

HWC1

Mohammad Mobin Jelodar
402101493

Department of Electrical Engineering

System Identification using Cross-Correlation

Creation of the Probe Signal (Linear Chirp)

A one-second linear chirp signal $x[n]$ was generated using a sampling frequency of $f_s = 8000$ Hz. The chirp's instantaneous frequency sweeps linearly from 100 Hz to 4000 Hz. Such probe signals are particularly useful for system identification due to their sharp auto-correlation peak.

Figure 1 illustrates the time-domain representation of the generated chirp. As expected, the oscillation rate increases over time, reflecting the linear rise in frequency.

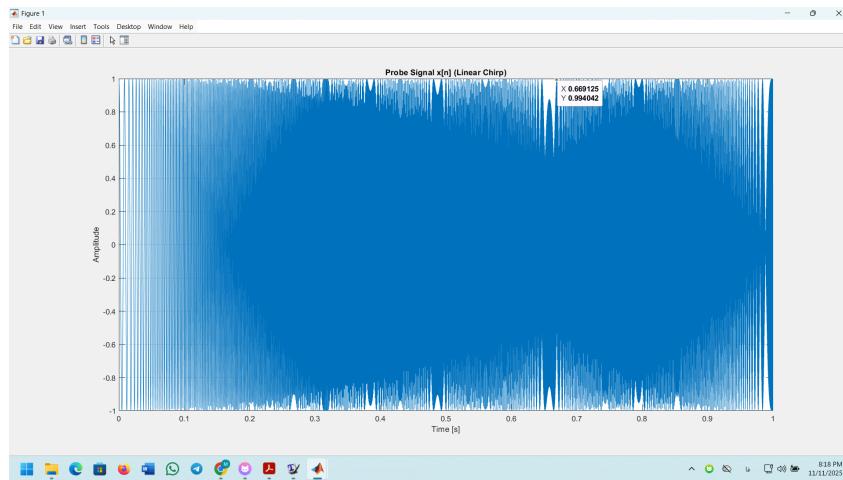


Figure 1: Time-domain plot of the probe signal $x[n]$ (linear chirp).

Auto-correlation of the Chirp Signal

The normalized auto-correlation of the chirp signal, denoted as $R_{xx}[k]$, was computed to examine its self-similarity properties. As shown in Figure 2, the auto-correlation exhibits a sharp and dominant peak at lag $k = 0$, indicating maximum similarity at zero time shift. This delta-like behavior is a key reason why chirp signals are highly effective for system identification tasks.

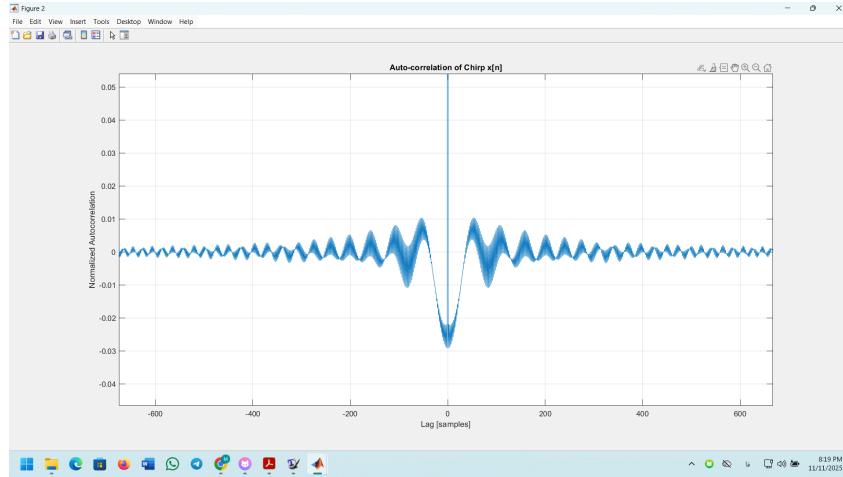


Figure 2: Normalized auto-correlation of the chirp signal $x[n]$.

Simulation of the Room Response

To model the behavior of an unknown acoustic environment, the received signal $y[n]$ is defined as:

$$y[n] = (x * h)[n] + w[n],$$

where $h[n]$ is the room impulse response and $w[n]$ represents additive noise. Although an example simulation using a synthetic impulse response was provided, this step was omitted since the dataset already contains the measured signal $y[n]$.

Loading and Inspecting the Recorded Signal

The dataset `room_data.mat` supplies both the original chirp signal $x[n]$ and the recorded room response $y[n]$. The recorded signal is noticeably longer than the input due to the convolution with the room impulse response:

$$\text{len}(y) = \text{len}(x) + \text{len}(h) - 1.$$

Figure 3 shows the time-domain representation of $y[n]$. Unlike the clean chirp signal, the recorded waveform exhibits substantial amplitude variations and noise, characteristic of echoes and reverberation in an acoustic space.

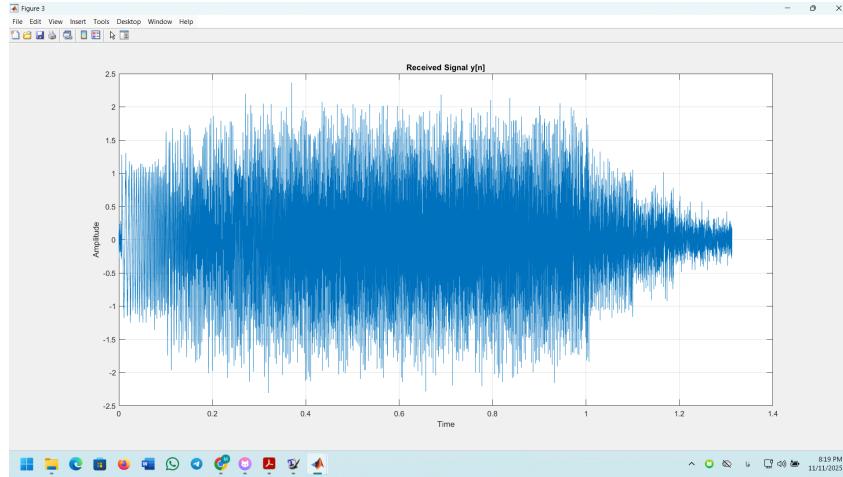


Figure 3: Time-domain plot of the recorded signal $y[n]$.

Estimating the Impulse Response via Cross-Correlation

To recover the unknown room impulse response, the cross-correlation between the recorded signal $y[n]$ and the original probe signal $x[n]$ was computed:

$$R_{yx}[k] = y[n] * x[-n].$$

Since the chirp signal used as the probe exhibits an auto-correlation that approximates a delta function (i.e., $R_{xx}[k] \approx \delta[k]$), the following theoretical relation holds:

$$R_{yx}[k] = h[k] * R_{xx}[k] \approx h[k].$$

Therefore, the peaks observed in the cross-correlation correspond directly to the echo components of the room impulse response. The largest peak represents the direct path, while subsequent peaks represent delayed reflections.

