Advanced Topics in Speech and Language Processing

Moschogiannis Paschalis   Dimitriou Vasileios

Department of Electrical and Computer Engineering

UNIVERSITY OF THESSALY

Information

Course: MDE671 Advanced Topics in Speech and Language Processing

Computational Assignment

Professor: Potamianos Gerasimos

Academic Year: 2023-24

Submission Date: February 25, 2024

Recording

For recording our name using the **SoX** tool with a sampling frequency of 16kHz and a sample depth of 16 bits, for a duration of 5 seconds:

```
sox -t alsa default -c 1 -r 16000 -b 16 -e signed-integer -t wav
recorded_audio.wav remix 1 trim 0 5
```

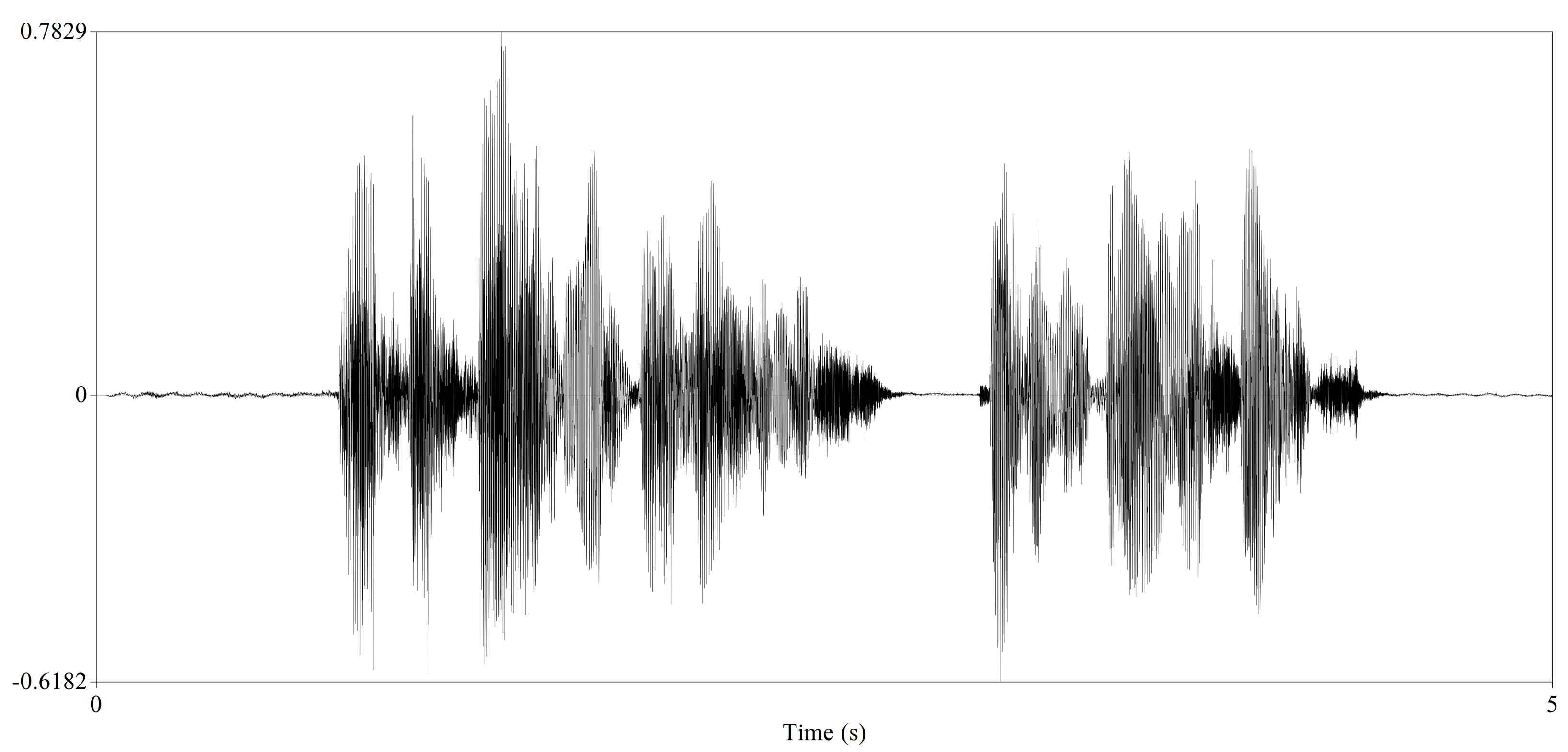


Figure 1. Waveform representation of the phrase "Moschogiannis Paschalis and Dimitriou Vasilis" in Praat [1].

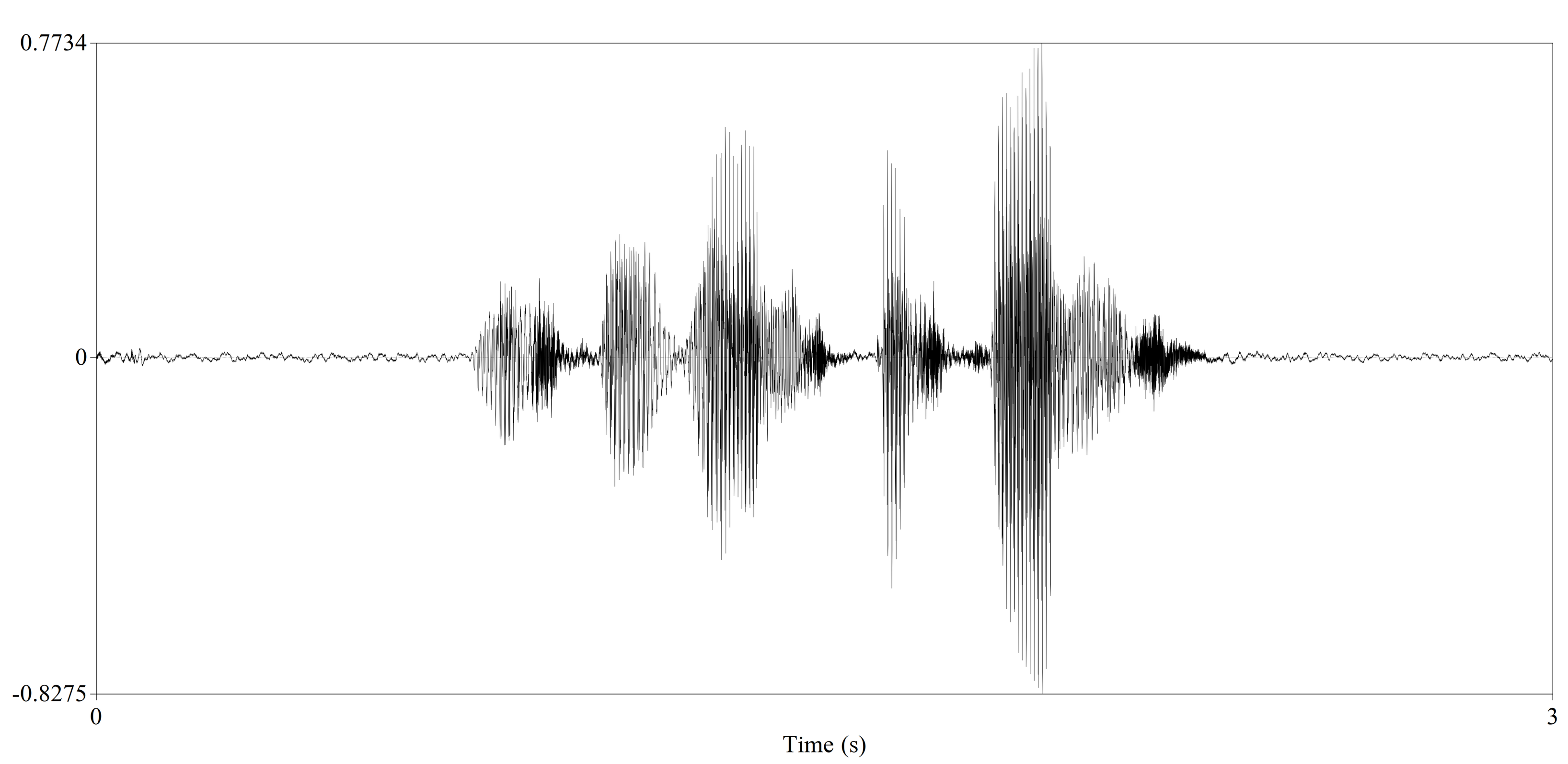


Figure 2. Waveform of one name "Moschogiannis Paschalis" in Praat [1].

Spectrogram

Creating spectrograms using Fourier transformation on audio signals allows the analysis of frequency content over time, providing a visual representation of frequency changes throughout the recording. The use of different Hamming window sizes affects the analysis, with smaller windows improving temporal resolution and larger windows improving frequency resolution.

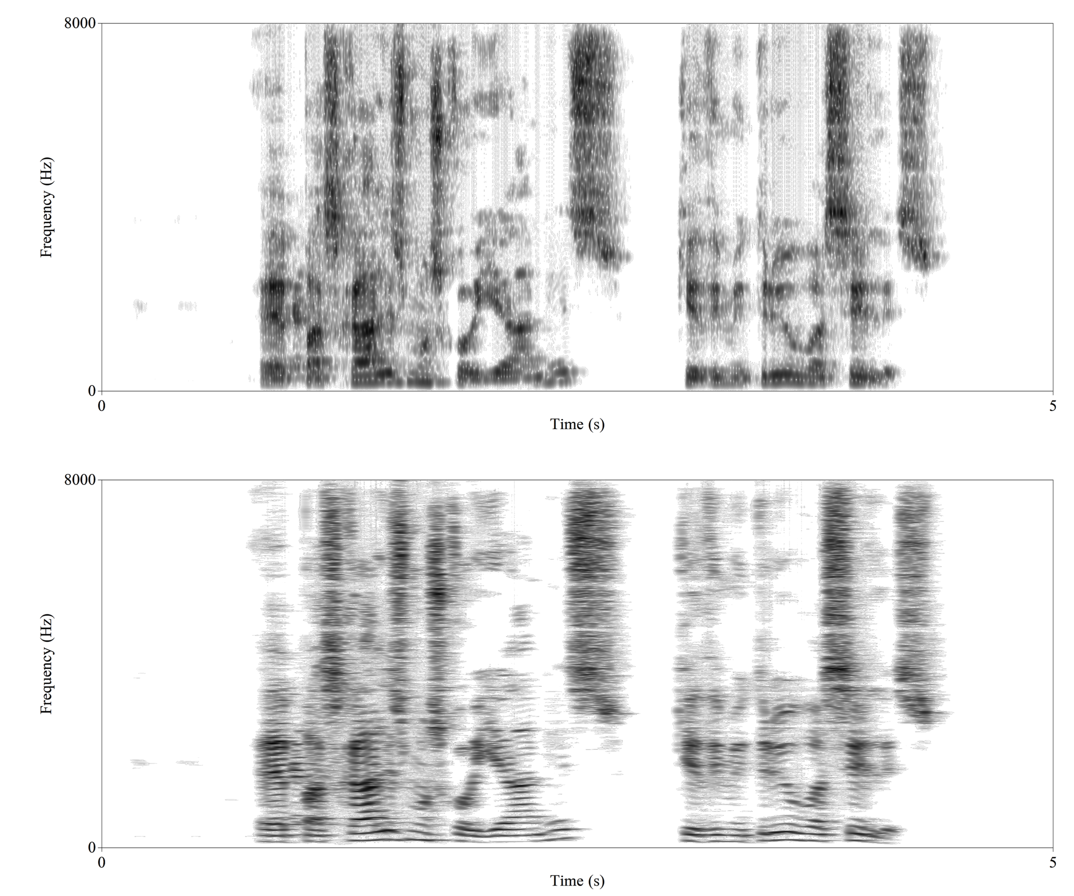


Figure 3. Spectrograms for 10ms and 100ms respectively in Praat [1] for the phrase "Moschogiannis Paschalis and Dimitriou Vasilis".

