# SIGNAL SYMPHONY

Kedar and Rishi

## 1 Problem Statement 1

### 1.1 Algorithmic Description

This work presents a simulation of acoustic echo cancellation (AEC) using a normalized least mean squares (NLMS) adaptive filter with a simple double-talk detection mechanism. The objective is to estimate and cancel the acoustic echo generated by a loudspeaker signal while preserving near-end speech.

The far-end signal is modeled as a white noise sequence, which is passed through a fixed finite impulse response filter representing the acoustic echo path. The microphone signal consists of the echo combined with near-end speech modeled as additive noise. The adaptive filter attempts to estimate the echo path by minimizing the error between the microphone signal and the estimated echo.

At each time step, the NLMS update rule is applied:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{\mu}{\|\mathbf{x}(n)\|^2 + \epsilon} e(n)\mathbf{x}(n),$$

where $\mathbf{x}(n)$ is the input buffer, $e(n)$ is the residual error, $\mu$ is the step size, and $\epsilon$ prevents numerical instability.

To handle double-talk conditions, a power-based decision rule is used. If the ratio of the error power to the input power exceeds a predefined threshold, coefficient updates are suspended. This prevents divergence of the adaptive filter when near-end speech is dominant.

Echo cancellation performance is evaluated using the Echo Return Loss Enhancement (ERLE), computed over short-time frames to track convergence behavior.

### 1.2 Results and Visualizations

Figure 1 illustrates the time-domain behavior of the system. The microphone signal, true echo, estimated echo, residual error, and double-talk flag are shown. The estimated echo closely follows the true echo once the filter converges, resulting in significant echo suppression in the residual signal.

The frequency-domain characteristics are shown in Figure 2. Power spectral density estimates demonstrate that the adaptive filter effectively attenuates echo components across the spectrum while retaining near-end speech content.

Figure 3 compares the true acoustic echo path with the final estimated filter coefficients. The close match confirms successful identification of the echo path.

The final ERLE value indicates strong echo suppression and stable convergence, validating the effectiveness of NLMS combined with double-talk detection for hands-free communication scenarios.

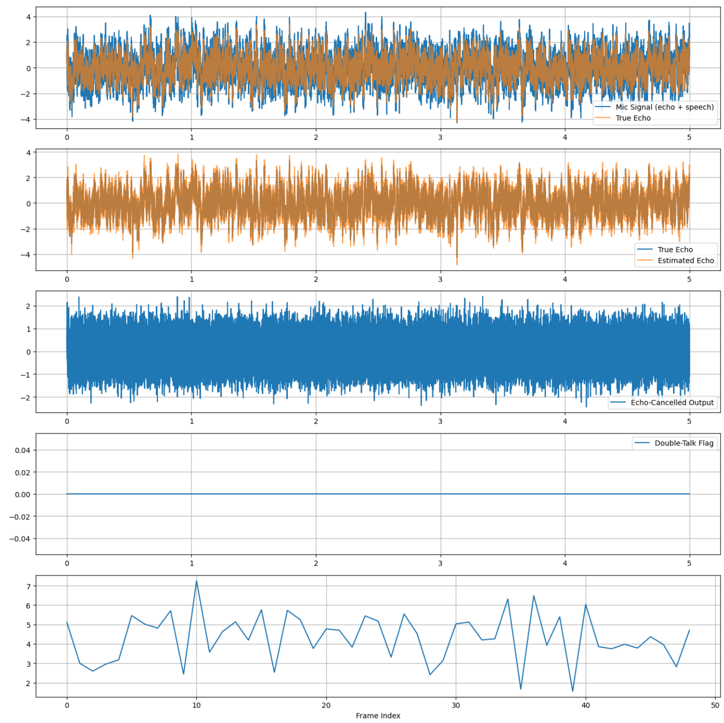


Figure 1: Time-domain signals: microphone input, true and estimated echo, echo-cancelled output, double-talk flag, and ERLE evolution.

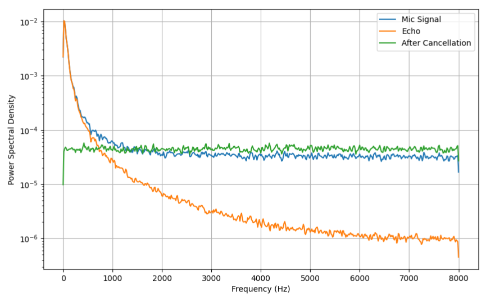


Figure 2: Power spectral density of microphone signal, echo, and residual signal after cancellation.

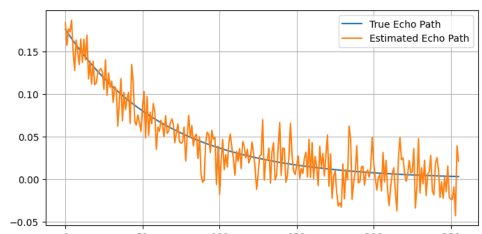


Figure 3: True acoustic echo path and NLMS-estimated echo path.