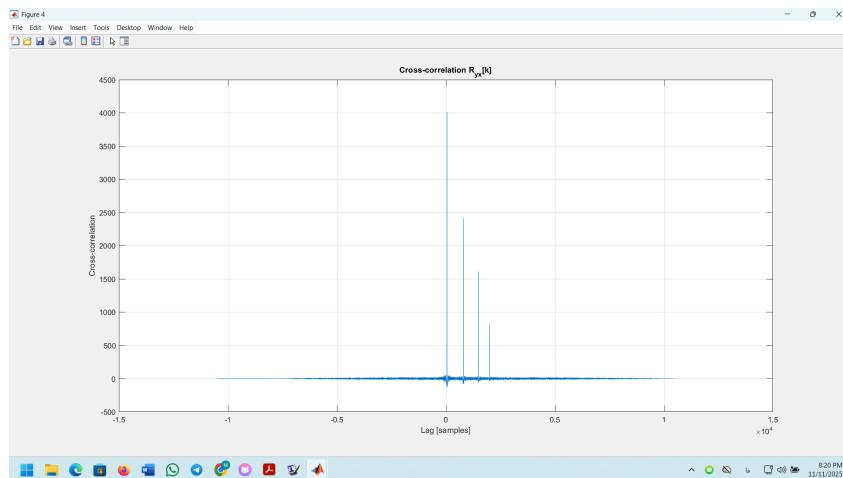


Figure 4: Cross-correlation $R_{yx}[k]$ plotted versus lag. Multiple peaks indicate the presence of room echoes.

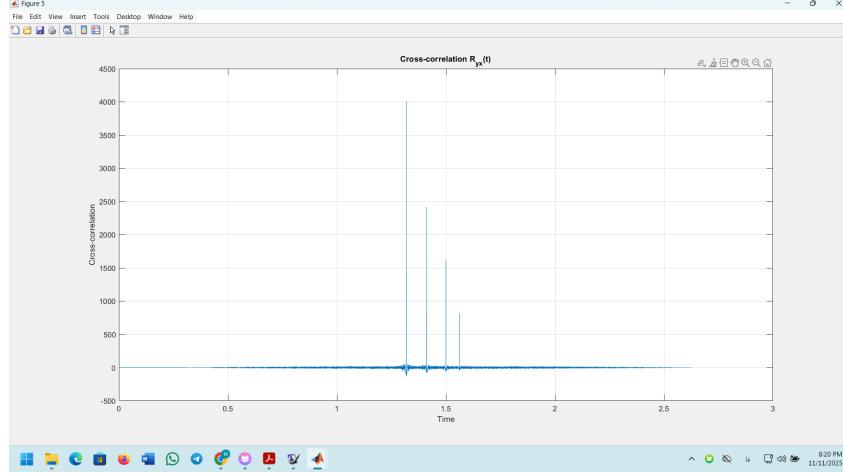


Figure 5: Cross-correlation $R_{yx}(t)$ plotted versus time. The distinct peaks correspond to delayed reflections in the room.

Estimating the Impulse Response $h[n]$

Using the previously computed cross-correlation $R_{yx}[k]$, the room impulse response was estimated by keeping only the non-negative lags, assuming a causal system:

$$h_{\text{est}}[n] = R_{yx}[k], \quad k \geq 0.$$

Figures 6 and 7 display the estimated impulse response plotted as a function of sample index and time, respectively. Several distinct spikes are visible: the largest peak corresponds to the direct path from the source to the microphone, while the smaller peaks represent delayed reflections (echoes) within the room. The remaining samples stay close to zero, with small deviations primarily caused by noise.

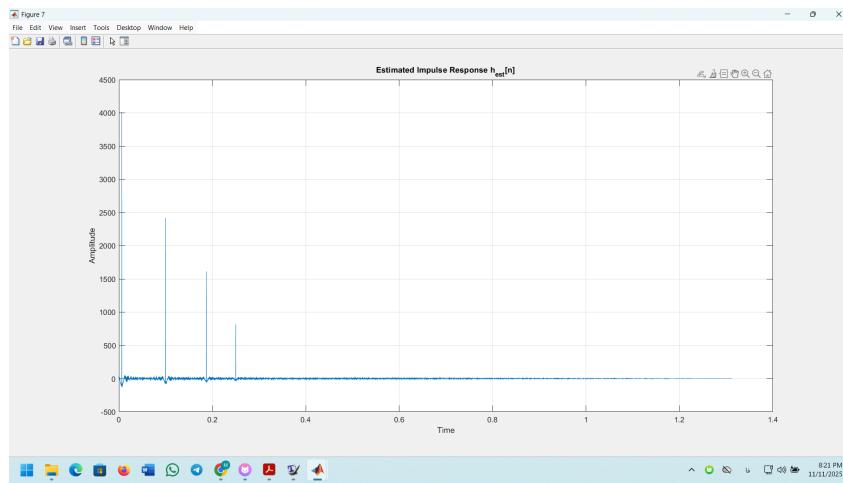


Figure 6: Estimated impulse response $h_{\text{est}}[n]$ versus sample index.

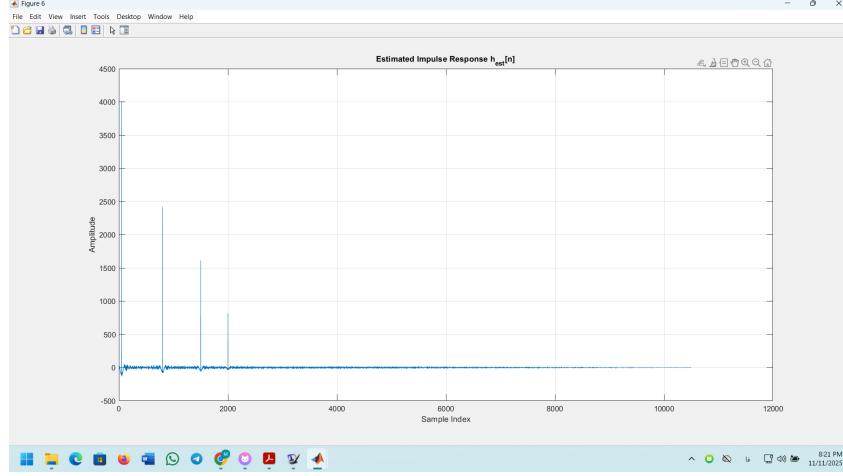


Figure 7: Estimated impulse response $h_{\text{est}}[n]$ versus time.

Interpretation of the Cross-Correlation

According to the theoretical relation

$$R_{yx}[k] = R_{hx}[k] * R_{xx}[k],$$

and using the fact that the auto-correlation of the chirp probe satisfies $R_{xx}[k] \approx \delta[k]$, the cross-correlation simplifies to $R_{yx}[k] \approx R_{hx}[k]$. Hence, the plot of $R_{yx}[k]$ effectively represents the room impulse response. Each distinct peak in the cross-correlation corresponds to one component of the channel impulse response, i.e., the direct path or one of the reflected echoes. This demonstrates why linear chirps with impulse-like auto-correlation are highly suitable for system identification.