In the MATLAB environment in Figure **4**, the same spectrograms are depicted with the X-axis representing time and the Y-axis representing frequencies. The intensity of each frequency at each time is displayed using different shades or colors.

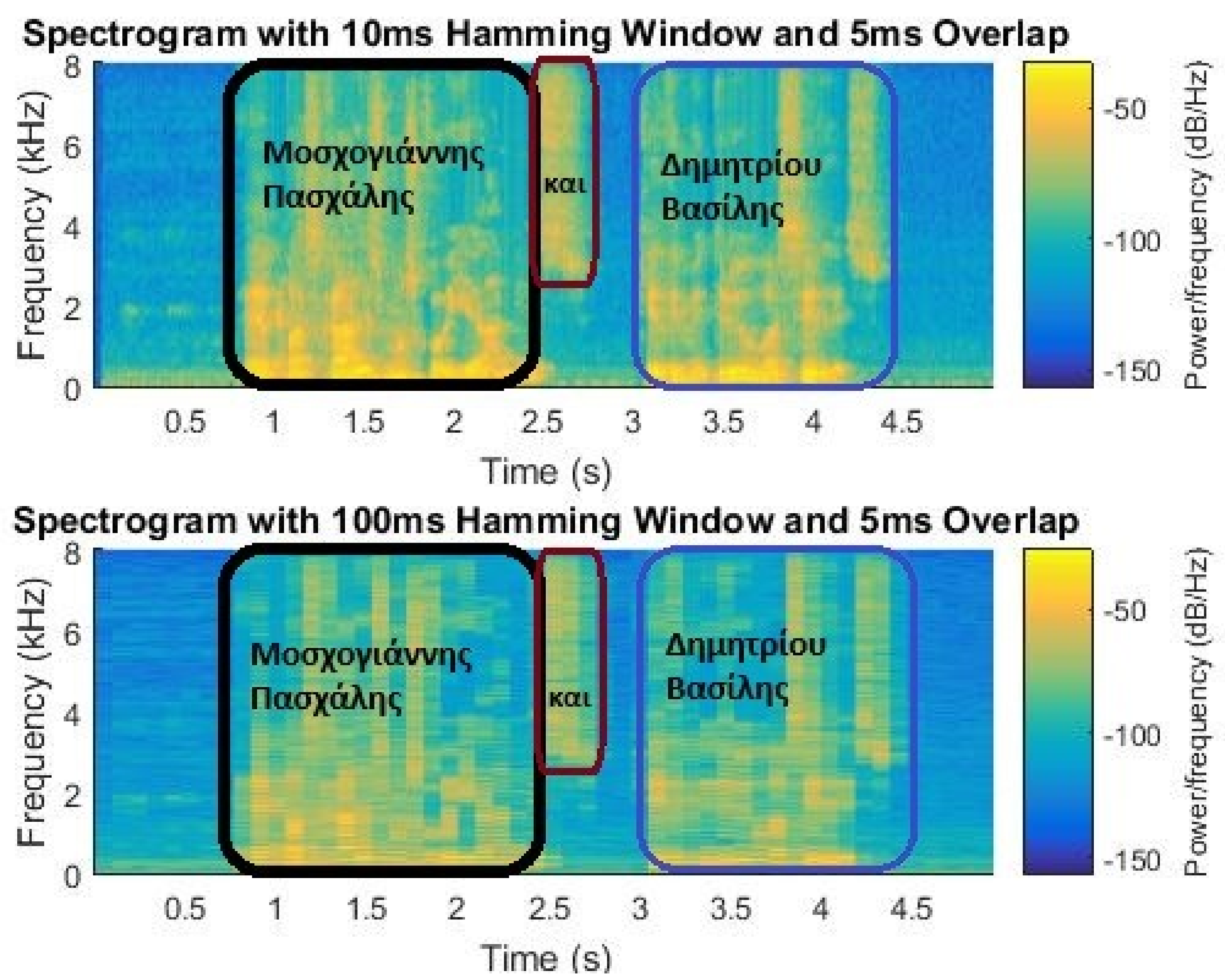


Figure 4. Indication of the content (words) of the spectrograms.

Separation of Voiced, Unvoiced, and Silent Sounds

In the case of the audio signal of one name, an algorithm was developed to separate the voiced, unvoiced, and silent segments of an audio signal using energy analysis and zero-crossing rate (ZCR), improving accuracy in speech processing and analysis.

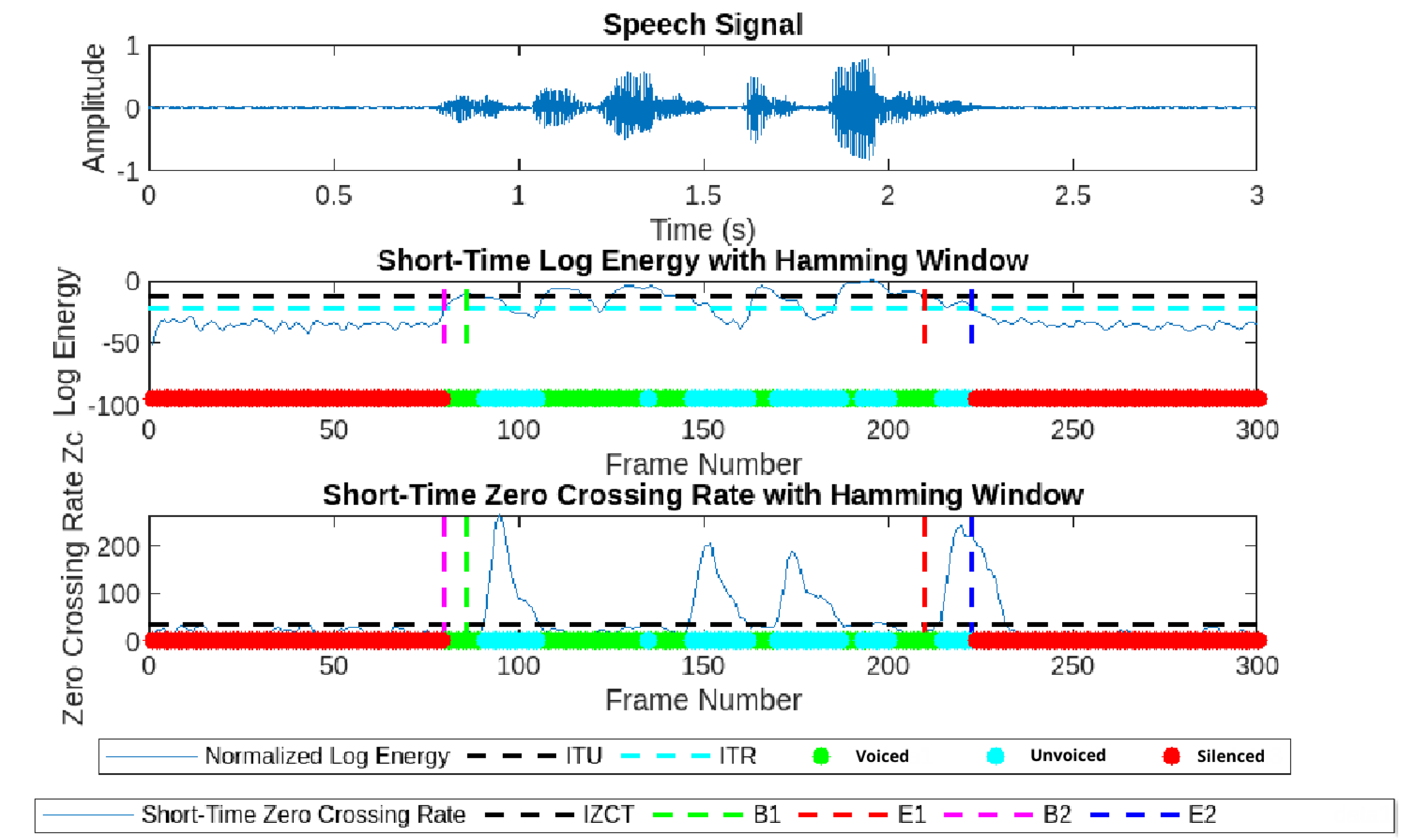
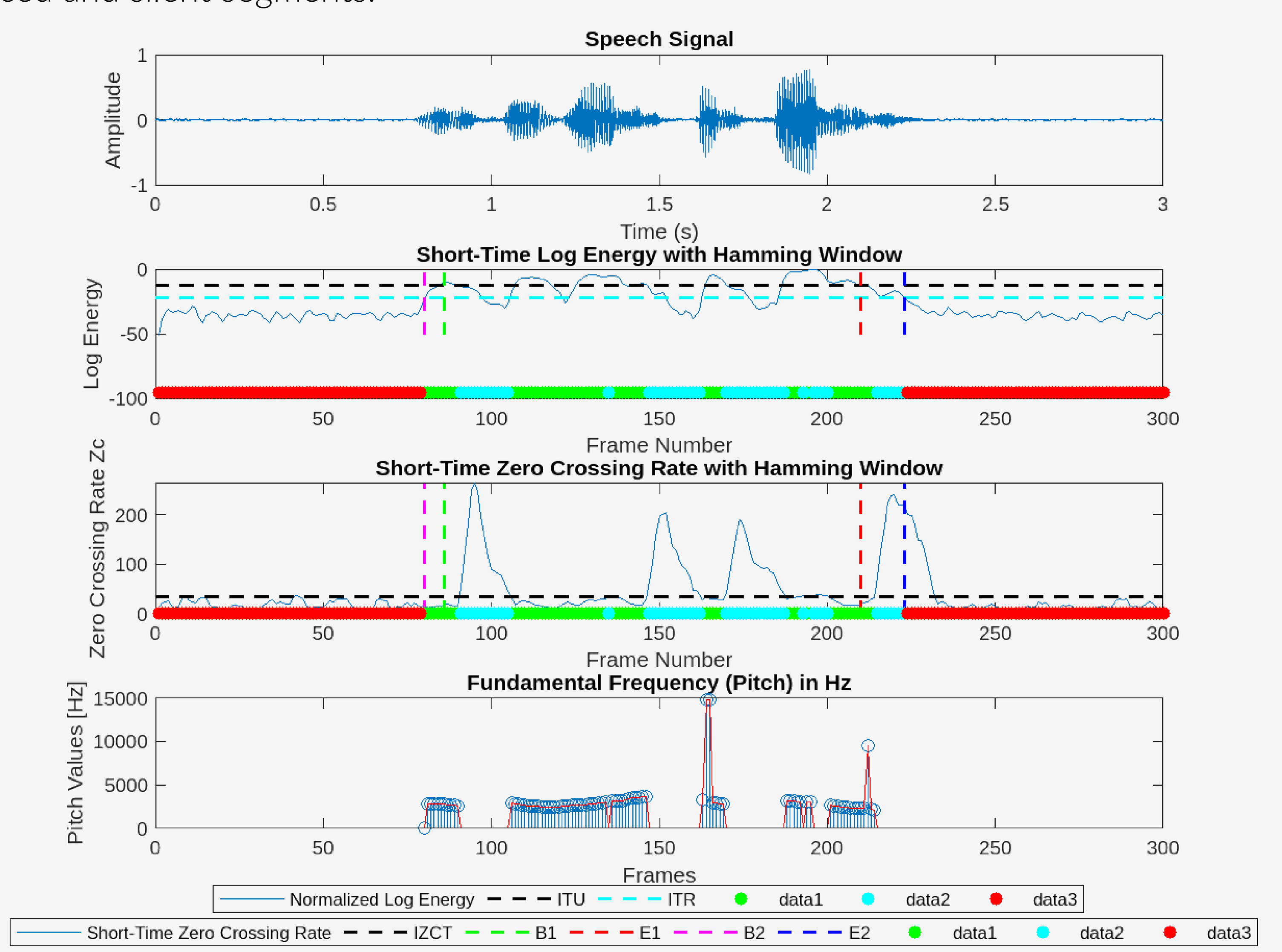


Figure 5. Analysis of energy and zero-crossing rate (ZCR).

