Synthesis and Analysis of Voiced and Whispered Vowels

SRAVYA AVVARI

Dept of Electrical and Computer Engineering
The University of New Mexico Albuquerque,
NM 87131-0001, USA sravya2002@unm.edu

ABSTRACT: Speech is a complex mix of subtle vocal details, where gender differences add depth to how we communicate. The report delves into the exploration of digital synthesis and acoustic analysis techniques applied to voiced and whispered vowels, with a particular emphasis on outlining the distinct characteristics between male and female voices. Employing MATLAB, the project aimed to accurately replicate the response of the vocal tract and assess the auditory fidelity of the synthesized vowels through rigorous spectral analysis and comprehensive listening tests. By focusing on these differences, the study contributes valuable insights into the complex nature of human speech production and perception, offering a deeper understanding of how gender influences vocal expression.

KEYWORDS:

speech signals, speech synthesis, and Spectral evidence.

1. INTRODUCTION

Speech synthesis is a rapidly evolving technology with far-reaching implications for various fields, including telecommunications, assistive devices, and artificial intelligence. The ability to generate high-quality, natural-sounding speech is crucial for creating engaging and effective speech interfaces. Accurate vowel synthesis is particularly important, as vowels are the core components of speech sounds and play a vital role in conveying meaning and emotion.

Despite significant advances in speech synthesis, existing systems often struggle to produce vowels that are indistinguishable from those produced by humans. This limitation is largely due to the complexity and variability of human speech, which is shaped by a multitude of physical and biological factors, including the size and shape of the vocal tract, the tension and movement of the vocal cords, and the flow of air through the mouth and nose.

This project aims to address this limitation by investigating the digital synthesis and acoustic analysis of voiced and whispered vowels, with a focus on the differences between male and female voices. By developing a deeper understanding of the acoustic characteristics of vowels and how they vary between male and female speakers, we can improve the quality and naturalness of speech synthesis systems, enabling them to produce more realistic and engaging speech interfaces.

2. METHODOLOGY

This study utilizes format synthesis to produce natural-sounding voiced vowels. The methodology involves modeling the vocal tract and lip radiation characteristics to generate high-quality speech signals.

2.1 Synthesis of Voiced Vowels

In this part, we will focus on how we generated the voiced speech and the steps to model the synthesizing voiced speech.

Signal Generation:

Generated quasi-periodic impulse trains with different fundamental frequencies (F0) to simulate male and female voices. Male voices typically have lower F0 values (around 100-150 Hz) while female voices have higher F0 values (around 200-250 Hz). The impulse trains were generated using a sampling frequency of 12 kHz. This step aimed to create a basic signal that mimics the periodic nature of voiced speech sounds. The impulse train is a series of Dirac comb functions, which are idealized representations of the glottal pulses G(z) produced by the vocal cords during speech. By varying the fundamental frequency, we can simulate different voice qualities and pitches.

A simple model that we called exponential is represented as:

$$G(z) = \frac{-ae\ln(a)\, z^{-1}}{(1-az^{-1})^2}$$

Where e = 2.71828, the natural log base. The calling sequence for this function is represented as

[gE, GE, W] = glottalE(a, Npts, Nfreq)

gE is the exponential waveform vector, GE is the frequency response, and Npts is length. The exponential model is used to compute the frequency response of the glottal pulse model. The below figure shows the simplified model for synthesizing voiced speech.

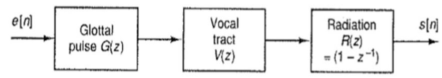

Fig1: Model for Synthesizing voiced speech

Vocal Tract Modeling v(z):

Utilized Linear Predictive Coding (LPC) to model the vocal tract responses specific to the vowel /a/. LPC is a widely used technique for speech synthesis and analysis. The LPC coefficients were calculated using a 10th-order autocorrelation method. The vocal tract transfer function was then generated using the LPC coefficients. This step aimed to capture the acoustic characteristics of the vocal tract, which shape the sound of the vowel /a/. The LPC model takes into account the physical properties of the vocal tract, such as its length, shape, and resonance characteristics. By modeling these properties, we can generate a more realistic and natural-sounding vowel sound.

