Acoustic Echo Cancellation Using Partitioned Block Frequency Domain Adaptive Filtering

DSP Design Challenge Submission ,Team - NeuralNinja

Abstract

Acoustic Echo Cancellation (AEC) is a critical component of hands-free communication systems where microphone signals are corrupted by delayed reflections of loudspeaker output. This paper presents a complete, classical Digital Signal Processing (DSP) based AEC system using a Partitioned Block Frequency Domain Adaptive Filter (PBFDAF) with NLMS adaptation. Robustness is achieved using coherence-based and energy-based double-talk detection, along with a frequency-domain non-linear post-processor (NLP) for residual echo suppression. The system is implemented without the use of machine learning or pre-trained models. Experimental evaluation using synthetic yet realistic test signals demonstrates stable convergence and significant Echo Return Loss Enhancement (ERLE).

1 Introduction

Hands-free audio devices such as smart speakers, conferencing systems, and in-vehicle voice assistants are highly susceptible to acoustic echo. This phenomenon occurs when audio emitted by a loudspeaker propagates through the environment and is captured by the device microphone along with the desired near-end speech. If left untreated, acoustic echo results in degraded speech quality and poor user experience.

Adaptive filtering has long been established as an effective solution to the acoustic echo problem. However, practical AEC systems must address several challenges, including long and time-varying echo paths, limited computational resources, and the presence of double-talk scenarios where near-end and far-end speakers are active simultaneously.

This work focuses on a purely DSP-driven AEC solution that emphasizes algorithmic transparency and real-time feasibility. The design avoids machine learning approaches and instead relies on frequency-domain adaptive filtering, robust double-talk detection, and classical residual echo suppression techniques.

2 Signal Model

The microphone signal is modeled as:

$$d(n) = y(n) + s(n) + v(n) \tag{1}$$

where $y(n)$ is the acoustic echo, $s(n)$ is the near-end speech signal, and $v(n)$ denotes background noise.

The echo signal is given by:

$$y(n) = \sum_{k=0}^{L-1} h(k)x(n-k) \tag{2}$$

where $x(n)$ is the far-end reference signal and $h(k)$ represents the room impulse response (RIR).

The goal of the AEC system is to estimate $\hat{y}(n)$ such that the residual signal:

$$e(n) = d(n) - \hat{y}(n) \tag{3}$$

contains minimal echo energy while preserving near-end speech.

3 System Architecture

The proposed AEC system consists of the following components:

- Pre-emphasis filtering
- Overlap-save block processing
- Partitioned Block Frequency Domain Adaptive Filter (PBFDAF)
- NLMS-based weight adaptation
- Double-talk detection (DTD)
- Non-linear post-processing (NLP)
- Performance evaluation using ERLE

Block-based frequency-domain processing is used throughout the system to ensure computational efficiency and low latency.

4 Partitioned Block Frequency Domain Adaptive Filtering

Long acoustic impulse responses make time-domain adaptive filters computationally expensive. The PBFDAF addresses this challenge by dividing the adaptive filter into multiple partitions and performing convolution in the frequency domain.

The echo estimate in the frequency domain is computed as:

$$\hat{Y}(k) = \sum_{p=0}^{P-1} W_p(k)X_p(k) \tag{4}$$

where $W_p(k)$ are frequency-domain filter weights and $X_p(k)$ are delayed spectra of the far-end signal.

4.1 NLMS Adaptation

Filter weights are updated using the Normalized Least Mean Squares (NLMS) algorithm:

$$W_p(k) \leftarrow W_p(k) + \mu \frac{X_p^*(k)E(k)}{\sum_p |X_p(k)|^2 + \epsilon} \tag{5}$$

To ensure stability, the system incorporates:

- Leakage-based weight decay
- Global weight norm clipping
- Adaptive step-size reduction after convergence

5 Double-Talk Detection

During double-talk, the adaptive filter must be frozen to avoid divergence.

5.1 Coherence-Based Detection

Magnitude-squared coherence between the far-end and microphone signals is computed as:

$$\gamma^2(k) = \frac{|P_{xd}(k)|^2}{P_{xx}(k)P_{dd}(k)} \tag{6}$$

A low average coherence indicates double-talk, causing adaptation to be suspended.

5.2 Energy-Based Detection

A secondary energy-based detector compares the error energy with far-end signal energy. This detector provides a lightweight fallback mechanism during rapid near-end activity changes.

6 Residual Echo Suppression

Residual echo remains even after adaptive filtering due to modeling errors and non-linearities. A frequency-domain non-linear post-processor (NLP) is applied:

$$G(k) = \frac{|E(k)| - \alpha|\hat{Y}(k)|}{|E(k)| + \epsilon} \tag{7}$$

The gain is bounded to avoid excessive speech distortion. Temporal exponential smoothing is applied to reduce musical noise artifacts.

7 Experimental Setup

7.1 Test Signal Generation

The far-end signal is taken from real speech recordings. A synthetic room impulse response with exponential decay is generated to simulate the echo path. Background noise is added to replicate realistic conditions.

7.2 System Parameters

Table 1: System Parameters

Parameter	Value
Sampling Rate	16 kHz
Block Size	256
FFT Size	512
Filter Partitions	16
Step Size	0.15

8 Performance Evaluation

Echo Return Loss Enhancement (ERLE) is used to quantify echo suppression:

$$\mathrm{ERLE} = 10\log_{10}\left(\frac{E[y^2(n)]}{E[e^2(n)]}\right) \tag{8}$$

Evaluation is restricted to steady-state operation by excluding convergence periods and inactive far-end segments.

9 Conclusion

This work presents a complete and robust acoustic echo cancellation system based entirely on classical DSP principles. The combination of PBFDAF, NLMS adaptation, coherence-based double-talk detection, and frequency-domain non-linear post-processing achieves effective echo suppression while maintaining stability under challenging conditions. The design is computationally efficient and suitable for real-time and embedded deployment.

10 Future Work

Future enhancements include real-time implementation on embedded hardware, adaptive step-size scheduling, perceptually motivated post-processing, and integration with speech enhancement modules.