Pitch Analysis using Autocorrelation and Hamming

A method was implemented to find the fundamental frequency of voiced sounds in the recorded signal using autocorrelation and applying a Hamming window. The results are presented visually, distinguishing fundamental frequencies in voiced segments and considering the value as zero in unvoiced and silent segments.



pmoschogiannis@uth.gr | vasildimitriou@uth.gr

UNIVERSITY OF THESSALY

Advanced Topics in Speech and Language Processing

Moschogiannis Paschalis Dimitriou Vasileios

Department of Electrical and Computer Engineering

Analysis of Sound Signals with LPC Model

Using the Linear Predictive Coding (LPC) model, sound signals of 30 msec duration are analyzed into voiced and unvoiced segments, estimating LPC characteristics and applying filtering to a voiced and an unvoiced frame.

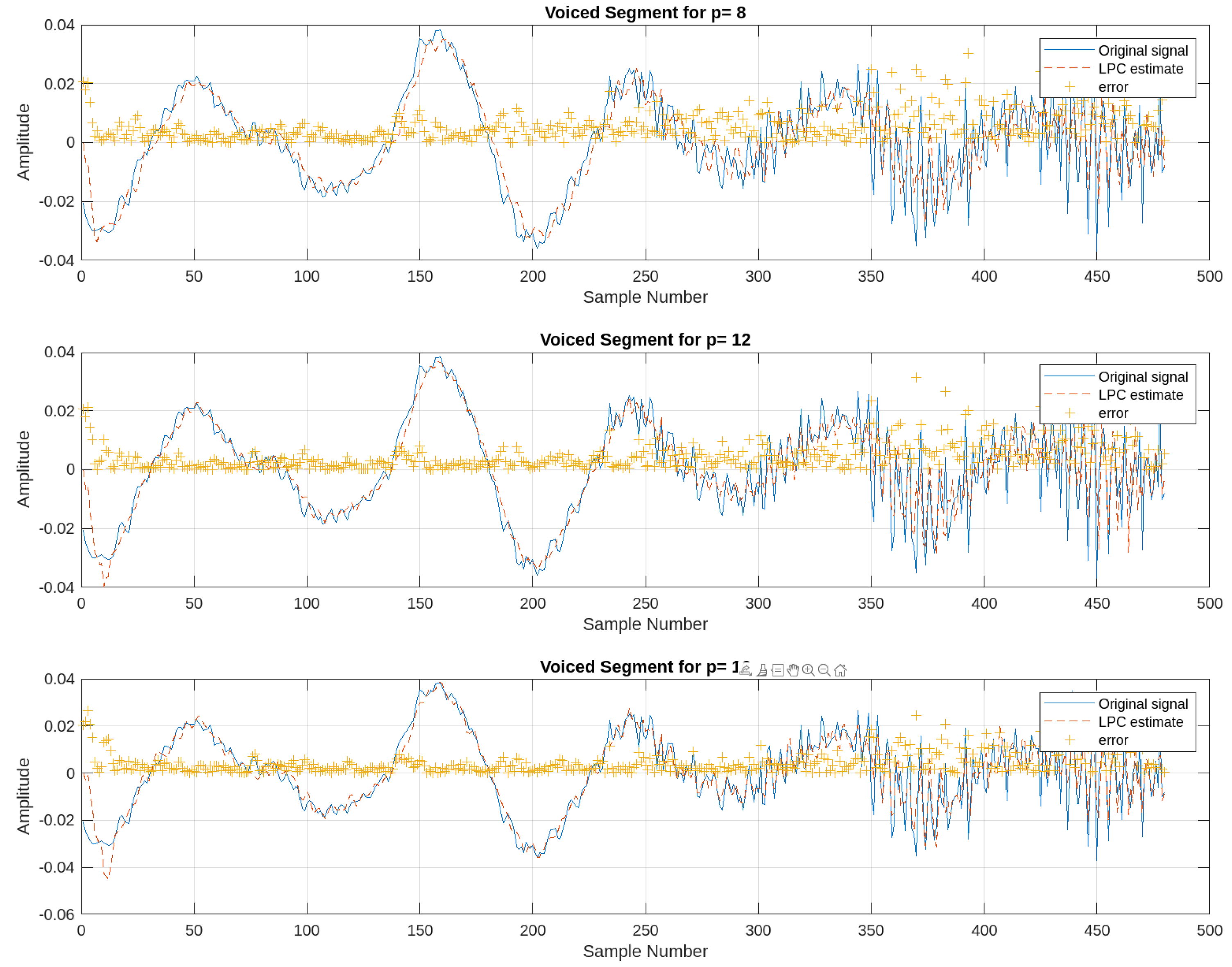


Figure 6. Example of a voiced segment.

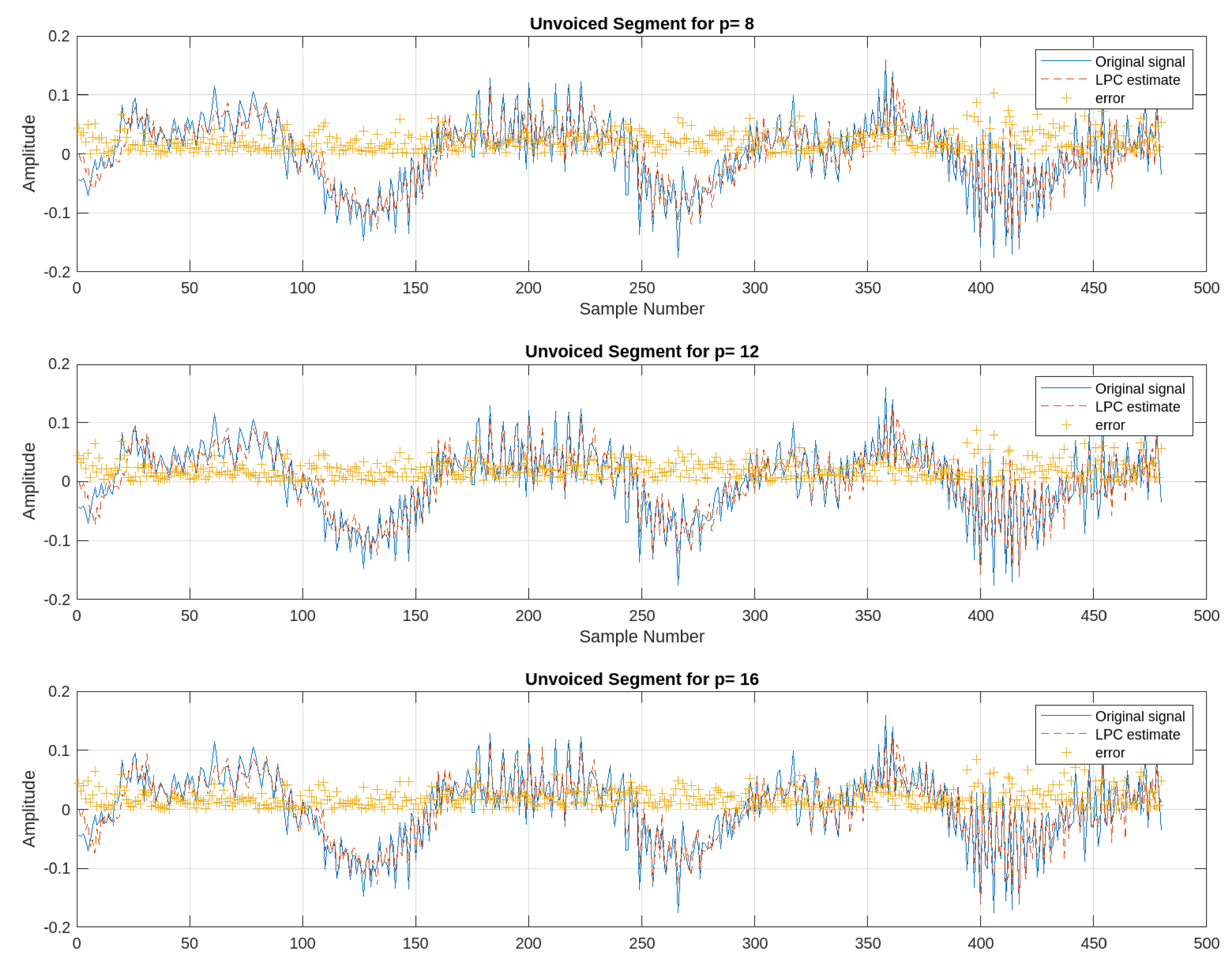


Figure 7. Example of an unvoiced segment.

Synthetic Voice Generation with Pitch Excitation Model

Create a synthetic sound of 3 seconds with a sampling frequency of 10 kHz, including six vowels each of 0.5 sec duration, using a pitch period of 8 msec and excitation from the equation $p[n] = \sum_{k=0}^{\infty} \beta^k \delta[n - kNp]$. Design the $p[n]$, the spectrum $P(e^{j\omega})$, and the pole-zero plot of $P(z)$. $P[z] = \sum_{k=0}^{\infty} \beta^k z^{kNp} = \frac{1}{1-bz^{-Np}}$

