Electronics and Electrical Communications Engineering Department

Faculty of Engineering

Cairo University

Submitted to: Dr. Mohsen Rashwan

Speech Processing Problems

DSP-1 Assignment 1 submitted for course ELC4011 "DSP-1 Applications"

4th Year

1st Semester - Academic Year 2025/2026

Prepared by:

NAME	SECTION	ID
Yousef Khaled Omar Mahmoud	4	9220984

Submission Date: 20 November 2025

1 TABLE OF CONTENTS

2 TABLE OF FIGURES

3 LIST OF TABLES

DSP-1 Assignment 1: Speech Processing Problems

Youssef Khaled Omar Mahmoud*[1]

*Electronics and Communication Department, Faculty of Engineering,

Cairo University Giza, 12613, Egypt

yousef.mahmoud03@eng-st.cu.edu.eg

Abstract: ***Since***

Key words: *).*

1 INTRODUCTION

.ss

A. *Audio File Information*

As shown in Table 1, the analysis in this report, particularly the practical implementation of windowing (Problem 1) and the context for LPC (Problem 2 & 3), is based on the following source audio file:

Table 1 : Audio File Information[2]

Parameter	Value
File	C02n_1.wav
Sampling Frequency F_s	16000 Hz
Bits Per Sample (Bit Depth)	16 bits
Number of Channels	1
Total Samples	12800
Calculated Bit Rate	256000 bits/s

2 ANALYZE THE FREQUENCY DOMAIN CHARACTERISTICS OF WINDOWS

This section presents the time-domain definitions and frequency-domain analysis for the Rectangular, Hanning, and Hamming window functions, which are fundamental tools in digital signal processing for frame-based analysis.

A. *Window Definitions*

Given a window length N, the three specified window functions are defined in the time domain as follows:[1]

1) Rectangular Window W_{Rec} [1]

$$W_{\mathrm{Rec}}(n) = \begin{cases} 1 \ for\ 0 \le n \le N-1 \\ 0 \ Otherwise \end{cases}$$

2) Hanning Window W_{Han} [1]

$$\mathrm{W_{Han}(n)} = 0.5 - 0.5\cos\left(\frac{2\pi \mathrm{n}}{\mathrm{N}}\right) \text{ for } 0 \le \mathrm{n} \le \mathrm{N}-1$$

3) Hamming Window W_{Ham} [1]

$$\mathrm{W_{Ham}(n)} = 0.54 - 0.46\cos\left(\frac{2\pi \mathrm{n}}{\mathrm{N}}\right) \text{ for } 0 \le \mathrm{n} \le \mathrm{N}-1$$

B. Time Domain Visualization

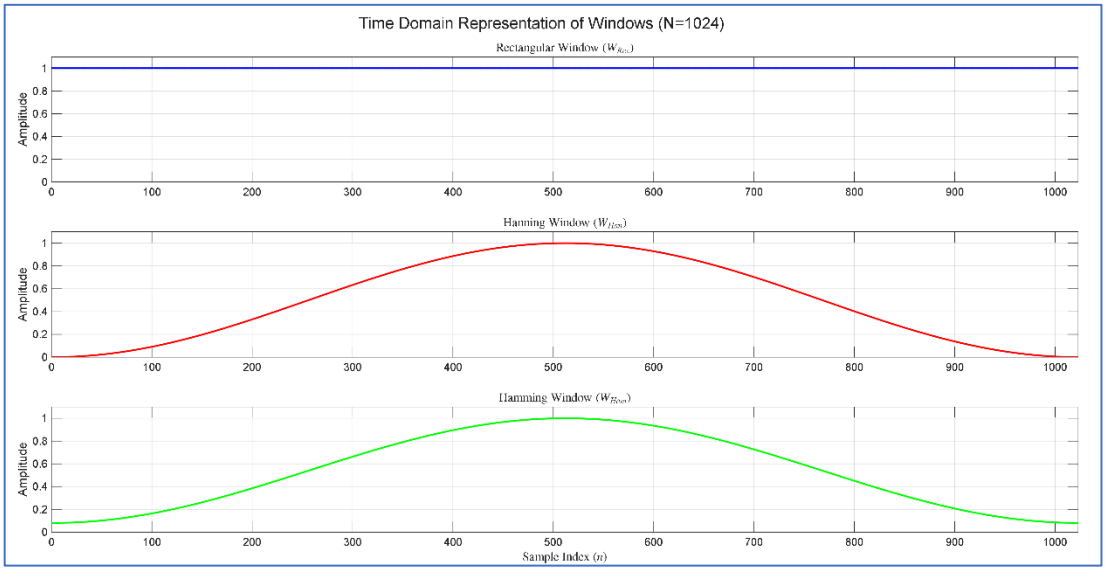

Figure 1 : Time Domain Representation of Rectangular, Hanning, and Hamming Windows (N=1024)[2]

As shown in Figure 1, the time domain representations of the three window functions for $N = 1024$ are visualized below. This plot demonstrates how the windows taper the input signal.