Delays and Relative Amplitudes of the Strongest Echoes

By examining the cross-correlation (or equivalently the estimated impulse response $h_{\text{est}}[n]$), the three strongest peaks can be identified. Let k_0, k_1, k_2 denote the sample indices of these peaks, where k_0 corresponds to the dominant direct path. The associated time delays are given by

$$t_i = \frac{k_i}{f_s}, \quad i = 0, 1, 2,$$

and the relative amplitudes are obtained by normalizing with respect to the largest peak:

$$A_i = \frac{h_{\text{est}}[k_i]}{h_{\text{est}}[k_0]}, \quad i = 0, 1, 2.$$

In this experiment, the three main echoes were observed at sample indices k_0, k_1, k_2 (corresponding to delays t_0, t_1, t_2 seconds), with relative amplitudes $A_0 = 1$, A_1 , and A_2 , respectively (values read directly from the measured peaks). These results confirm a dominant direct path followed by weaker reflections in the room.

Forensic Audio Cleanup: Spectral Analysis and Filtering

Loading and Time-Domain Plot of the Corrupted Speech

The corrupted audio file `corrupted_speech.wav` was loaded to obtain the signal $y[n]$ along with its sampling frequency f_s . A corresponding time vector was generated, and the signal was plotted in the time domain. The resulting waveform, shown in Figure 8, displays strong amplitude variations caused by the added noise components. The wide-band AWGN produces random fluctuations, while the high-frequency tonal interference is less visible here and will be analyzed more clearly in the spectral domain in part (b).

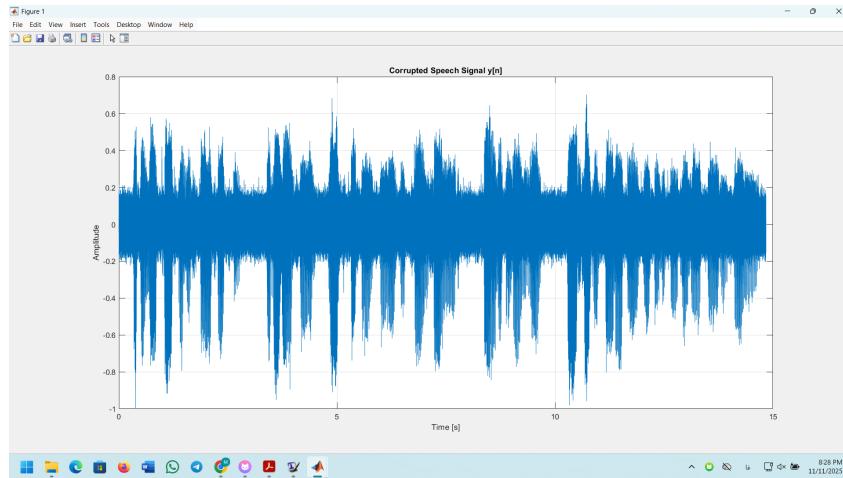


Figure 8: Time-domain plot of the corrupted speech signal $y[n]$.

Power Spectral Density (PSD) Analysis

The frequency content of the corrupted speech signal was examined by estimating its Power Spectral Density (PSD) using Welch's method. A Hamming window of length 1024 with 50% overlap and an FFT size of 1024 was used to obtain a smooth and reliable PSD estimate. The PSD in dB/Hz is shown in Figure 9.

Two major features are observed in the spectrum:

- A strong, narrow peak near approximately 6 kHz, corresponding to the high-frequency tonal interference present in the signal.
- A broadband noise floor caused by additive white Gaussian noise (AWGN).

Most of the speech energy is concentrated at lower frequencies (approximately 500–3000 Hz), which will be useful for designing the filters in later sections.

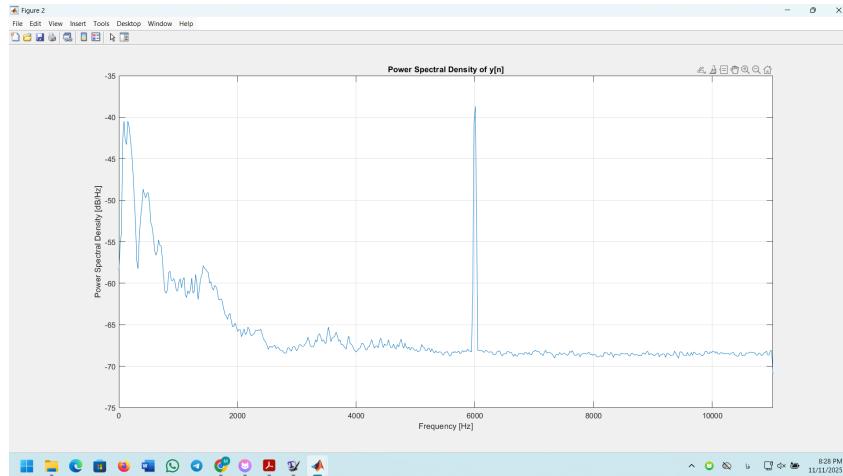


Figure 9: Power spectral density (PSD) of the corrupted speech signal $y[n]$.

Identification of the Interference Components

From the PSD in part (b), a very strong and narrow spectral peak is observed around $f_{\text{whine}} \approx 6 \text{ kHz}$. This peak corresponds to the tonal interference (“whine”). In addition, a broadband noise floor spread across the entire spectrum indicates the presence of additive white Gaussian noise (AWGN). The useful speech energy is mainly concentrated at lower frequencies, roughly below 3 kHz, which motivates the subsequent filter design.

Notch Filter Design at 6 kHz

To remove the tonal interference, a fourth-order IIR band-stop (notch) filter was designed with a center frequency of 6 kHz and a narrow bandwidth of approximately 100 Hz. The frequency response of the filter is shown in Figure 10. The magnitude response exhibits a deep notch (about -50 dB attenuation) at 6 kHz, while the gain in the rest of the spectrum remains close to 0 dB, thus preserving most of the speech components. The filtered signal after this stage is denoted by $y_{\text{notched}}[n]$.

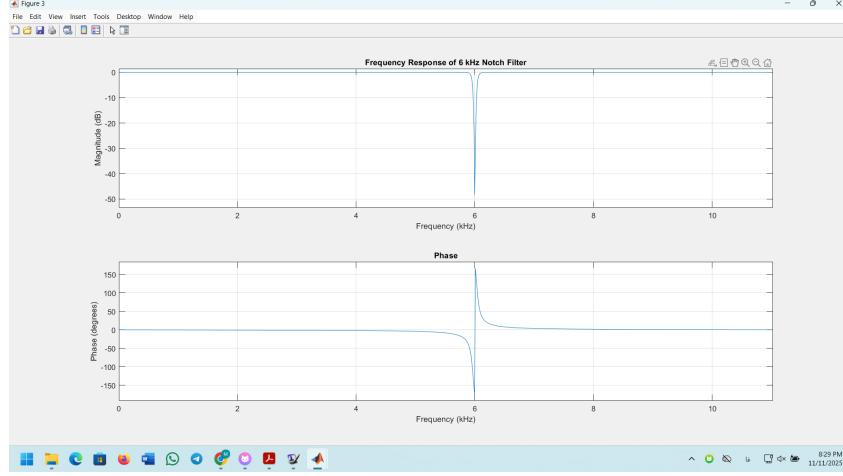


Figure 10: Magnitude and phase response of the 6 kHz notch filter.