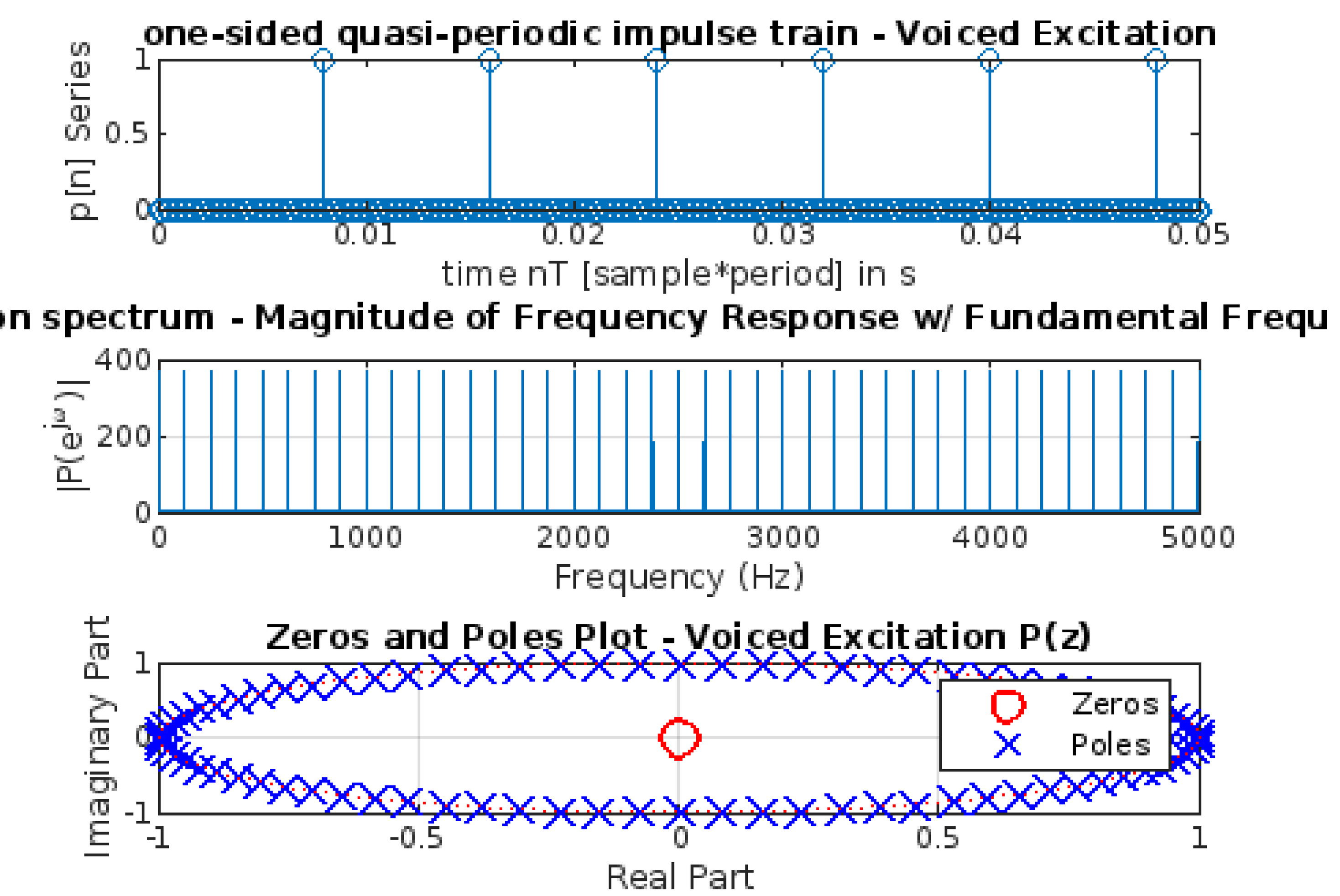


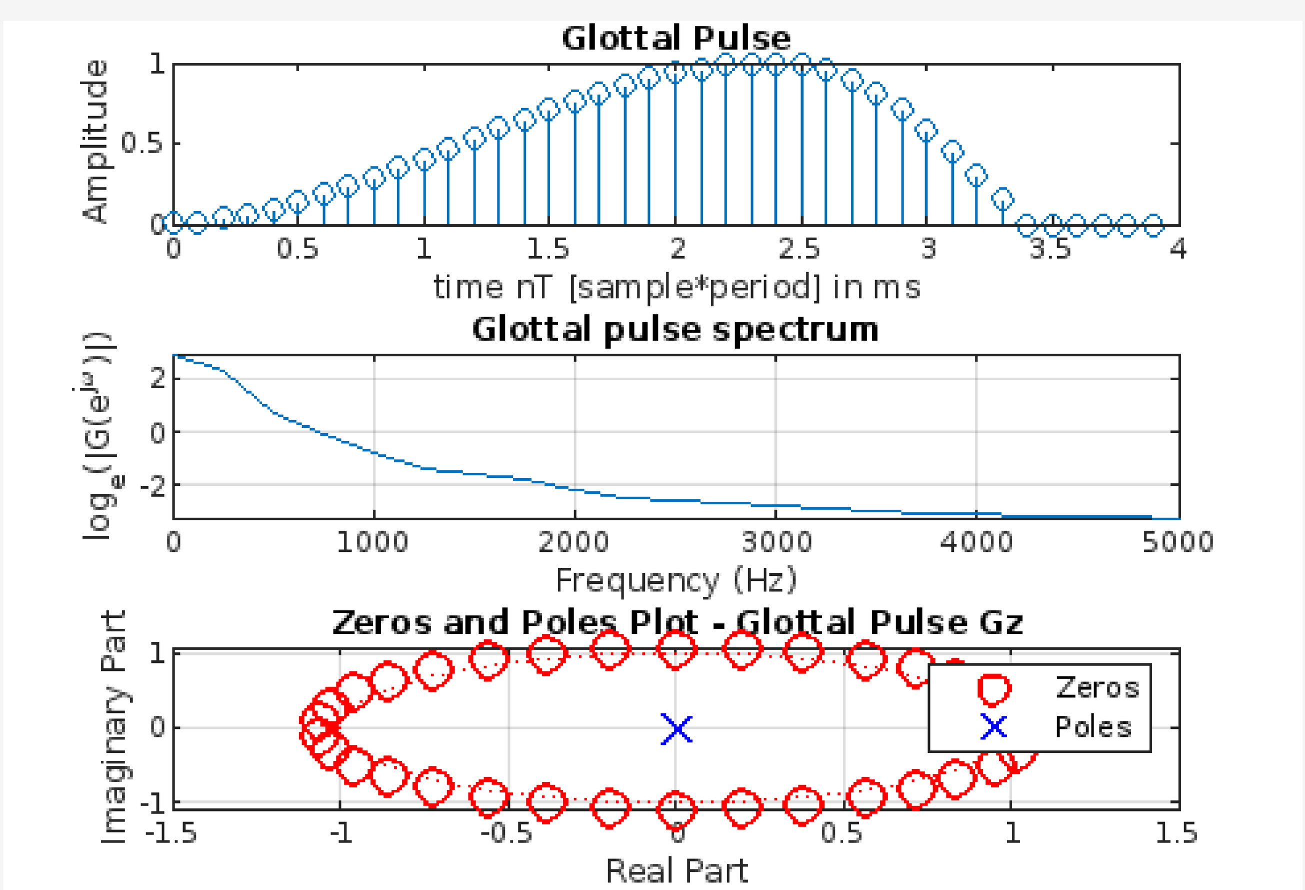
Figure 8. Plot of $p[n]$.

Glottal Pulse Analysis: Time Series, Spectrum, and Pole-Zero Plot

Design the model of the glottal pulse $g[n]$ based on the equation

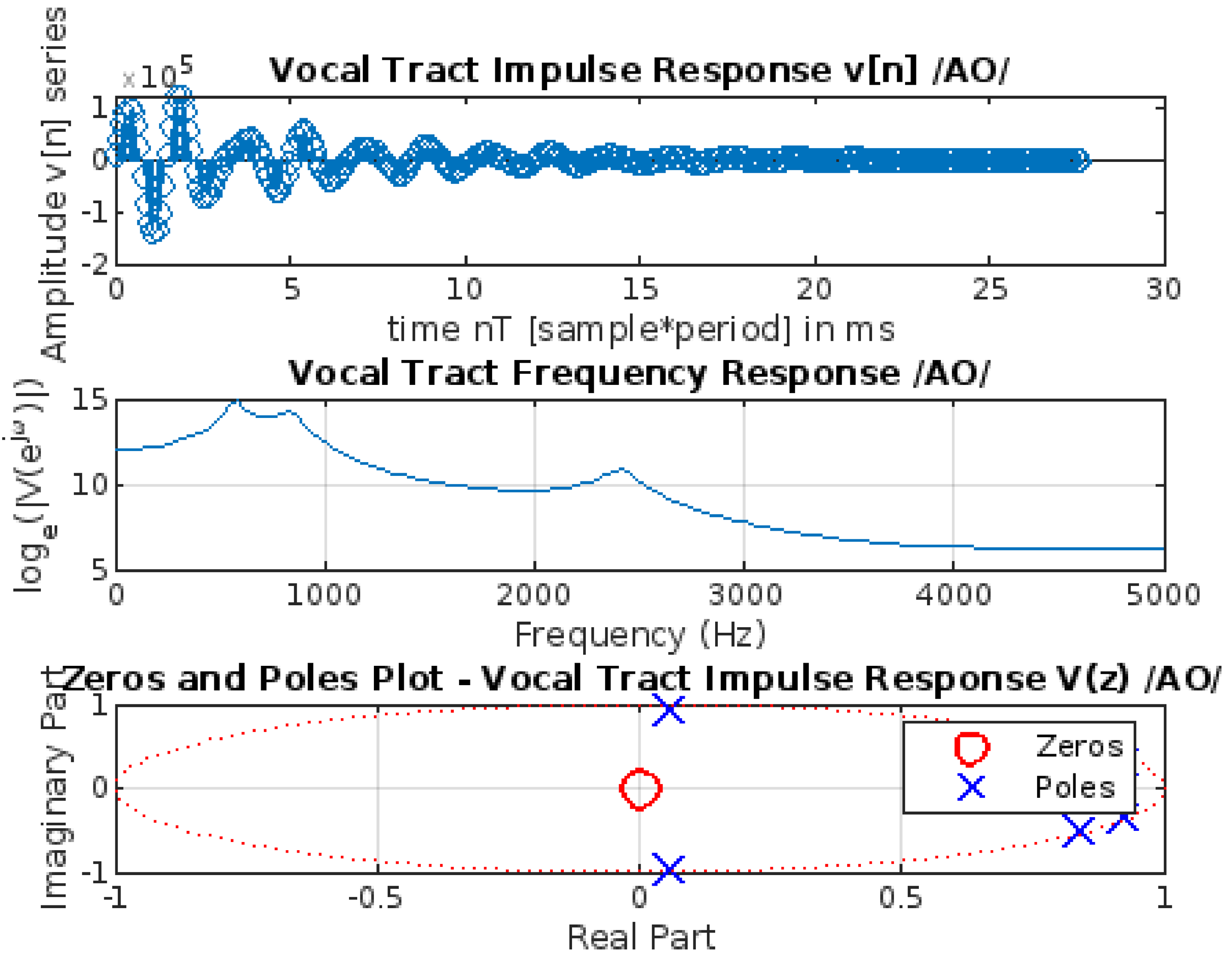
$$g[n] = \begin{cases} 0.5(1\cos(\pi(n+1)/25)), & \text{for } 0n24 \\ \cos(0.5(n24)/10), & \text{for } 25n33 \\ 0, & \text{elsewhere} \end{cases}$$

, the spectrum $G(e^{j\omega})$, and the pole-zero plot of $G(z)$.



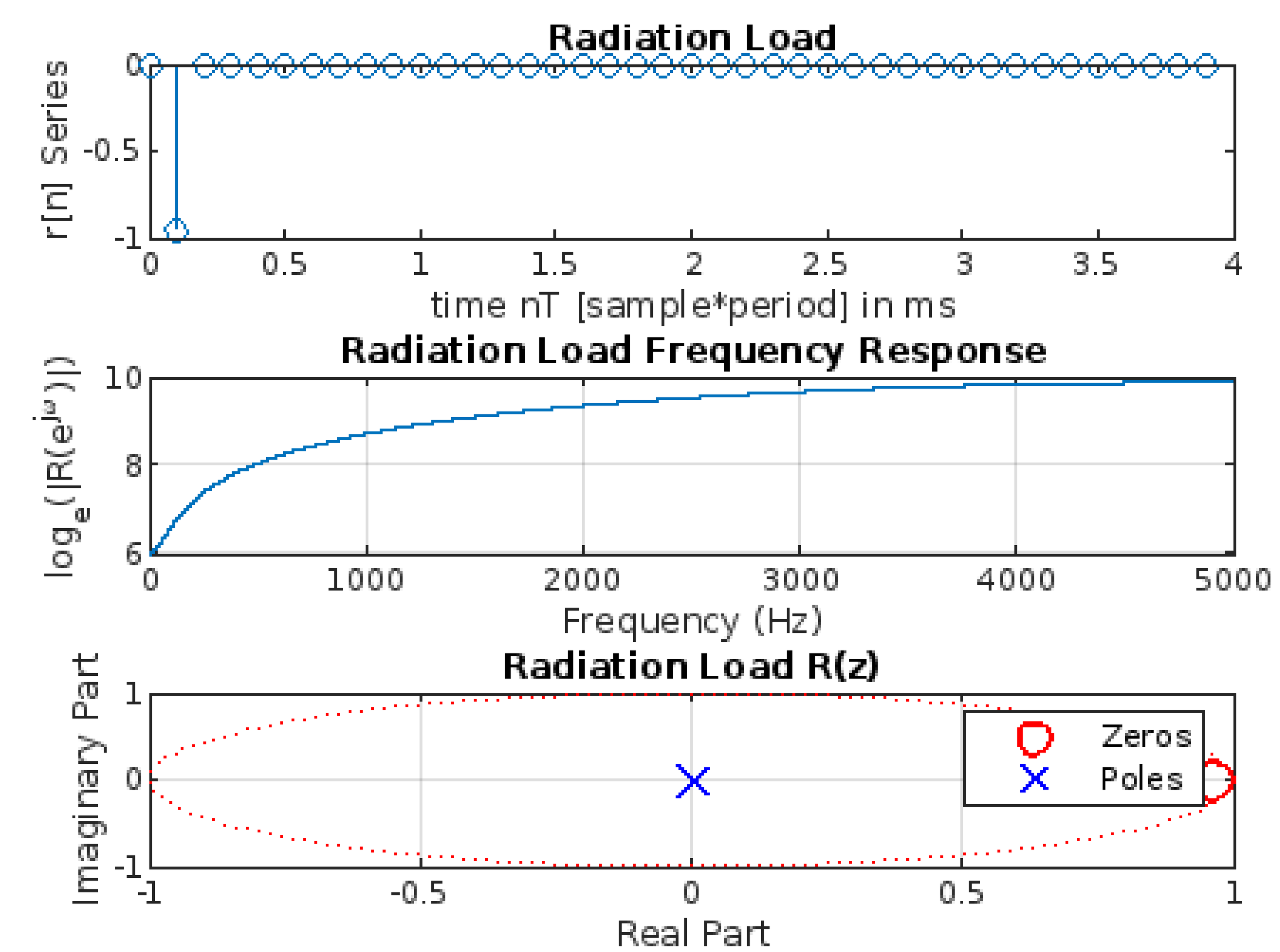
Analysis of Vocal Tract for /AO/: Spectrum and Pole-Zero Plot