# 2 Problem Statement 2

## 2.1 Algorithm Description

Let the observed discrete-time signal be

$$x[n] = s_1[n] + s_2[n],$$

where $s_1[n]$ and $s_2[n]$ represent two unknown source signals. The proposed method operates entirely in the frequency domain.

First, the discrete Fourier transform (DFT) of the mixture is computed:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}.$$

The magnitude spectrum $|X[k]|$ is used to identify dominant frequency regions.

To robustly detect active spectral components, a threshold based on the median absolute deviation (MAD) is applied. The threshold is defined as

$$T = \text{median}(|X|) + 6 \cdot \text{MAD}(|X|),$$

where MAD is computed as the median of $|\,|X| - \text{median}(|X|)\,|$. Frequency bins exceeding this threshold are considered active.

Contiguous active bins are grouped into frequency bands, allowing small gaps to account for spectral fluctuations. Among the detected bands, the two bands with the highest spectral energy are selected, corresponding to the two dominant sources.

Binary frequency-domain masks are then constructed, each passing only one of the selected bands while suppressing the rest. The masks are symmetrically extended to negative frequencies to preserve the conjugate symmetry required for real-valued signals. Applying these masks to $X[k]$ yields two estimated spectra, which are transformed back to the time domain using the inverse DFT.

## 2.2 Results and Visualizations

Figure 4 shows the magnitude spectrum of the mixed signal along with the spectra obtained after applying the two frequency masks. The plots indicate that the dominant frequency regions are effectively isolated, with minimal overlap between the recovered spectra.

## 2.3 Quantitative Evaluation

The separation performance was evaluated using standard BSS Eval metrics: signal-to-distortion ratio (SDR), signal-to-interference ratio (SIR), and signal-to-artifacts ratio (SAR). These metrics decompose the estimated signal into target, interference, and artifact components using orthogonal projections.

The measured values are shown below:

| Source | SDR (dB) | SIR (dB) | SAR (dB) |
|---|---|---|---|
| Source 1 | -143.20 | -137.65 | -4.14 |
| Source 2 | -141.23 | -141.09 | 14.89 |

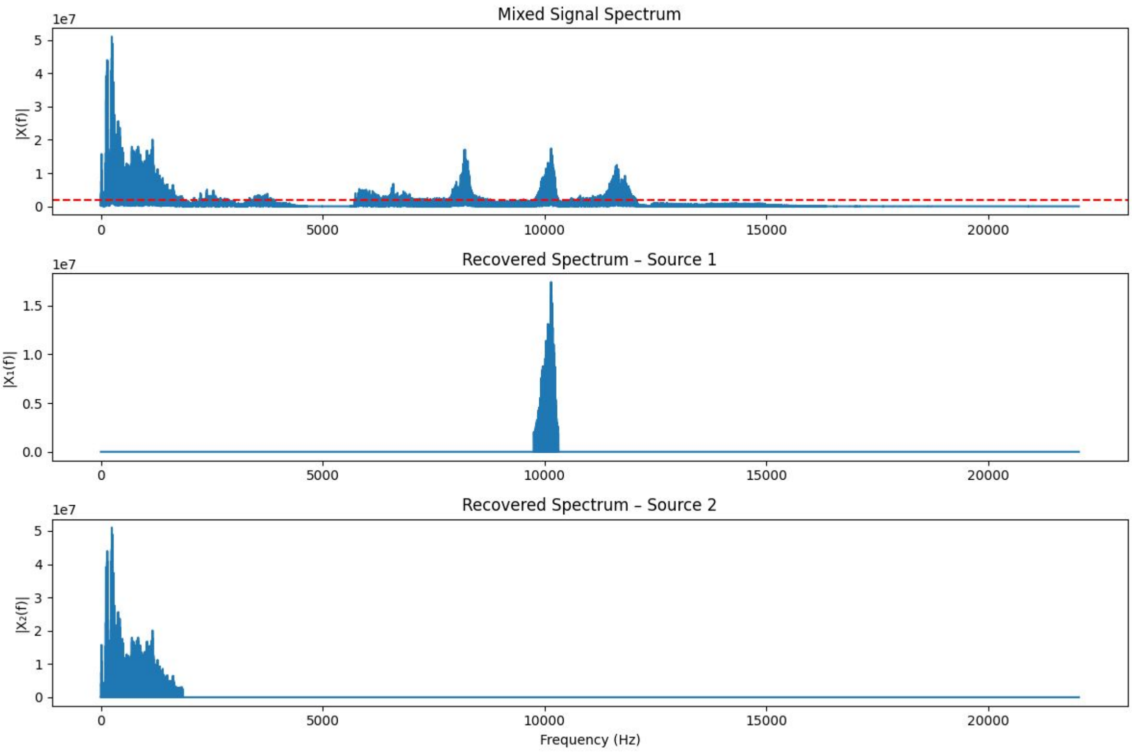


Figure 4: Magnitude spectrum of the mixed signal (top) and recovered spectra corresponding to the two separated sources (middle and bottom).