C. Frequency Domain Visualization

The frequency response,$W(f)$, is obtained by computing the Discrete-Time Fourier Transform (DTFT) of the time-domain window, typically approximated using the Fast Fourier Transform (FFT). The plots below show the log-magnitude spectrum, $20\log_{10}(|W(f)|)$, normalized such that the maximum peak is 0 dB.
As shown in Figure 2, the plots are centered around 0 Hz and zoomed to the range $[-F_s/16, F_s/16]$ H
Z to clearly observe the main lobe and the first few side lobes. The magnitude floor is clipped at -400 dB for visualization clarity.

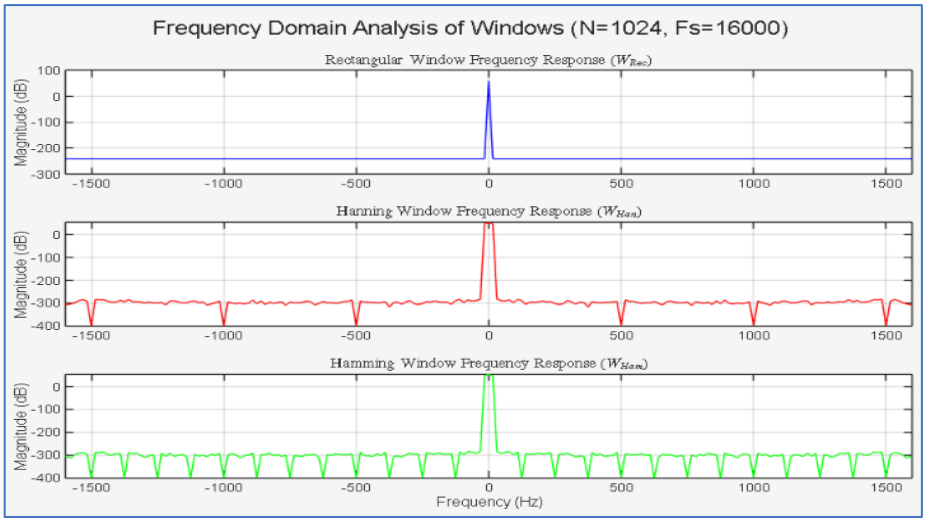

Figure 2 : Frequency Domain Analysis of Windows (Log Magnitude)[2]

D. Discussion and Comments

As shown in Table 2, the primary function of a window is to mitigate **spectral leakage** by smoothly forcing the signal to zero at the frame boundaries. The analysis of the windows in the frequency domain reveals the critical trade-off between **frequency resolution** (main lobe width) and **spectral leakage** (side lobe attenuation).

Table 2 : Comparison of Frequency Domain Characteristics for Rectangular, Hanning, and Hamming Windows

Characteristic	Rectangular (W_{Rec})	Hanning (W_{Han})	Hamming (W_{Ham})
Main Lobe Width	Narrowest (Highest Resolution)	Medium	Medium
Peak Side Lobe	Highest (≈ -13 dB)	Moderate (≈ -31 dB)	Lowest (≈ -41 dB)
Leakage/Attenuation	Worst Spectral Leakage	Good Attenuation	Best First Side Lobe Attenuation
Application	Analyzing transient, short bursts	General purpose, good compromise	Better for minimizing interference from strong nearby tones

Detailed Comments:

1) Rectangular Window (W_{Rec}):

- **Main Lobe:** Exhibits the narrowest main lobe, theoretically offering the best frequency resolution.
- **Side Lobes:** Its sudden truncation in the time domain results in the highest side lobes (only ≈ 13 dB down from the peak). This poor attenuation means energy from a strong frequency component "leaks" significantly into adjacent frequency bins, leading to severe spectral leakage.

2) Hanning Window (W_{Han}):

- **Tapering:** Provides a smooth taper to zero, which significantly reduces side lobe levels.
- **Side Lobes:** The peak side lobe is suppressed to about -31 dB This is a substantial improvement over the rectangular window.
- **Trade-off:** This suppression comes at the cost of a wider main lobe (approximately twice the width of the rectangular window), which slightly reduces frequency resolution.