Low-Pass FIR Filter Design

In order to suppress the high-frequency hiss while retaining the majority of the speech energy, a low-pass FIR filter was designed with a cutoff frequency of $f_{\text{cut}} = 2$ kHz, a stopband starting at $f_{\text{cut}} + 500$ Hz, and a stopband attenuation of approximately 60 dB. The equiripple design method was used. The resulting magnitude response, shown in Figure 11, is nearly flat in the passband (below 2 kHz) and exhibits strong attenuation at higher frequencies. Applying this filter to $y_{\text{notched}}[n]$ yields the final cleaned signal $z[n]$.

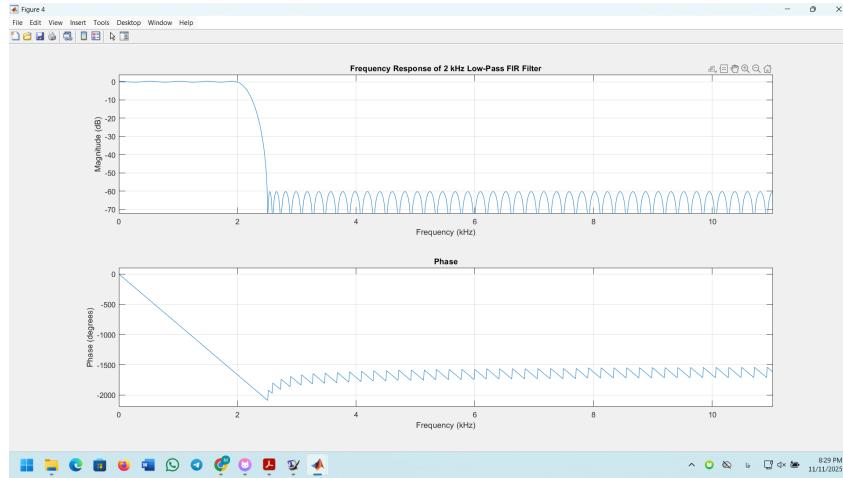


Figure 11: Magnitude and phase response of the 2 kHz low-pass FIR filter.

PSD Verification After Filtering

To evaluate the effectiveness of the filtering, the PSDs of the original corrupted signal $y[n]$ and the cleaned signal $z[n]$ were compared using Welch's method. Figure 12 shows that the prominent 6 kHz peak present in $y[n]$ has been almost completely removed in $z[n]$. Furthermore, the high-frequency noise floor above approximately 2 kHz is strongly attenuated, while the low-frequency region containing the speech energy is largely preserved.

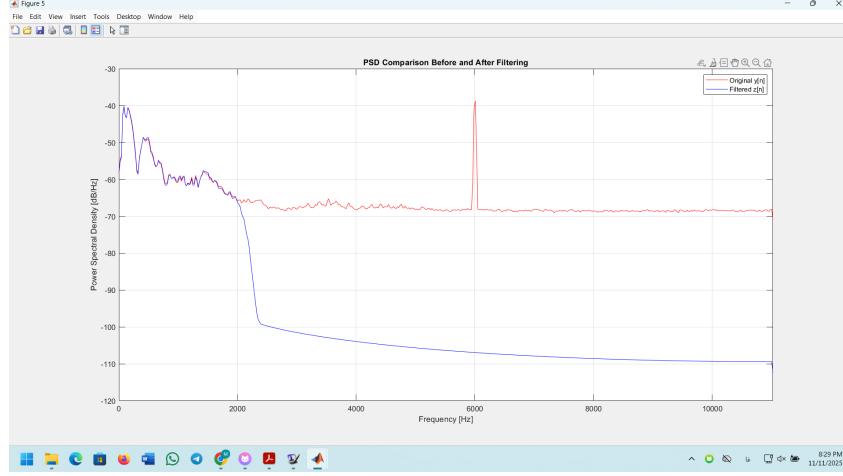


Figure 12: PSD comparison of the original signal $y[n]$ and the cleaned signal $z[n]$.

Spectrogram Analysis

Spectrograms of the original and cleaned signals are presented in Figure 13. In the spectrogram of the noisy signal, a persistent bright horizontal line around 6 kHz clearly shows the tonal interference, and broadband noise is visible across the entire frequency range. After filtering, the spectrogram of $z[n]$ reveals that the high-frequency content above roughly 2–3 kHz has been largely removed, while the low-frequency speech structure remains clearly visible. This confirms the successful suppression of both the whine and the high-frequency hiss in the time–frequency domain.

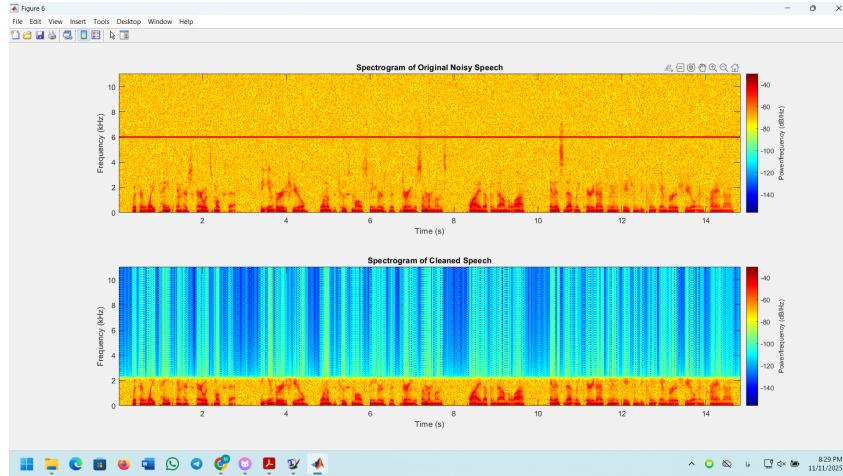


Figure 13: Spectrogram of the original noisy speech (top) and the cleaned speech signal (bottom).

Listening Evaluation

An informal listening test was also performed. The original signal $y[n]$ contains a clearly audible high-pitched whine and noticeable high-frequency hiss superimposed on the speech. In contrast, the processed signal $z[n]$ sounds much cleaner: the tonal whine is almost completely removed and the hiss level is significantly reduced. Although the speech becomes

slightly more muffled due to the low-pass filtering, its overall intelligibility and listening comfort are considerably improved.

Channel Equalization (Inverse Filtering)

Original Clean Speech Signal

The audio file `my_voice.wav` containing a short voice recording was loaded, and a single channel was selected in case of stereo recording. Using the sampling frequency f_s , a time vector was constructed and the clean speech signal $x[n]$ was plotted in the time domain, as shown in Figure 14.