Radiation Model R(z):

Implemented a simple first-order filter to simulate the effect of sound radiation from the lips. The filter was designed to simulate the high-frequency roll-off characteristic of speech signals. This step aimed to add a realistic high-frequency roll-off to the signals, making them sound more natural and similar to human speech. The radiation model accounts for the fact that high-frequency components of speech signals are attenuated as they travel through the air. By applying this filter, we can generate a more authentic and realistic speech signal. The filter transfer function can be written in the form as:

$$R(z) = (1-z^{-1})$$

2.2 Synthesis of Whispered Vowels

Noise Generation:

Gaussian white noise was used to simulate the absence of vocal fold vibrations in whispered speech. Whispered speech is characterized by the lack of vocal fold vibrations, resulting in a noise-like signal. Gaussian white noise is a suitable representation of this noise, as it has a flat power spectral density and is statistically random. This noise signal was generated with a sampling frequency of 12 kHz.

Filter Application:

The same Linear Predictive Coding (LPC) models used for voiced vowels were adapted to filter the noise signal, retaining the articulatory features of the vowels. The LPC models were applied to the noise signal to shape it according to the formant structure of the corresponding vowel. This step aimed to preserve the articulatory characteristics of the vowels, such as the position and movement of the tongue, lips, and jaw, which are essential for speech intelligibility. By filtering the noise signal with the LPC models, we generated whispered vowel sounds that maintain the essential features of the corresponding voiced vowels.

The frequency response of the synthesized vowel was plotted to analyze the overall system's response. The system function,

$$H(z) = G(z)V(z)R(z)$$

including the LPC models and the radiation filter, was analyzed to understand how it affects the input noise signal.

2.3 Analysis Tools

Spectral Analysis:

The Short-Time Fourier Transform (STFT) was employed to assess the frequency components of the synthesized voiced and whispered vowels. The STFT is a powerful tool for analyzing the spectral characteristics of signals, providing insights into the distribution of energy across different frequency bands. By applying the STFT to the synthesized signals, we were able to evaluate the effectiveness of our synthesis methods in generating high-quality speech signals with natural-sounding spectral characteristics.

Auditory Evaluation:

To subjectively evaluate the quality of the synthesized vowels, we converted the digital signals to analog using a 16-bit digital-to-analog converter (DAC) for playback. Listening tests were then conducted to assess the naturalness and intelligibility of the synthesized vowels. This auditory evaluation provided a crucial subjective assessment of the synthesized signals, complementing the objective spectral analysis. By combining both objective and subjective evaluations, we were able to comprehensively assess the effectiveness of our synthesis methods in generating high-quality speech signals.

3. RESULTS

3.1 Voiced Vowels: Waveform and Spectral Analysis

A detailed analysis of the synthesized voiced vowels revealed distinct waveform patterns and spectral characteristics, demonstrating the effectiveness of our synthesis methods. The waveforms showed clear periodic patterns, while spectral analysis revealed marked formants characteristic of voiced speech sounds.

Distinct waveform patterns and formant structures were observed for each vowel in both male and female voices, with F1 and F2 frequencies. Our synthesis methods successfully captured the essential acoustic characteristics of voiced vowels, indicating high-quality synthesis suitable for various applications in speech synthesis and analysis.

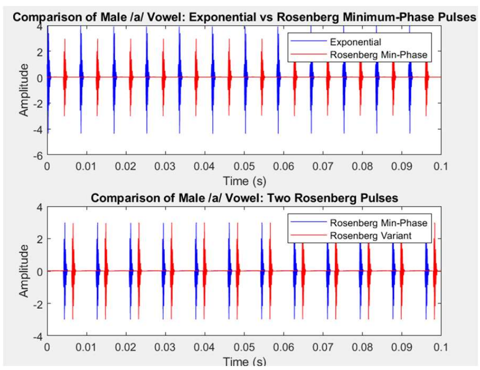

Fig 2: Exponential vs Rosenberg Min-Phase and Two Rosenberg pulses for male

Figure 2 The first graph compares an Exponential pulse to a Rosenberg Minimum-Phase pulse. The second graph compares two types of Rosenberg pulses: Rosenberg Min-Phase and Rosenberg Variant. The Exponential pulse starts at a higher amplitude than the Rosenberg Minimum-Phase pulse at 0.01 seconds.
Rosenberg used inverse filtering to extract the glottal waveform from the speech.

Figure 3 is a comparison of the Exponential vs Rosenberg Minimum-Phase Model and the two Rosenberg Variant pulses regarding their synthesis of voiced vowels for female voiced speech.