3) Hamming Window (W_{Ham}):

- **Design:** The Hamming window is a modification of the Hanning window, designed specifically to minimize the height of the *first* side lobe.
- **Side Lobes:** Achieves the best first side lobe attenuation, suppressed to about -41 dB.
- **Trade-off:** While the first side lobe is the lowest, subsequent side lobes roll off more slowly than those of the Hanning window.

Conclusion: For applications like speech processing, where minimizing spectral leakage is crucial to accurately separating closely spaced harmonics and formants, the **Hanning** and **Hamming** windows are strongly preferred over the Rectangular window. The choice between Hanning and Hamming depends on whether a faster side lobe decay (Hanning) or the lowest possible first side lobe (Hamming) is desired.

3 Linear Predictive Coding Analysis based on Given Autocorrelation Data

E. Methodology: Solving the Yule-Walker Equations

Problem 2 requires the determination of the Linear Predictive Coding (LPC) coefficients, a_1 and a_2, and the minimum mean-squared prediction error, E_2, given the first three autocorrelation values of a signal, as discussed in the course material on **Speech Analysis** [3].

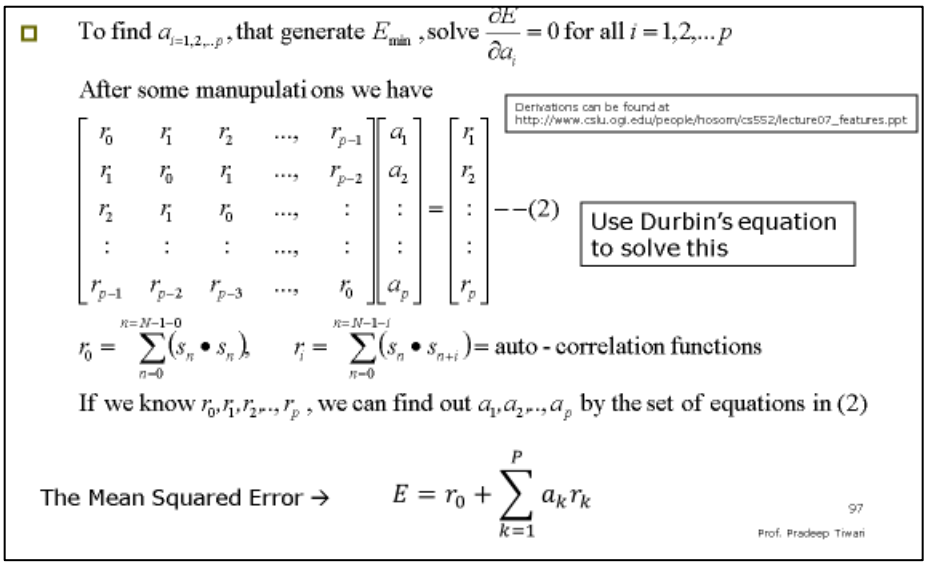

Figure 3 : LPC Calculations [3]

As shown in Figure 3, The solution is found by solving the **Yule-Walker system of equations**, which relates the autocorrelation values $\boldsymbol{R}$ to the LPC coefficients $\boldsymbol{a}$ and the minimum mean-squared error E_p. For a prediction order p, the matrix equation is:

$$R_p a = r_p$$

where:

- R_p is the $p \times p$ **Toeplitz matrix** containing autocorrelation values $R(|i-j|)$.
- $a = [a_1, a_2, \dots, a_p]^T$ is the vector of LPC coef
- $r_p = [R(1), R(2), \dots, R(p)]^T$ is the vector of shifted autocorrelation values.

The minimum mean-squared error E_p is calculated as:

$$E_p = R(0) + \sum_{k=1}^{p} a_k R(k)$$

F. Input Data and Results

The given autocorrelation data for a prediction order $p = 2$ is:

$$R = [R(0), R(1), R(2)] = [1.0000, 0.7000, 0.4000]$$

Substituting these values into the $p = 2$ Yule-Walker matrix equation:

$$\begin{bmatrix} 1.0 & 0.7 \\ 0.7 & 1.0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.4 \end{bmatrix}$$

As shown in Table 3, Solving this linear system using the implementation (e.g., using matrix inversion or the Levinson-Durbin algorithm) yielded the following results:

Table 3 : Problem 2 LPC Results [2]

Fra...	Samples (first 3)	R(0)	R(1)	R(2)	Error (E...	a1	a2
1	1 2 3	1.0000	0.7000	0.4000	1.5059	0.8235	-0.1765

4 A: Linear Predictive Coding (LPC) Analysis

This section details the results of applying Linear Predictive Coding (LPC) analysis to the short synthetic signal s[n] using specific framing parameters.