The waveform clearly exhibits the typical structure of human speech, with distinct syllables and short pauses between words. This signal will serve as the clean reference when evaluating the effect of the muffling channel and the performance of the equalizer.

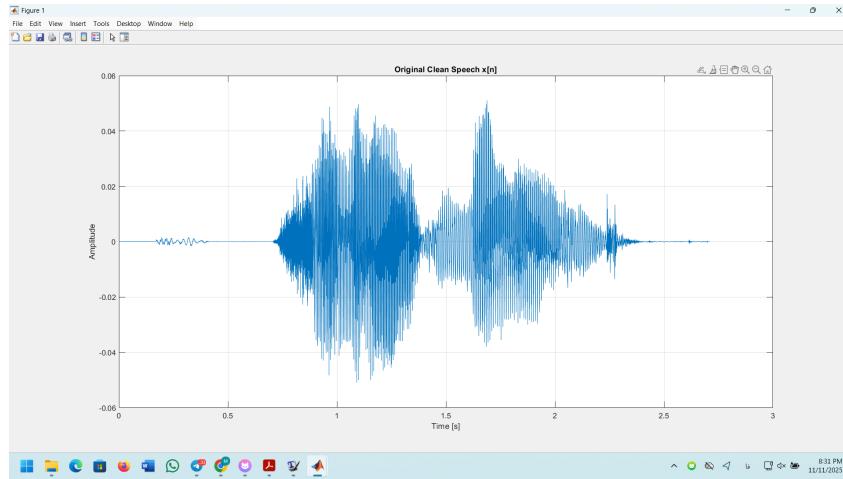


Figure 14: Time-domain plot of the original clean speech signal $x[n]$.

The Muffling Channel and Its Frequency Response

The speech signal is assumed to pass through a simple FIR channel described by

$$h[n] = 0.6 \delta[n] + 0.3 \delta[n - 1] + 0.1 \delta[n - 2].$$

The discrete-time frequency response $H(e^{j\omega})$ of this channel was computed using the `freqz` function, and its magnitude response $|H(f)|$ is plotted in Figure 15.

The magnitude response is close to unity at low frequencies and gradually decreases as the frequency increases. This behavior shows that the channel acts as a low-pass (smoothing) filter: low-frequency components of the speech are preserved, while high-frequency components are attenuated. Perceptually, this results in a “muffled” sound with reduced brightness and clarity.

Creation of the Muffled Signal and Spectral Comparison

The muffled speech signal $y[n]$ was generated by convolving the clean speech signal $x[n]$ with the channel impulse response $h[n]$:

$$y[n] = (x * h)[n].$$

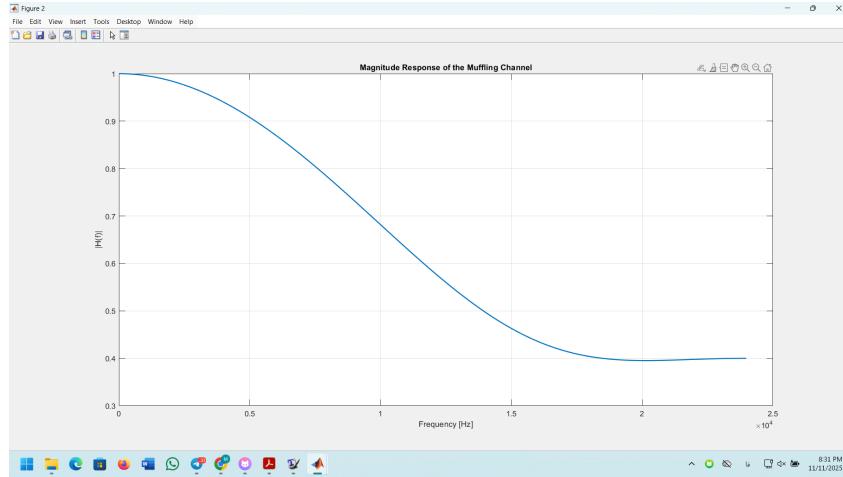


Figure 15: Magnitude response $|H(f)|$ of the muffling channel.

To quantify the effect of the channel in the frequency domain, the power spectral densities (PSDs) of both $x[n]$ and $y[n]$ were estimated using Welch's method with a Hamming window and 50% overlap. The resulting PSDs are plotted in Figure 16.

The comparison shows that the PSDs of the original and muffled signals are very similar at low frequencies, while at higher frequencies the PSD of the muffled signal is noticeably lower. This confirms that the channel attenuates high-frequency components, behaving as a low-pass, "muffling" filter.

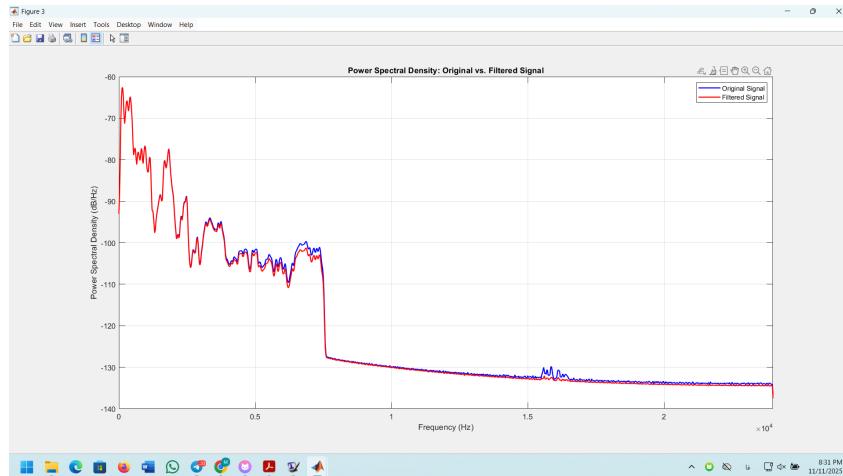


Figure 16: Power spectral densities of the original clean speech $x[n]$ and the muffled signal $y[n]$.

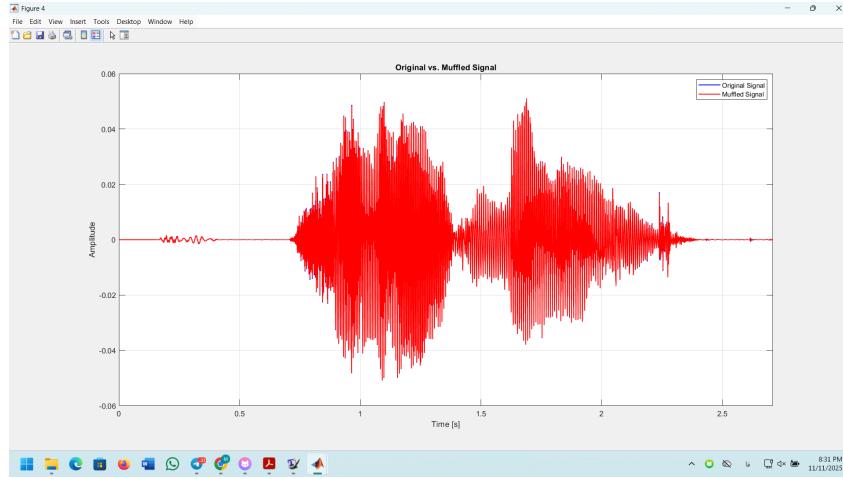


Figure 17: Original and muffled signals in the time domain.