Analyze the vocal tract system as a filter $V(z)$, using three formants for the vowel /AO/ with frequencies F1=570Hz, F2=840Hz, F3=2410Hz and bandwidths 60Hz, 100Hz, 120Hz respectively. Design the spectrum $V(e^{j\omega})$ and the pole-zero plot of $V(z)$.



Radiation Load Analysis: Time, Spectrum, and Pole-Zero Plot

Design $r[n]$, the spectrum $R(e^{j\omega})$, and the pole-zero plot of $R(z)$ for the radiation load $r[n] = \delta[n] - 0.96\delta[n-1]$.



pmoschogiannis@uth.gr | vasildimitriou@uth.gr

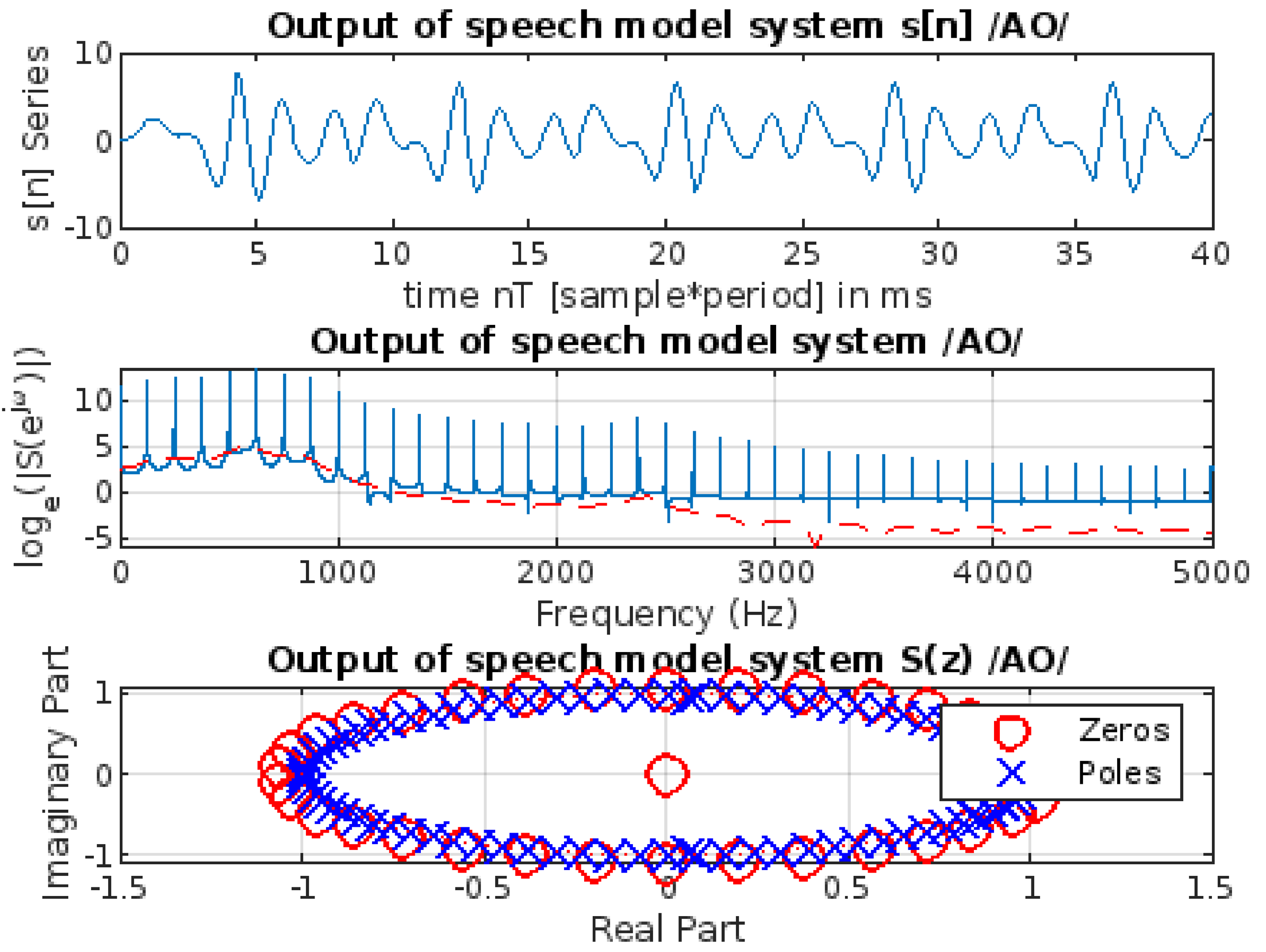
Advanced Topics in Speech and Language Processing