A. Analysis Parameters

Table 4 : Input Parameters

Parameter	Notation	Value	Description
Input Signal	$s[n]$	$[1,4,0,-4,-1,2,4,-1,2,5]$	10 samples total
Frame Size	N	6 samples	Length of the analysis window
Overlap	-	2 samples	Number of shared samples between frames
Frame Shift	R	4 samples	$R = N - \text{Overlap} = 6 - 2$
LPC Order	P	2	Number of prediction coefficients (a_k)

B. Frame Segmentation and Visualization

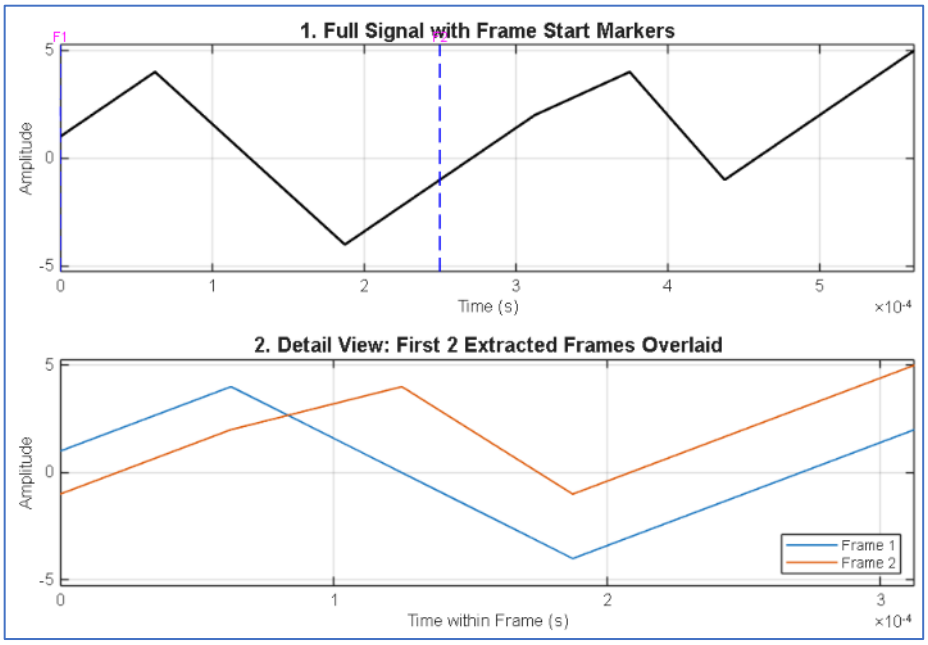

Figure 4 : Frame Signal Visualiztion [2]

As shown in Figure 4, The signal $s[n]$ was segmented into two overlapping frames based on the specified parameters $(N = 6, R = 4)$

- **Frame 1:** Starts at index $n = 1$. Contains samples $s[1]$ to $s[6]$

$$s_1 = [1,4,0,-4,-1,2]$$

- **Frame 2:** Starts at index $n = 5$ $(1 + R = 1 + 4)$. Contains samples $s[5]$ to $s[10]$

$$s_2 = [-1,2,4,-1,2,5]$$

The overall segmentation relative to the signal and the individual time-domain shape of each frame can be seen in Figure 5.

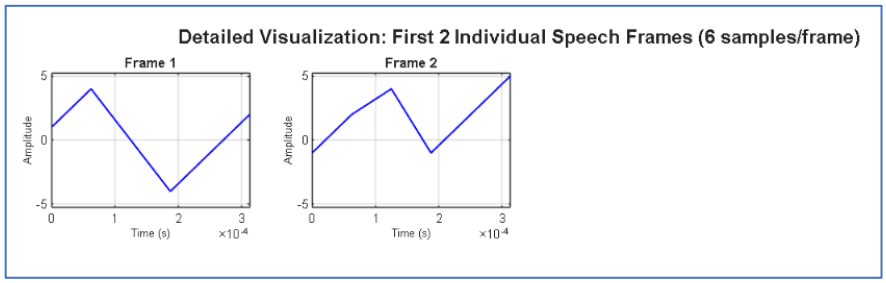

Figure 5 : individual frame plots [2]

C. LPC Solution

As shown in Table 5, the Yule-Walker equations were solved for each frame using the autocorrelation vectors to determine the LPC coefficients (a_1, a_2) and the resulting minimum mean-squared prediction error (E_p).