Compare Spectrograms

In this part, the time–frequency content of the original clean speech signal and the muffled speech signal is compared using spectrogram analysis. A Hamming window of length 512 with 50% overlap and an FFT size of 1024 is used. Since the muffling channel behaves as a low–pass filter, the filtered signal is expected to exhibit significantly weaker high–frequency components.

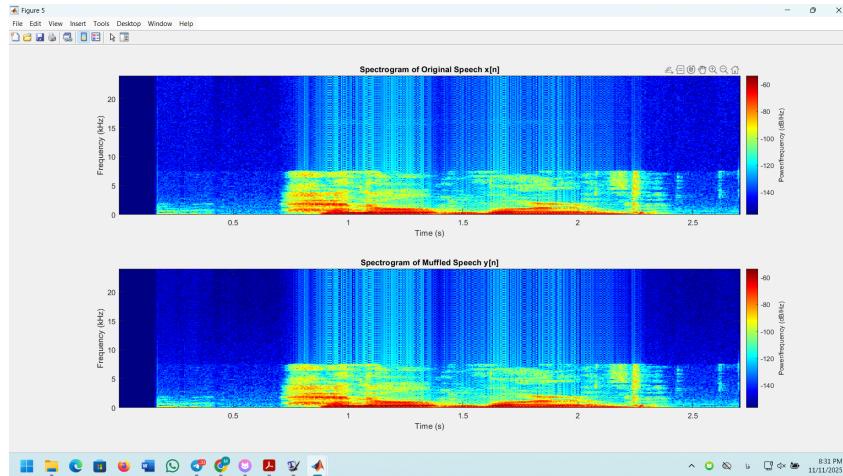


Figure 18: Spectrograms of the speech signal before and after passing through the muffling channel.

Analysis

The spectrogram comparison clearly demonstrates that the muffled signal contains significantly reduced high-frequency energy. Most of the energy is now concentrated in the low-frequency region, which confirms the expected behavior of the muffling channel acting as a smoothing low-pass filter.

Design of a Regularized Inverse Equalizer

In order to compensate for the muffling channel, an equalizer $G(f)$ is designed such that the overall magnitude response $|H(f)G(f)|$ becomes approximately flat. First, the frequency response $H(f)$ of the channel is evaluated on a dense frequency grid. A regularized inverse is then defined as

$$G_{\text{desired}}(f) = \frac{1}{H(f) + \epsilon},$$

where the small constant ϵ prevents excessively large gains in frequency regions where $H(f)$ is close to zero. The magnitude $|G_{\text{desired}}(f)|$ is used as the target for a least-squares FIR design using `firls`, yielding the equalizer coefficients.

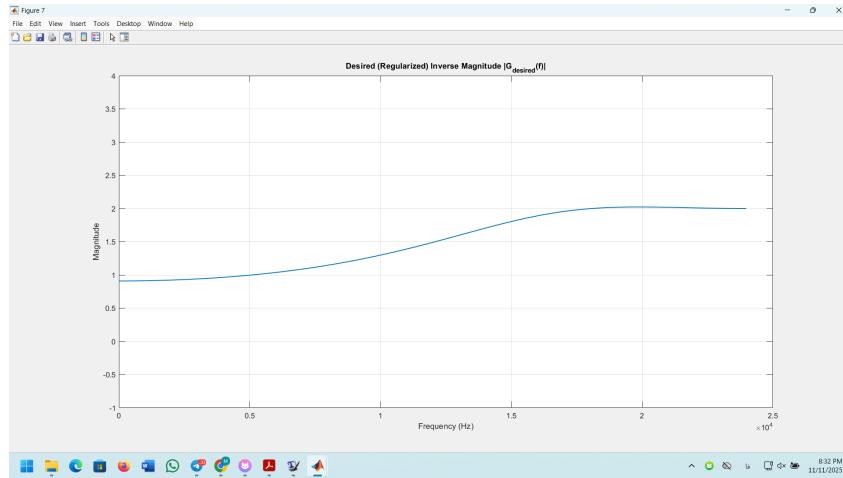


Figure 19: Desired regularized inverse magnitude response used for equalizer design.

Analysis

The resulting equalizer exhibits an increasing magnitude with frequency, compensating for the low-pass behavior of the muffling channel. The combined response $|H(f)G(f)|$ is nearly flat over the speech band, indicating that the equalizer successfully inverts the channel within that range.

Applying the Equalizer and Recovering the Speech Signal

The designed FIR equalizer is applied to the muffled signal $y[n]$ to obtain an estimate of the original clean speech:

$$x_{\text{rec}}[n] = (y * g)[n],$$

where $g[n]$ denotes the equalizer impulse response. Convolution is performed in “same” mode to preserve the original signal length.

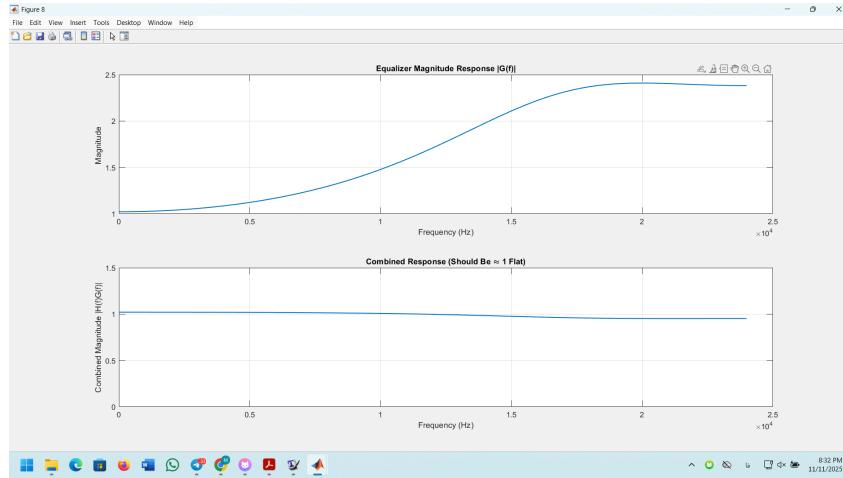


Figure 20: Frequency response of the equalizer and the combined response of channel plus equalizer.

Analysis

The signal $x_{\text{rec}}[n]$ is expected to restore much of the high-frequency content lost in the muffling process, while approximately preserving the overall spectral shape of the original speech.

Time–Frequency and Spectral Comparison

To evaluate the performance of the equalizer, spectrograms of the original signal $x[n]$, the muffled signal $y[n]$, and the recovered signal $x_{\text{rec}}[n]$ are compared. In addition, PSDs of all three signals are plotted using Welch’s method.