Moschogiannis Paschalis Dimitriou Vasileios

Department of Electrical and Computer Engineering

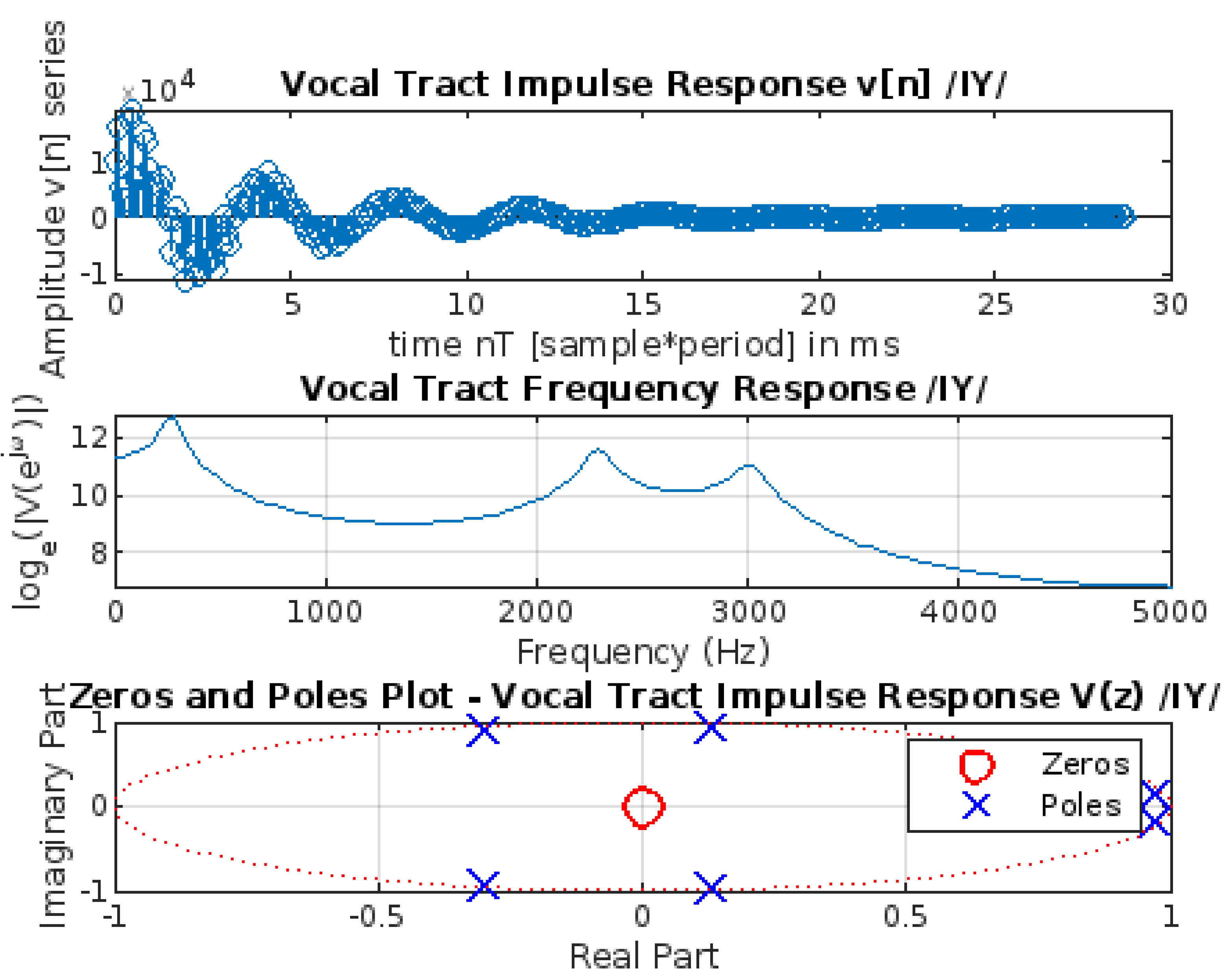
Spectral Analysis of Voice Signal

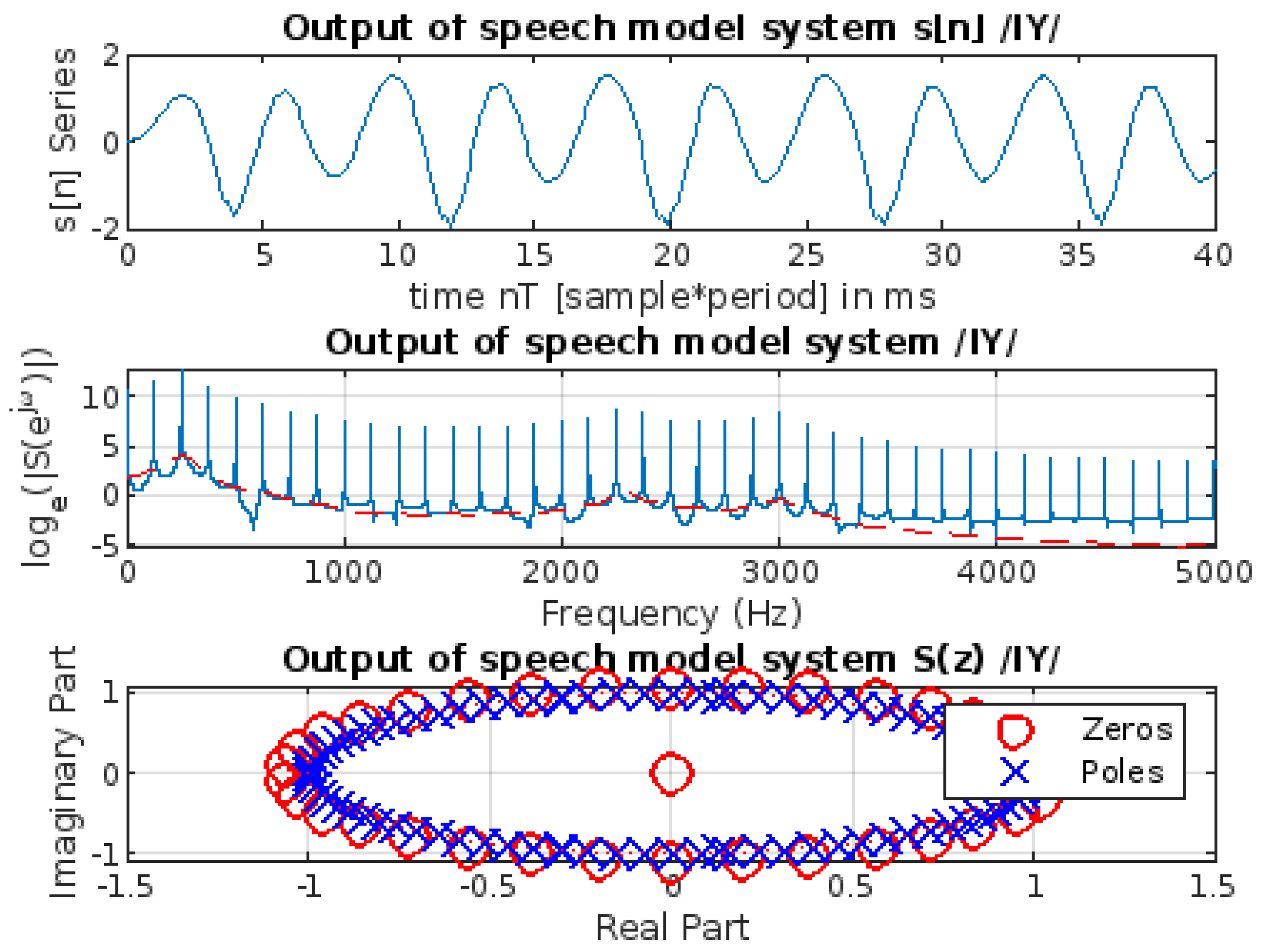
Design the spectrum $S(e^{j\omega})$ and the pole-zero plot of $S(z)$ for the total voice signal $s[n]$, obtained from the convolution of $p[n]$, $g[n]$, $v[n]$, and $r[n]$ with gain $A = 5000$.

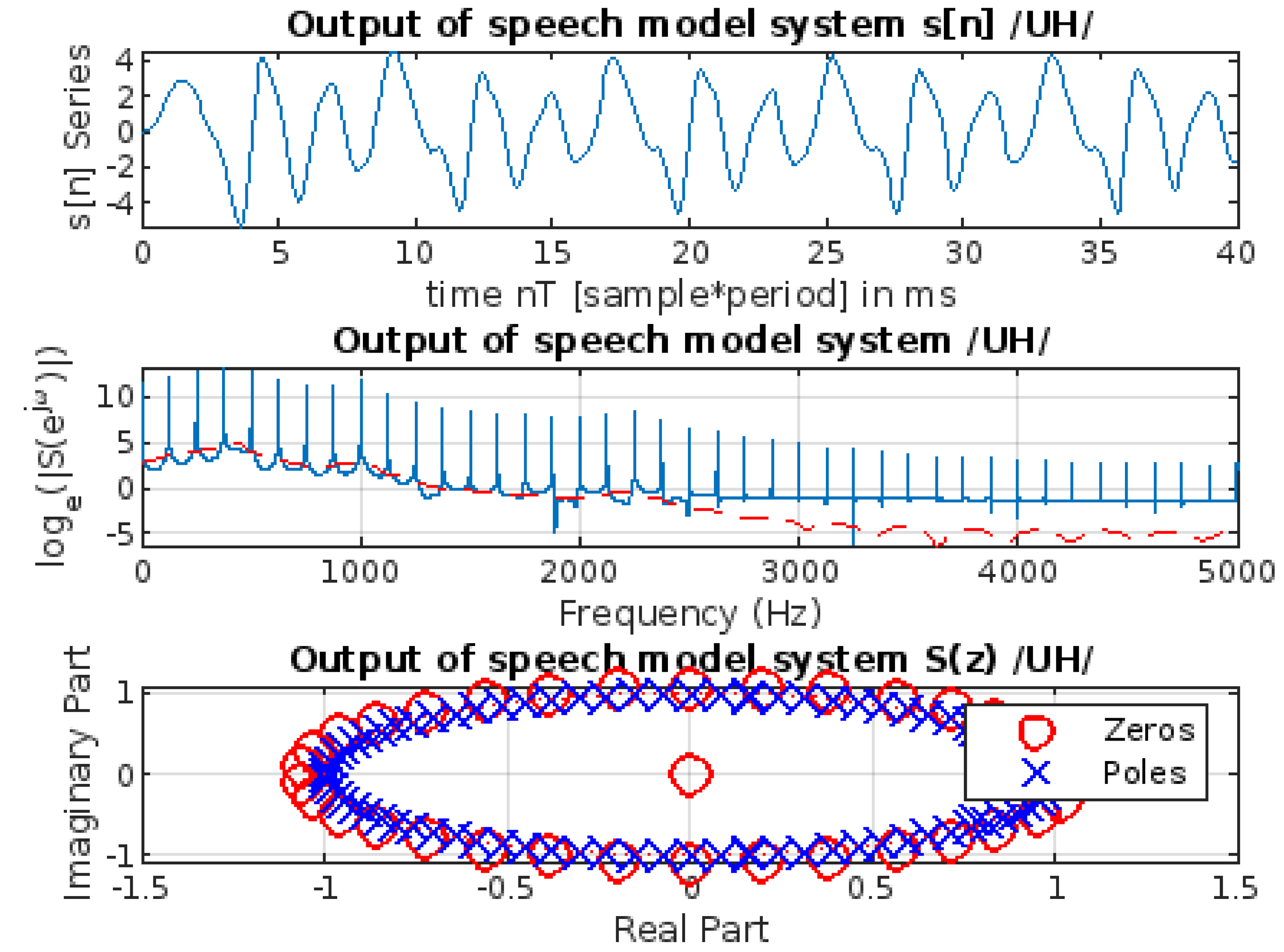


Synthesis and Analysis of Vowels: Spectra and Pole-Zero Plots

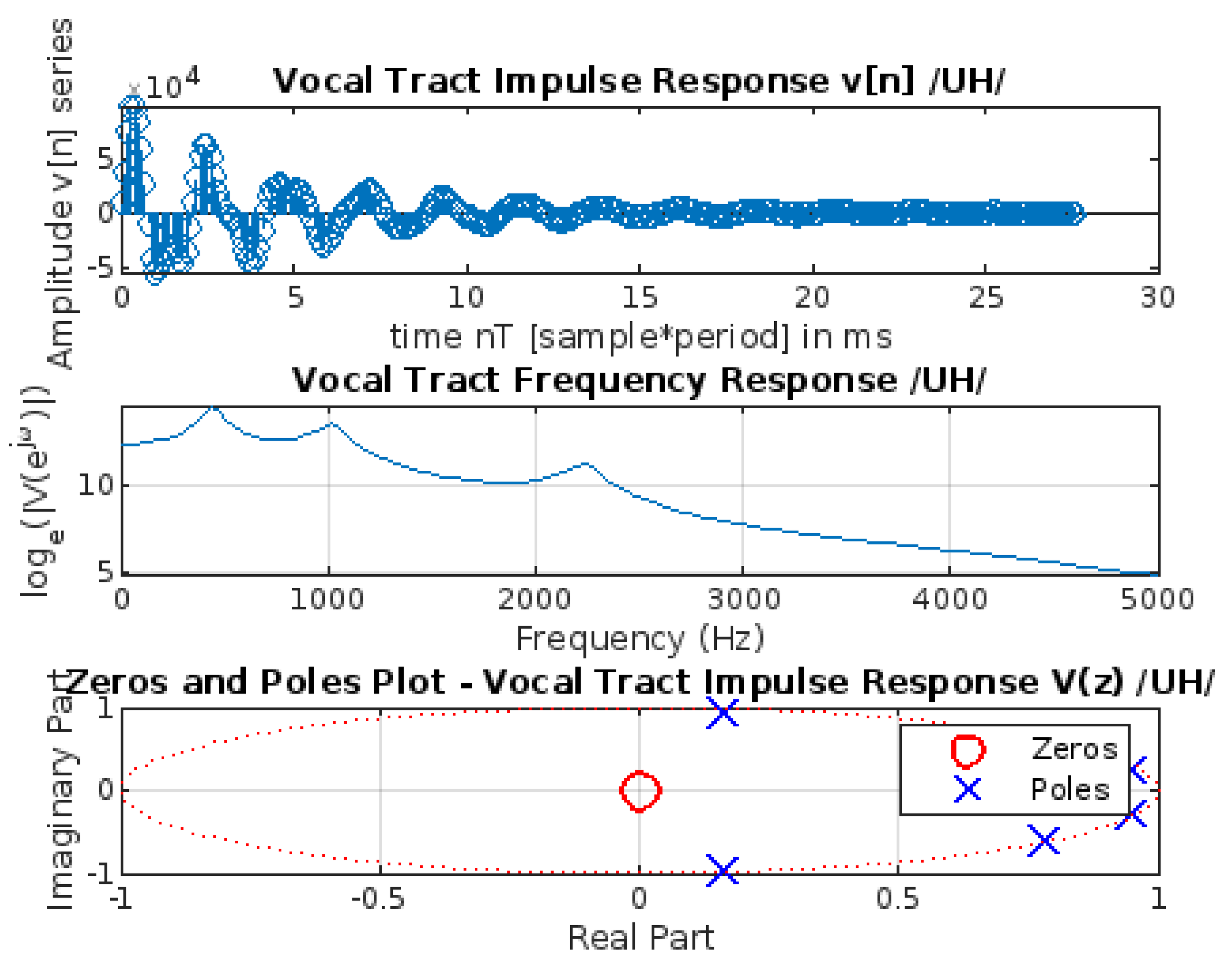
Synthesize 0.5-second voice signals for the vowels /IY/, /UH/, /EH/, /AH/, /IH/, adjusting the frequencies F1, F2, F3 accordingly. Design the spectra $V(e^{j\omega})$ and $S(e^{j\omega})$ and the pole-zero plots $V(z)$ and $S(z)$ for each vowel.