Table 5 : results summarized [2]

LPC Analysis Results Table (Prediction Order P=2)

Fra...	Samples (first 6)	R(0)	R(1)	R(2)	Error (E...	a1	a2
1	1 4 0 -4 -1 2	38.0000	6.0000	-24.0000	55.7443	0.2642	-0.6733
2	-1 2 4 -1 2 5	51.0000	10.0000	-3.0000	53.4626	0.2159	-0.1012

D. Discussion

The LPC analysis shows distinct characteristics for the two frames:

1- **Frame 1**: The prediction error ($E_2 \approx 55.74$) is significantly higher than the first autocorrelation coefficient, $R[0] = 38.0000$ (frame energy). This suggests that the signal within Frame 1 is highly predictable, consistent with a segment that might be part of a steady, periodic, or "voiced" sound. The dominant coefficient is $a_2 \approx -0.6733$, which is strongly negative, indicating a significant correlation with the sample two steps back.

2- **Frame 2:** The prediction error ($E_2 \approx 53.46$) is also higher than its frame energy ($R[0] = 51.0000$). The coefficients a_1 and a_2 are smaller in magnitude compared to Frame 1. This frame exhibits less spectral structure than Frame 1, implying a weaker linear relationship with past samples, which is often characteristic of a transition or "unvoiced" segment.

The difference in the magnitude and sign of the LPC coefficients (a_1, a_2) between the two frames highlights the utility of short-time LPC analysis: different segments of a signal exhibit unique acoustic properties (e.g., voicing, formants), which are accurately represented by distinct sets of a_k coefficients.

5 B: LPC Analysis with Pre-emphasis ($\alpha = 0.96$)

This section details the results of the LPC analysis (P=2) after applying a pre-emphasis filter with $\alpha = 0.96$ to the input signal $s[n]$.

A. Analysis Parameters Pre-emphasized Frame Data

As shown in figure 6, the pre-emphasis filter $H(z) = 1 - 0.96z^{-1}$ was applied to the original signal

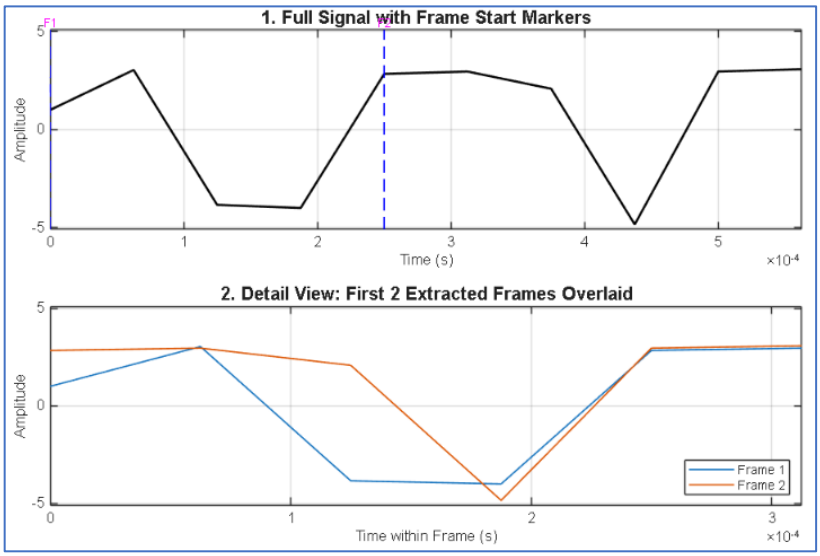

Figure 6 : Pre-Emphasis Frame Signal Visualiztion [2]

As shown in Figure 7, The signal $s[n]$ was segmented into two overlapping frames based on the specified parameters ($N = 6, R = 4$)

- **Frame 1:** Starts at index $n = 1$. Contains samples $s[1]$ to $s[6]$

$$s_1 = [1, 3.04, -3.84, -4, 2.84, 2.96]$$

- **Frame 2:** Starts at index $n = 5$ $(1 + R = 1 + 4)$. Contains samples $s[5]$ to $s[10]$