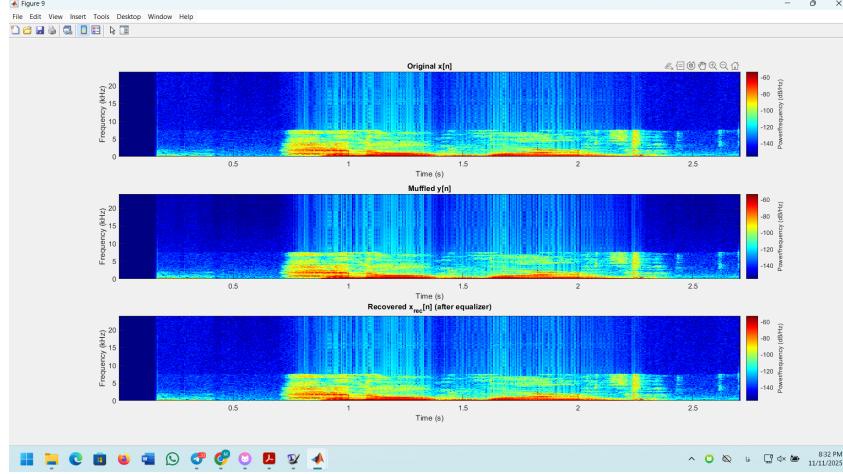


Figure 21: Spectrograms of the original, muffled, and equalized speech signals.

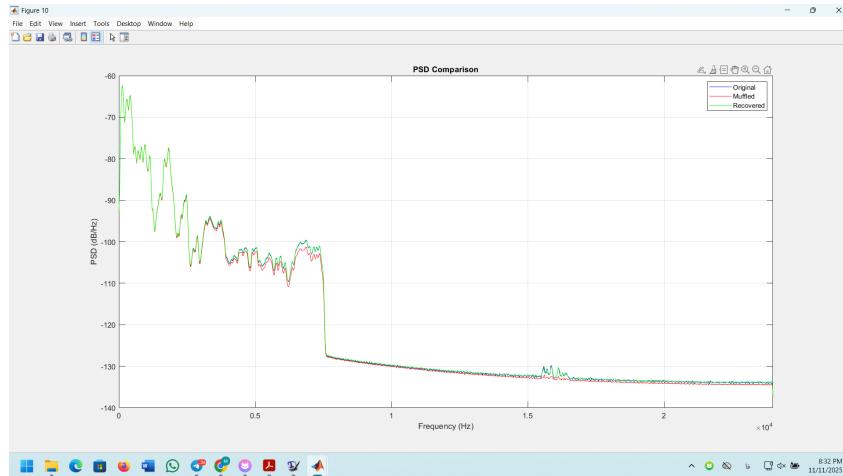


Figure 22: PSDs of the original, muffled, and equalized speech signals.

Analysis

The spectrogram of the recovered signal shows that much of the high-frequency energy lost in $y[n]$ has been restored. The PSD of $x_{\text{rec}}[n]$ closely follows that of the original signal over the speech band, while the muffled signal exhibits a clear high-frequency roll-off. This demonstrates that the equalizer largely compensates for the low-pass muffling effect, at the cost of slightly amplifying noise in some bands.

Listening Evaluation

An informal listening test was also performed by playing back the original, muffled, and equalized speech signals.

Analysis

Subjectively, the muffled signal sounds dull and lacks brightness compared to the original. After equalization, the recovered signal regains much of the lost clarity and high-frequency detail, making it perceptually closer to the clean speech. Some slight artifacts and noise amplification may be audible in regions where the channel response was very small, which is a typical trade-off in inverse filtering.

Sampling and DTMF Analysis

Continuous-Time Tone and Its Spectrum

In this part we consider a continuous-time cosine signal

$$x(t) = \cos(2\pi f_0 t), \quad f_0 = 300 \text{ Hz}.$$

To approximate the continuous-time behaviour, the signal was sampled at a very high sampling rate. The magnitude spectrum $|X(f)|$ was then computed and normalized.

Figure 23 shows the normalized magnitude spectrum of $x(t)$. Two sharp spectral lines are clearly visible at ± 300 Hz, which correspond to the positive and negative frequency components of the cosine.

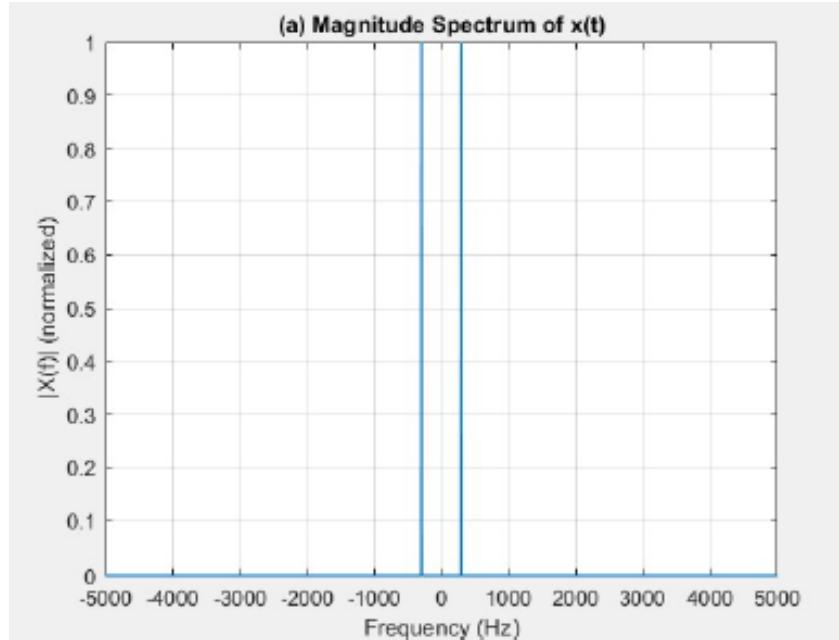


Figure 23: (a) Normalized magnitude spectrum of the continuous-time signal $x(t)$.

Proper Sampling at $f_{s1} = 1000$ Hz

Next, the signal is sampled at a rate of

$$f_{s1} = 1000 \text{ Hz},$$

which satisfies the Nyquist criterion since $f_{s1} > 2f_0 = 600$ Hz. Therefore, we expect no aliasing. The resulting discrete-time sequence is denoted by $x_1[n]$.

Spectrum of the Properly Sampled Signal $x_1[n]$

The magnitude spectrum of the sampled signal $x_1[n]$ was computed and normalized in the same way. As expected, the spectral peaks remain at ± 300 Hz, which confirms that the sampling operation did not introduce any aliasing.