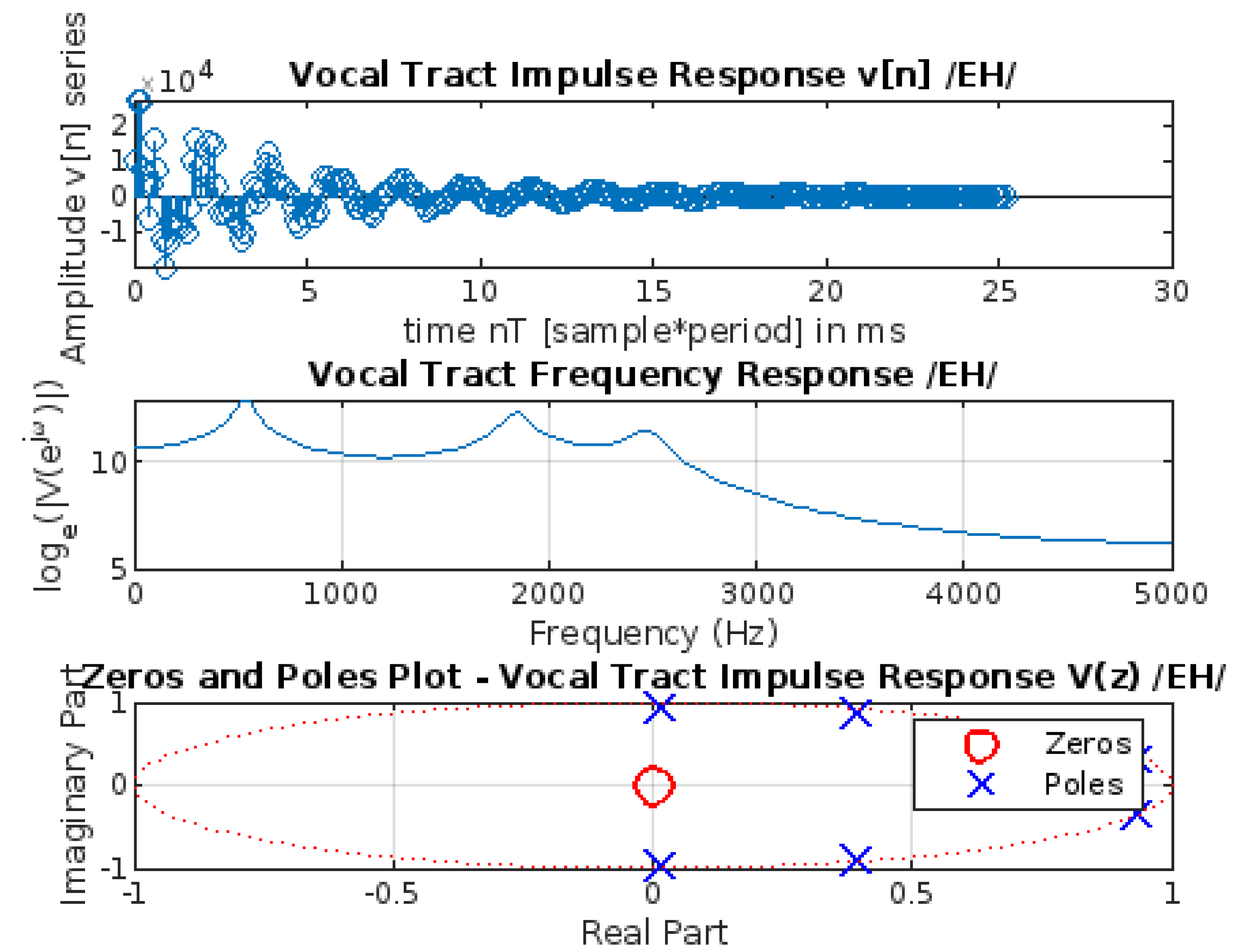




Vowel /UH/



Vowel /EH/



pmoschogiannis@uth.gr | vasildimitriou@uth.gr

Advanced Topics in Speech and Language Processing

Moschogiannis Paschalis Dimitriou Vasileios

Department of Electrical and Computer Engineering

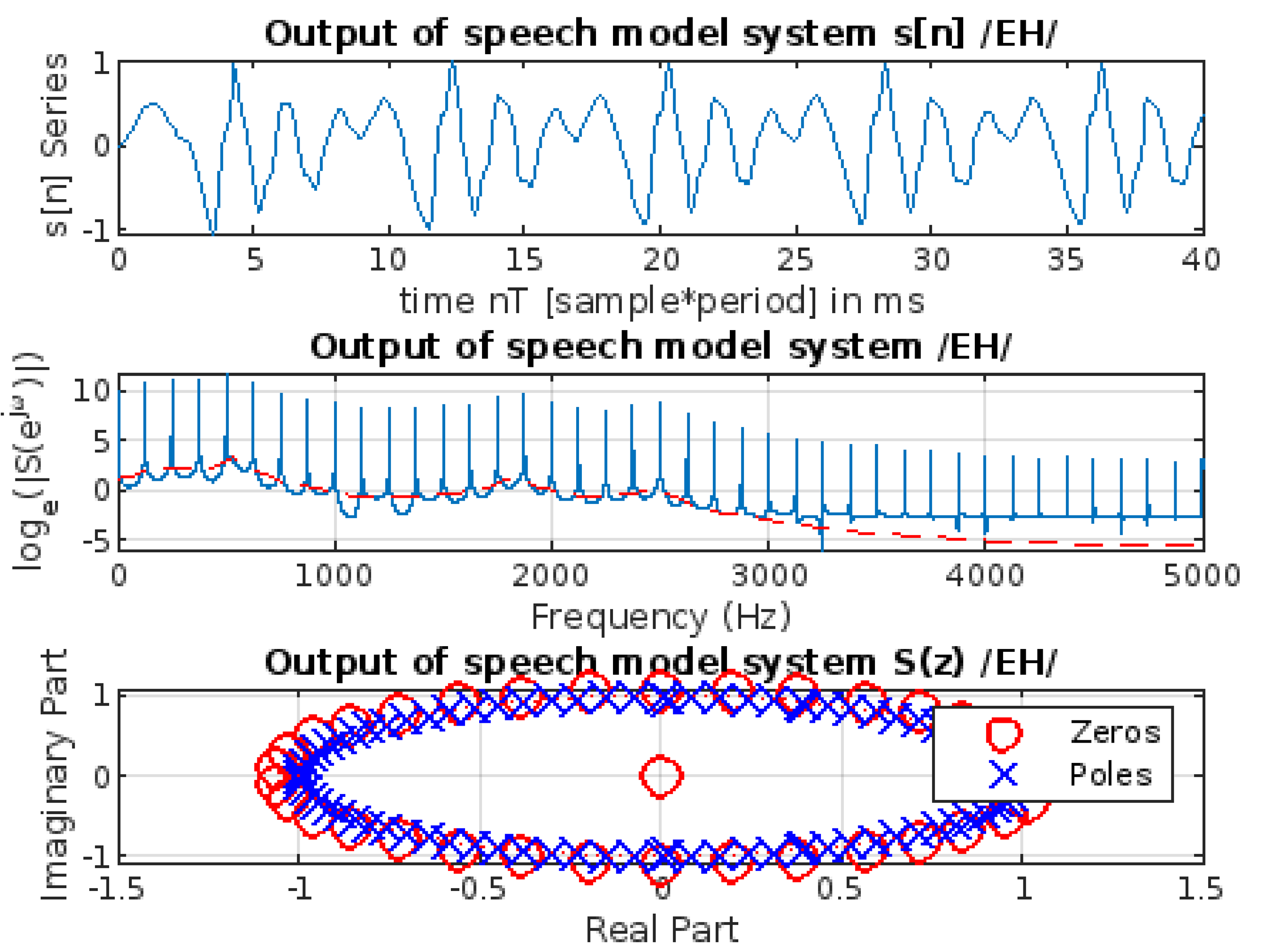


Figure 9. Plot of $s[n]$ for /EH/.

Vowel /AH/

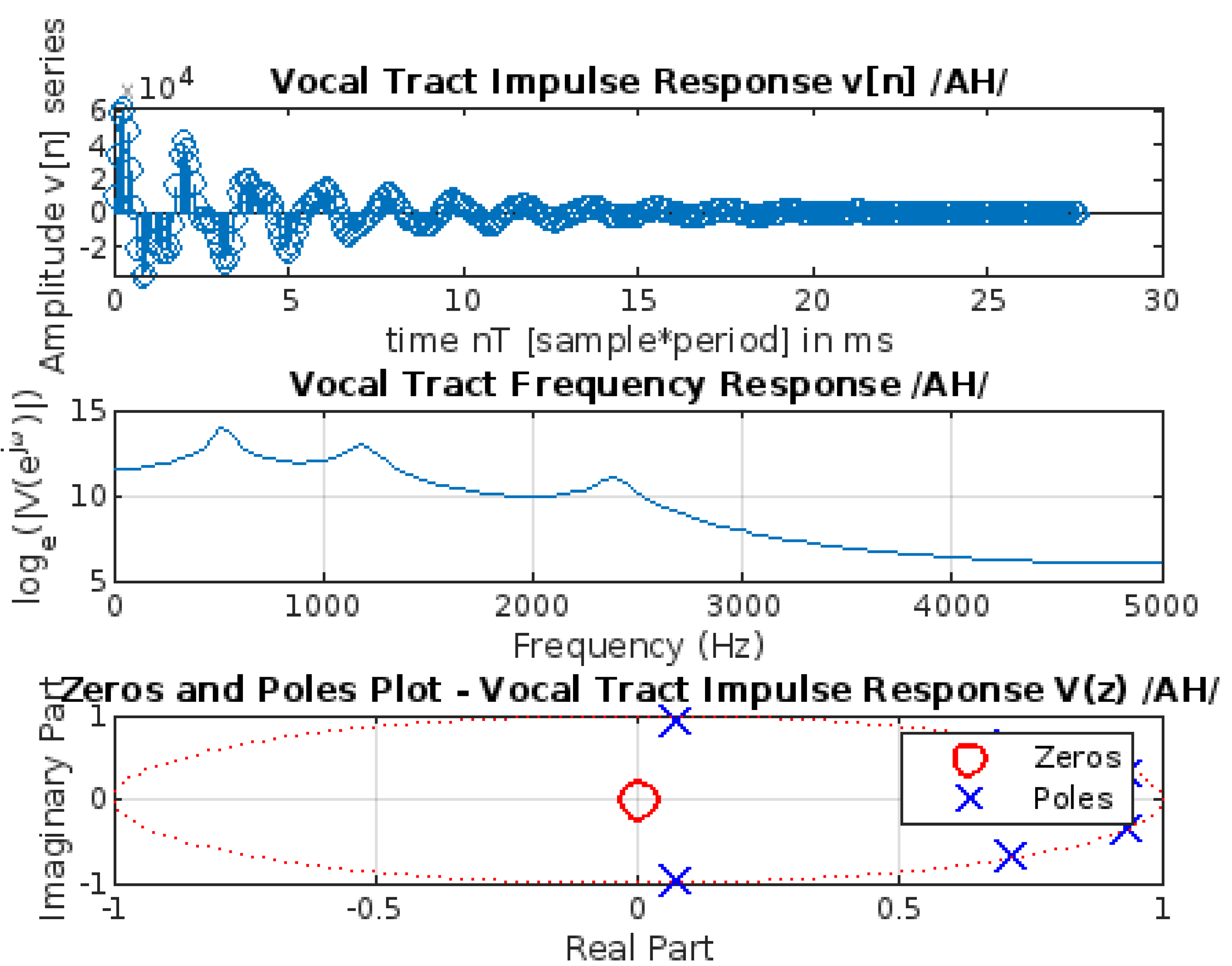


Figure 10. Plot of $V(z)$ for /AH/.