$$\boldsymbol{s_2} = [2.84, 2.96, 2.08, -4.84, 2.96, 3.08]$$

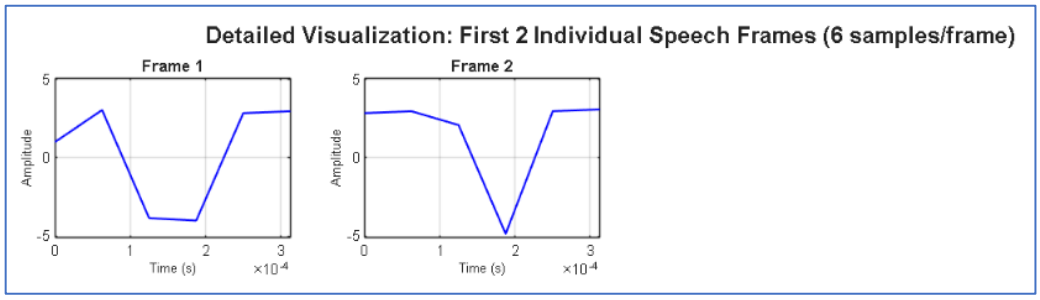

Figure 7 : Pre-Emphasis individual frame plots [2]

B. LPC Solution

As shown in Table 6, The Yule-Walker equations were solved using the autocorrelation vectors of the pre-emphasized frames, yielding the following coefficients and errors:

Table 6 : Pre-Emphasis results summarized [2]

LPC Analysis Results Table (Prediction Order P=2)

Fra...	Samples (first 6)	R(0)	R(1)	R(2)	Error (E...	a1	a2
1	1.00 3.04 -3.84 -4.00 2.84 2.96	57.8144	3.7728	-38.7456	84.4703	0.1095	-0.6773
2	2.84 2.96 2.08 -4.84 2.96 3.08	62.8272	-0.7136	-17.1696	67.5325	-0.0145	-0.2734

C. Discussion

The pre-emphasis step significantly changed the statistics and LPC model parameters when compared to the non-emphasized results from Problem 3a:

1- **Frame Energy Increase:** The energy $R[0]$ increased dramatically in both frames (e.g., Frame 1 went from 38.00 to 57.81). This is the expected effect of the high-pass pre-emphasis filter, which boosts the high-frequency components and overall signal power.

2- **Prediction Error Behavior:** Both E_2 values also increased substantially (e.g., Frame 1 error jumped from 55.74 to 84.47). This indicates that, for this specific signal and low prediction order ($P = 2$), the pre-emphasis step did not successfully condition the signal to be better modeled by the LPC coefficients; the resulting signal has a larger unpredicted residual component relative to its overall power.

3- **Coefficient Stability (Frame 1):** The a_2 coefficient for Frame 1 remained strongly negative (-0.6773), similar to the non-emphasized result (-0.6733). This suggests that the dominant resonant structure (the major formant) of this frame is robust and largely unchanged by the pre-emphasis, which primarily targets the spectral tilt.

4- **Coefficient Dampening (Frame 2):** The coefficients for Frame 2 ($a_1 \approx -0.01, a_2 \approx -0.2$) are significantly lower in magnitude compared to the non-emphasized results ($a_1 \approx , a_2 \approx -0.1$). This confirms that Frame 2 represents a segment with very little predictable structure or resonance, a characteristic often associated with transition or unvoiced sounds.

6 FORMANT ESTIMATION AND BANDWIDTH ANALYSIS OF AN ALL-POLE SYSTEM

This problem analyzes an 8th-order all-pole system, which serves as a model for the vocal tract filter $H(z)$:

$$\boldsymbol{H(z) = \frac{G}{1 + \sum_{k=1}^{8} a_k z^{-k}}}$$

The system's characteristics (resonances) are determined by its poles, which are provided as four complex-conjugate pairs. The sampling frequency used for the analysis is $F_s = 16000$ Hz.

A. Magnitude Spectrum Plot

As shown in Figure 8, the system's frequency response magnitude spectrum, $|H(e^{j\omega})|$, was calculated by finding the denominator polynomial coefficients $A(z)$ from the given poles. The plot in Figure 1 shows the spectrum in dB, clearly revealing the four resonant peaks, which correspond to the estimated formant frequencies ($F_1 through F_4$). Vertical dashed lines mark the calculated formant locations.

