_p |X_p(k)|^2 + \epsilon},$$

where μ is the step size and ϵ ensures numerical stability. This formulation provides fast convergence and scale invariance.

5 Double-Talk Detection

5.1 Need for Double-Talk Detection

During double-talk scenarios, where both near-end and far-end speech are present, updating the adaptive filter can lead to divergence and corruption of the echo path estimate. Therefore, adaptation must be suspended during such periods.

5.2 Coherence-Based Detection

Magnitude-squared coherence is defined as

$$\gamma^2(k) = \frac{|P_{xd}(k)|^2}{P_{xx}(k)P_{dd}(k)},$$

where $P_{xx}(k)$ and $P_{dd}(k)$ are auto-spectral densities and $P_{xd}(k)$ is the cross-spectral density.

A block is classified as double-talk if the average coherence across frequency bins falls below a predefined threshold:

$$\frac{1}{K}\sum_{k}\gamma^2(k) < \tau.$$

This method is robust, energy-independent, and widely adopted in commercial AEC systems.

6 Non-Linear Processing

Residual echo persists even after adaptive cancellation due to model mismatch and non-linearities. A classical non-linear processor is employed to suppress this residual echo.

6.1 Residual Echo Estimation

The residual echo magnitude is estimated as

$$\hat{E}_{res}(k) = \alpha|\hat{Y}(k)|,$$

where α controls suppression aggressiveness.

6.2 Gain Computation

The frequency-dependent gain is computed as

$$G(k) = \max\left(\frac{|E(k)| - \hat{E}_{res}(k)}{|E(k)| + \epsilon}, G_{\min}\right).$$

6.3 Gain Smoothing

To prevent musical noise and perceptual artifacts, temporal smoothing is applied:

$$\tilde{G}(k,n) = \beta\tilde{G}(k,n-1) + (1-\beta)G(k,n).$$

Gain smoothing is mandatory for real-time AEC systems.

7 Experimental Setup

7.1 Simulation Parameters

- Sampling rate: 16 kHz

- Block size: 256 samples
- FFT size: 512
- Echo path: synthetic exponentially decaying impulse response

7.2 Evaluation Metrics

Performance is evaluated using:

- Echo Return Loss Enhancement (ERLE)
- Adaptive filter norm
- Coherence-based DTD behavior
- Time-domain waveform comparison

8 Results and Analysis

8.1 ERLE vs Time

ERLE over time demonstrates convergence behavior and steady-state echo suppression performance.

8.2 Filter Norm vs Time

The ℓ_2-norm of the adaptive filter remains bounded, confirming numerical stability and absence of divergence.

8.3 Double-Talk Detector Behavior

Coherence remains high during far-end-only periods and drops sharply during double-talk, enabling reliable detection and safe suspension of adaptation.

8.4 NLP Gain Behavior

Smoothed gain trajectories prevent musical noise and ensure perceptually acceptable residual echo suppression.

8.5 Waveform Comparison

Time-domain waveform comparison shows significant attenuation of the echo component while preserving signal structure.

9 Real-Time Implementation

The system was implemented in real time using Python and PortAudio via the `sounddevice` interface. Key considerations included conservative step-size selection, numerical safeguards, and gating of adaptation during detected double-talk.

10 Limitations

- Linear echo path assumption
- No explicit modeling of loudspeaker non-linearities
- Fixed detection thresholds

11 Conclusion

This project demonstrates a complete, production-grade acoustic echo cancellation system using classical DSP techniques. The integration of PBFDAF, coherence-based DTD, and robust non-linear processing achieves reliable echo suppression under realistic operating conditions.

12 Future Work

- Embedded fixed-point implementation
- Dynamic threshold adaptation
- Loudspeaker non-linearity compensation
- Multi-microphone extensions