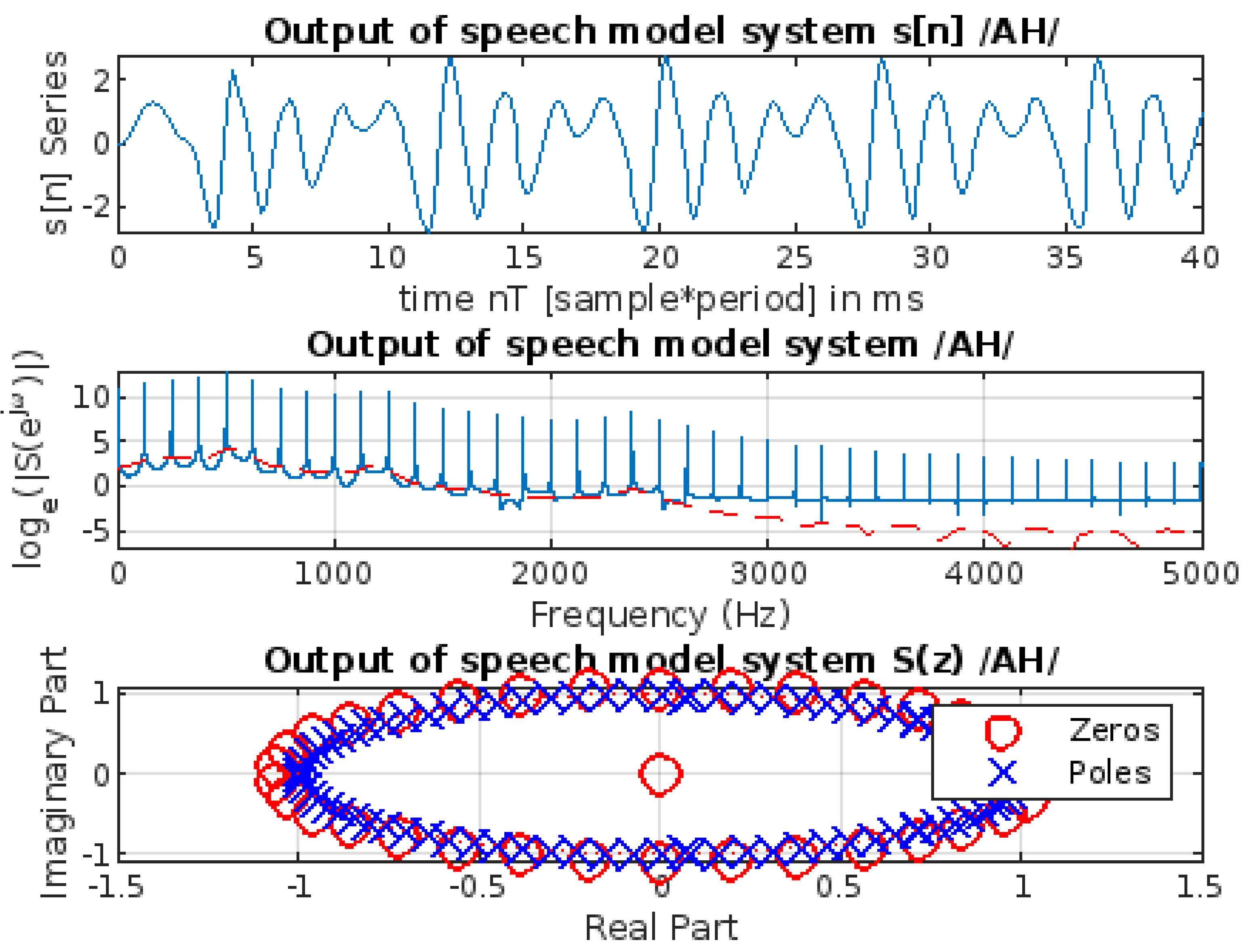
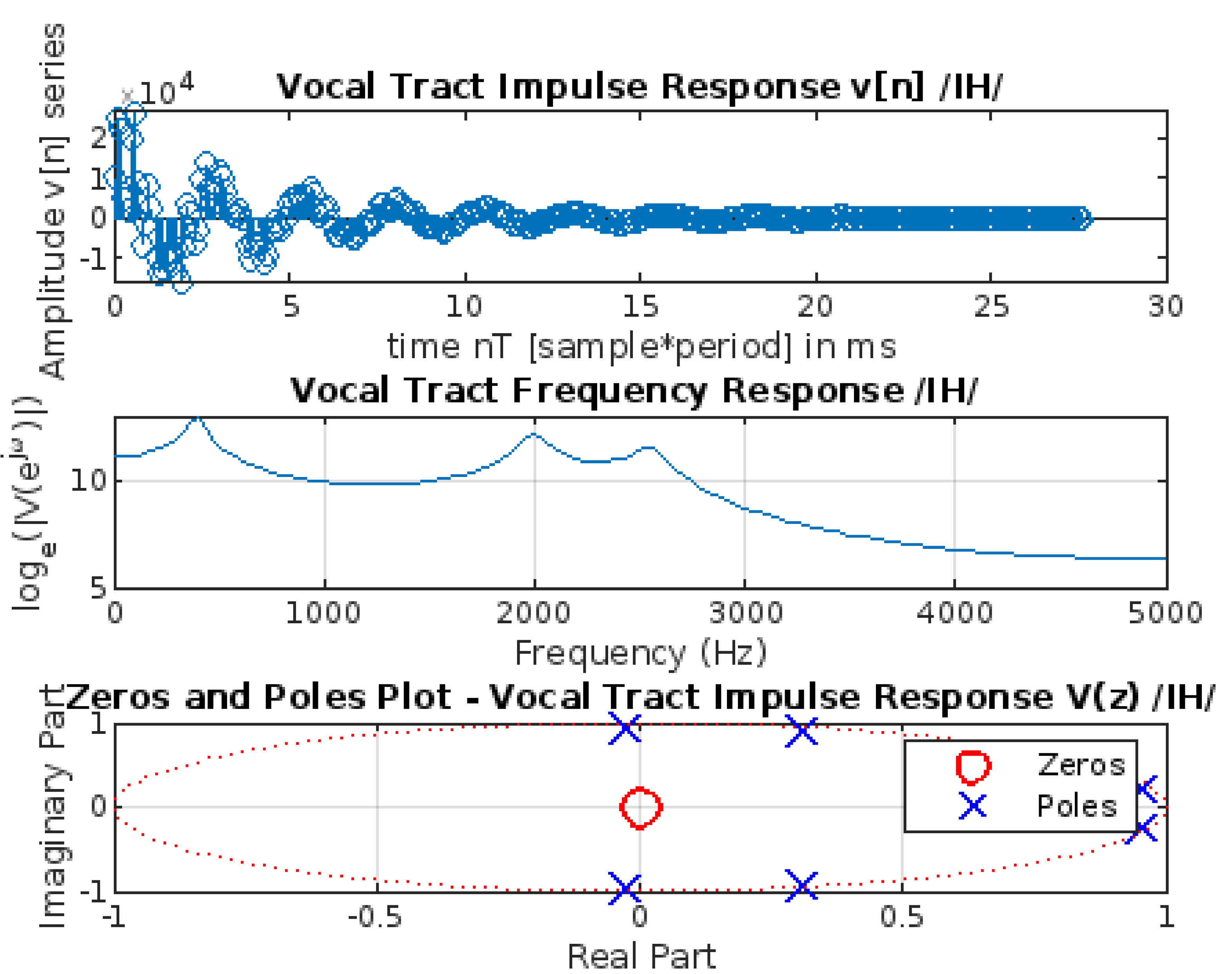
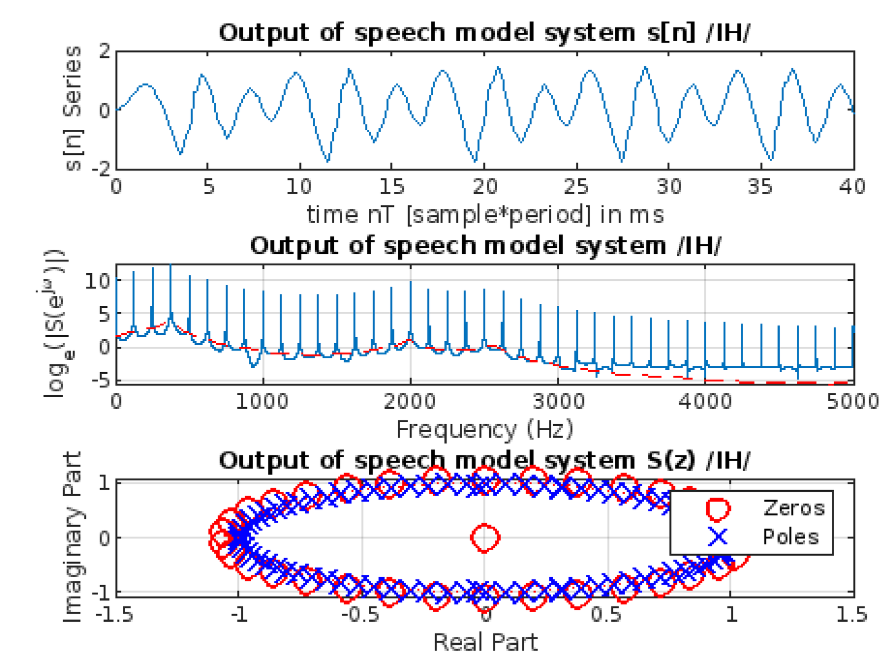


Figure 11. Plot of $s[n]$ for /AH/.

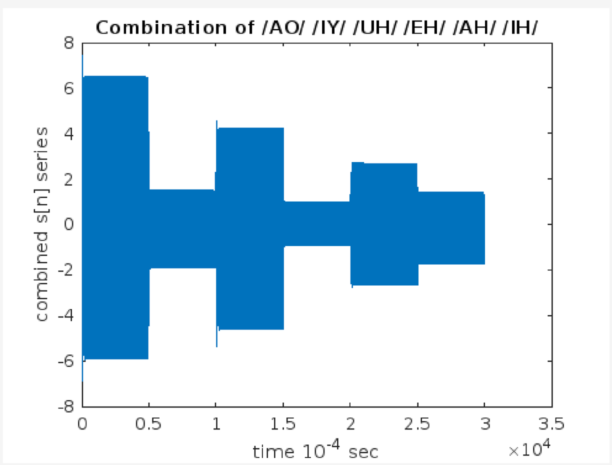
Vowel /IH/





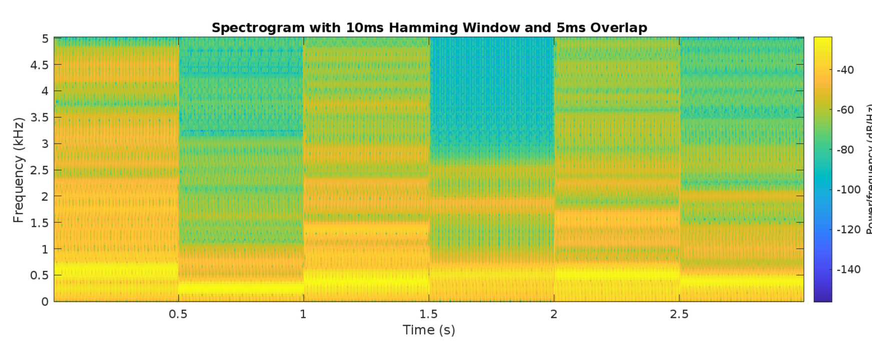
Comparison of Synthetic Vowel Signals

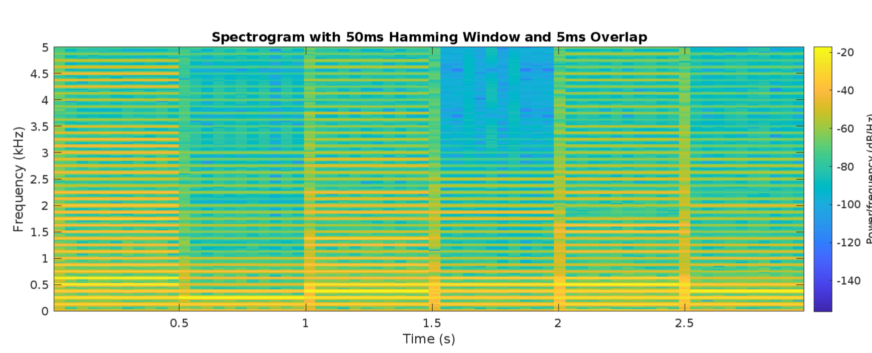
Design and comment on the six synthetic signals $s[n]$ (for 1000 samples) corresponding to the vowels, highlighting differences in their shapes.



Spectrograms of Synthetic vs Natural Voice

Design the spectrogram of the synthetic voice signal, record and analyze the natural voice signal with the same vowels, and compare the spectrograms.





Relative Bibliographic Sources

[1] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." [Computer program], 2024. Retrieved 6 January 2024 from `http://www.praat.org/`.

[2] L. R. Rabiner and R. W. Schafer, *Introduction to Digital Speech Processing*. 2007.

[3] The SoX Project, Available online at `http://sox.sourceforge.net`, *SoX: Sound eXchange*, 2024.

pmoschogiannis@uth.gr | vasildimitriou@uth.gr