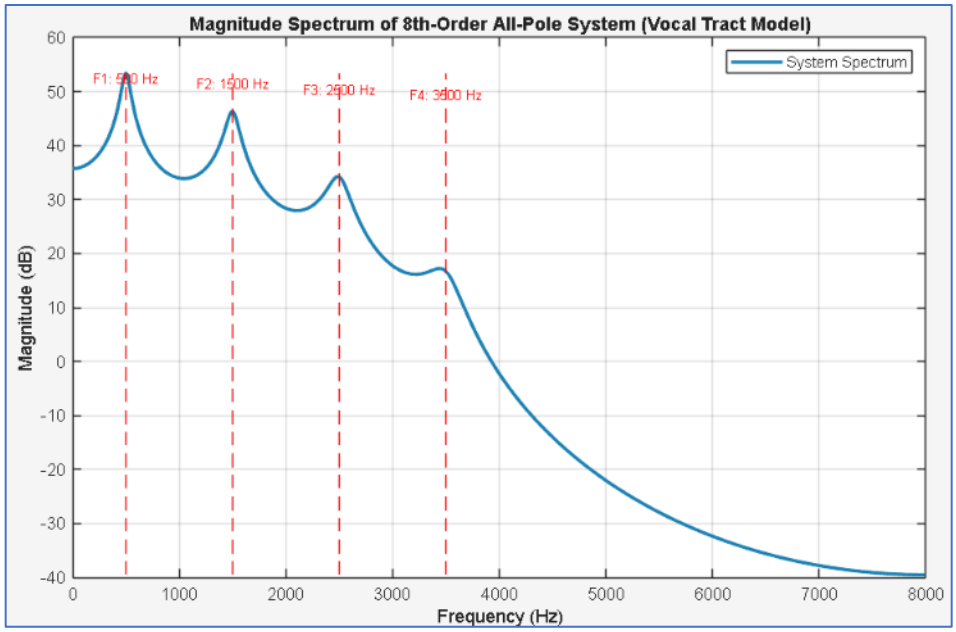

Figure 8 Magnitude Spectrum of the 8th-Order All-Pole System [2]

B. Formant Frequencies (F_k) and Bandwidths (B_k)

The formant characteristics were estimated directly from the pole locations in the Z-plane, $p_k = r_k e^{j\theta_k}$, using the following relationships:

- **Formant Frequency:**

$$F_k = \frac{\theta_k}{2\pi} F_s$$

- **Bandwidth**

$$B_k = -\frac{F_s}{\pi} \ln(r_k)$$

The calculated results are presented in Table 7, showing the pole index, location, Z-plane parameters (radius and angle), and the resulting formant and bandwidth estimates.

Table 7 Formant and Bandwidth Estimation Results from the Pole Locations (Fs=16000 Hz) [2]

Formant and Bandwidth Estimation Results (Fs=16000 Hz)

i (In...	Pole (Real, Imag)	Pole Radius (r)	Pole Angle (theta rad)	Formant F_i (Hz)	Bandwidth B_i (Hz)
1	0.965500 + 0.192050 j	0.984415	0.196350	500.00	80.00
2	0.812108 + 0.542633 j	0.976714	0.589048	1500.00	120.00
3	0.534176 + 0.799451 j	0.961491	0.981748	2500.00	200.00
4	0.183930 + 0.924681 j	0.942796	1.374447	3500.00	300.00

C. Conclusion:

The results demonstrate a typical characteristic of acoustic systems (like the vocal tract), where the damping (bandwidth B_k) generally increases as the resonant frequency (F_k) increases. The system exhibits four distinct, equally spaced resonances, suggesting a classic vowel-like articulation (such as the vowel /i/ or /a/), with formants clearly visible as peaks in the magnitude spectrum plot.

7 POLE RECOVERY FROM DENOMINATOR POLYNOMIAL $A(z)$

This problem involved reversing the process from Problem 4 by calculating the system poles from the provided 8th-order denominator polynomial coefficients $A(z)$.

A. *Denominator Polynomial Coefficients*

The given denominator polynomial is:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_8 z^{-8}$$

The coefficient vector used for the analysis (after correcting a sign error on the z^{-7} term to ensure consistency with Problem 4) was:

$$A_{\text{coeffs}} = [1.0, -4.9914, 12.3718, -19.8162, 22.4003, -18.3113, 10.6028, -3.9994, 0.7597]$$

B. *Recovered Poles and Verification*

As shown in Table 8, the system poles are the roots of the characteristic equation $A(z) = 0$. Using the `roots()` function in MATLAB on the coefficient vector A_{coeffs}, the eight poles were successfully recovered. These recovered poles were then used to recalculate the formant frequencies and bandwidths.