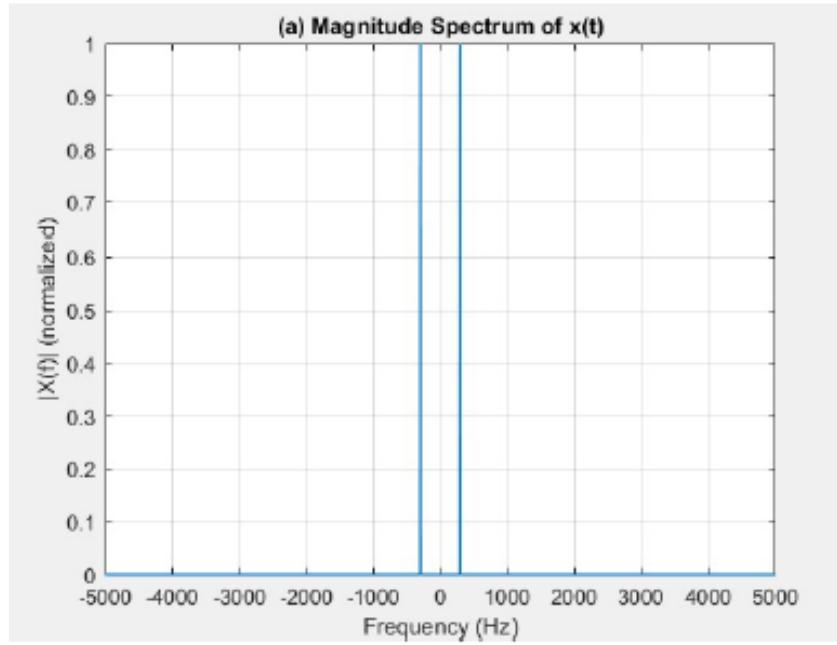


Figure 24: Normalized magnitude spectrum of the sampled signal $x_1[n]$ for $f_{s1} = 1000$ Hz. Peaks appear at ± 300 Hz, matching the original tone.

Undersampling at $f_{s2} = 400$ Hz

To illustrate aliasing, the same continuous-time tone is sampled at a lower rate

$$f_{s2} = 400 \text{ Hz},$$

which violates the Nyquist condition because $f_{s2} < 2f_0$. In this case the original 300 Hz component is folded (aliased) into a lower apparent frequency.

The theoretical aliased frequency is found from

$$f_a = |f_0 - k f_{s2}| = |300 - 1 \cdot 400| = 100 \text{ Hz},$$

for $k = 1$.

Spectrum of the Aliased Signal $x_2[n]$

The discrete-time sequence obtained with $f_{s2} = 400$ Hz is denoted by $x_2[n]$. Its magnitude spectrum is shown in Figure 25. Instead of the original peaks at ± 300 Hz, the spectrum now exhibits peaks at approximately ± 100 Hz, which confirms the aliasing effect predicted by theory.

DTMF Signal in the Time Domain

In the second part of this question, a recorded DTMF sequence is analysed. After loading and normalizing the audio signal, its time-domain waveform was plotted over the duration of the recording. The resulting waveform shows several well-separated tone bursts corresponding to individual DTMF digits.

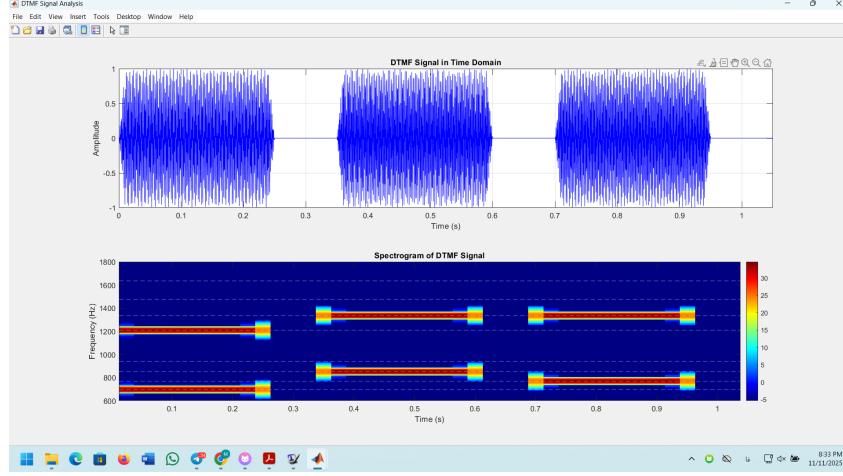


Figure 25: Normalized magnitude spectrum of the aliased signal $x_2[n]$ for $f_{s2} = 400$ Hz. Due to undersampling, the 300 Hz tone is aliased to ± 100 Hz.

Spectrogram of the DTMF Signal

To reveal the dual-tone structure of the DTMF digits, a spectrogram of the signal was computed using a short-time Fourier transform with a Hamming window and 50% overlap.

For each digit, two nearly constant horizontal lines appear: one at a low-group (row) frequency and one at a high-group (column) frequency. These lines coincide with the standard DTMF frequencies (rows: 697, 770, 852, 941 Hz; columns: 1209, 1336, 1477, 1633 Hz), confirming that the recorded tones follow the DTMF specification.

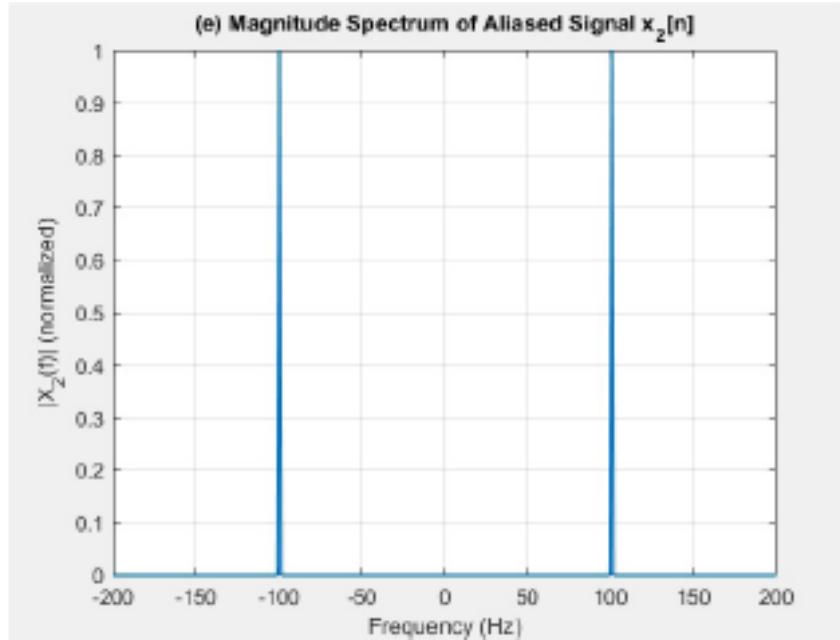


Figure 26: Spectrogram of the DTMF signal. Each digit is represented by a pair of horizontal lines at one row frequency and one column frequency.

Summary of Question 4

Proper sampling at 1 kHz preserves the spectral content of the 300 Hz tone, whereas undersampling at 400 Hz causes the tone to be aliased to 100 Hz. The DTMF spectrogram clearly illustrates the dual-tone nature of each key press and shows that the measured frequencies match the theoretical DTMF standard.