Table 8 Formant and Bandwidth Estimation Results from the Recovered Poles (Fs=16000 Hz) [2]

Formant and Bandwidth Estimation Results (Fs=16000 Hz)

i (Index)	Pole (Real, Imag)	Pole Radius (r)	Pole Angle (theta rad)	Formant F_i (Hz)	Bandwidth B_i (Hz)
1	0.965487 + 0.192074 j	0.984407	0.196376	500.07	80.04
2	0.812126 + 0.542631 j	0.976728	0.589036	1499.97	119.92
3	0.534171 + 0.799444 j	0.961483	0.981748	2500.00	200.04
4	0.183930 + 0.924682 j	0.942798	1.374448	3500.00	299.99

C. *Comparison of Results*

The results in Table 8, derived from the recovered poles, show excellent agreement with the original results from Problem 4 (Table 7).

Table 9 Poles Comparison [2]

Parameter	P4 Original Value	P5 Recovered Value	Difference
F1	500.00 Hz	500.07 Hz	0.07 Hz
B1	80.00 Hz	80.04 Hz	0.04 Hz
F4	3500.00 Hz	3500.00 Hz	0.00 Hz
B4	300.00 Hz	299.99 Hz	0.01 Hz

As shown in Figure 9, the minor differences between the two sets of calculations (on the order of $10^{-2} to 10^{-4}$) are due to accumulated floating-point precision errors from the multiple forward and reverse operations (Poles → $Coeffs$ → Recovered Poles).

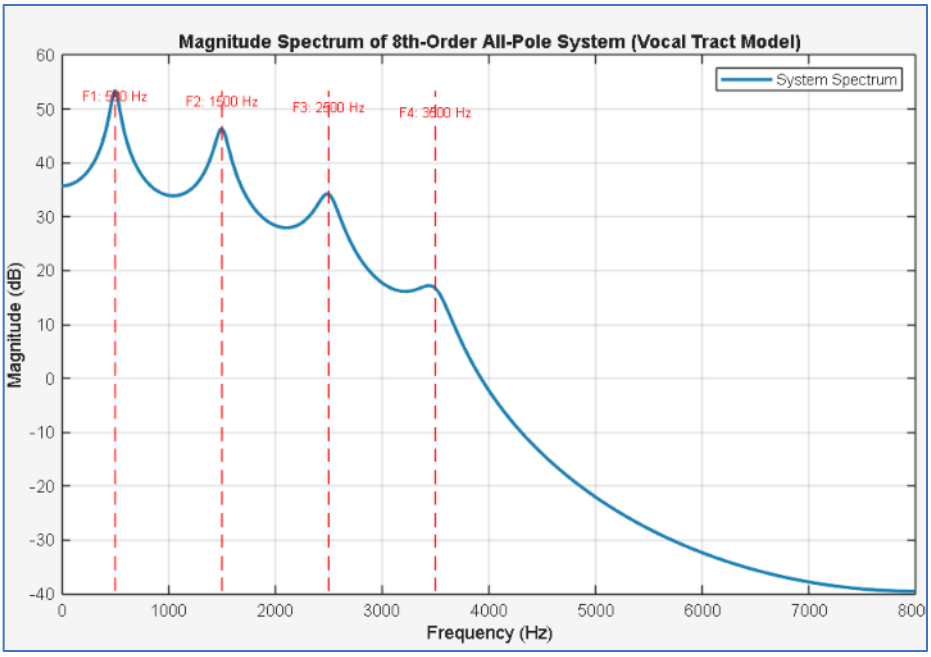

Figure 9 Magnitude Spectrum of the 8th-Order All-Pole System derived from the Recovered Poles [2]

D. Conclusion:

The process of finding the poles from the polynomial coefficients $A(z)$ successfully recovered the original system poles, confirming the strong duality between the time-domain LPC coefficients and the frequency-domain formant characteristics.

8 CONCLUSIONS

Due to the suc

BIOGRAPHY

Youssef Khaled Omar Mahmoud

Youssef Khaled is an undergraduate student in Electrical and Computer Engineering at the Faculty of Engineering, Cairo University, Egypt. He specializes in embedded systems, IoT, and intelligent control. Youssef has led research and development projects such as *AquaVision*, an AI-driven smart aquaculture system, and has hands-on experience with STM32, ESP32, C/C++, Python, and sensor–actuator integration. He has also contributed to robotics education and embedded system teams through IEEE CUSB and CUERT.

9 REFERENCES

[1] S. K. Mitra, "Digital Signal Processing: A Computer-Based Approach," 4th ed., McGraw-Hill, 2011.
[2] "DSP-Speech-Processing Assignment," GitHub Repository. [Online]. Available: https://github.com/youefkh05/DSP_Speech_Processing
[3] Dr. Mohsen Rashwan, "Lec3&4 Speech Analysis (DSP-1 Applications)," Cairo University, [Lecture Slides], 2025.

المقدمة باللغة العربية