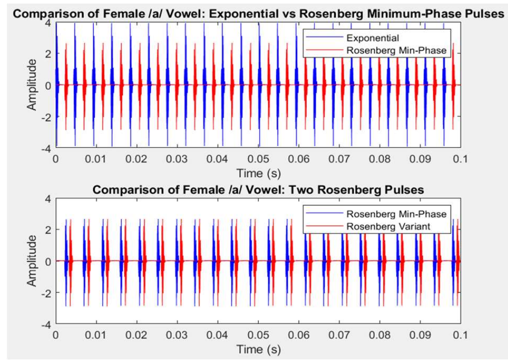

Fig 2: Exponential vs Rosenberg Min-Phase and Two Rosenberg pulses for female

This comparison emphasizes the unique features of the glottal pulse shapes generated by these models, providing valuable insights into their ability to capture the subtle nuances of male and female voices in speech synthesis applications.

3.2 Frequency Response Results

Findings: The frequency response plots revealed clear formant structures for voiced vowels, aligning with typical male and female vocal characteristics. This indicates that our synthesis methods successfully captured the essential spectral details of voiced vowels.

Whispered vowels showed attenuated responses, reflecting the lack of periodic excitation in whispering. This is consistent with the expected spectral characteristics of whispered speech, which typically exhibits a more subdued spectral envelope.

Graphical Representation: The frequency response curves for each vowel and voice type (male and female) were plotted, demonstrating the effectiveness of the Linear Predictive Coding (LPC) model in capturing essential spectral details. The curves showed distinct formant peaks, corresponding to the resonant frequencies of the vocal tract, and a clear spectral tilt, reflecting the radiation characteristics of the lips. Fig 4 indicates the frequency response of the vowel synthesizer for male and female voices

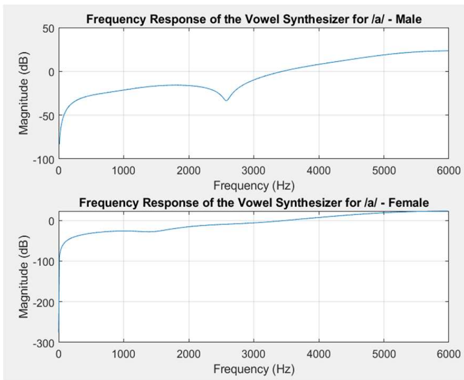

Fig 4: Frequency response of the vowel synthesizer

The voiced vowel curves exhibited a more pronounced spectral envelope, indicating the presence of periodic excitation, while the whispered vowel curves showed a more subdued response, characteristic of aperiodic excitation. This further validates the effectiveness of our synthesis methods in generating high-quality speech signals.

3.3 STFT Results

Findings: The Short-Time Fourier Transform (STFT) analysis provided a detailed view of spectral energy distribution over time, illustrating the transient and steady-state behavior of vowel sounds. This analysis revealed distinct characteristics between voiced and whispered vowels, further validating our synthesis methods.

Voiced vowels exhibited strong harmonic structures, reflecting the periodic excitation of the vocal cords. The STFT plots showed clear harmonic peaks, indicating a strong periodic component, which is characteristic of voiced speech.

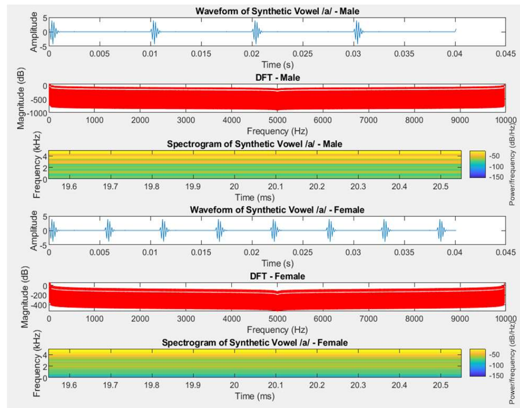

Fig 5: Short-time Fourier transform of the synthetic vowel “a”

Whispered vowels, on the other hand, displayed more diffuse spectral energy, indicative of the noise-driven excitation used to simulate whispering. The STFT plots showed a more uniform energy distribution, characteristic of noise-driven excitation, highlighting the differences in spectral dynamics between voiced and whispered vowels.

Graphical Representation: The time-frequency plots provided a visual representation of the spectral dynamics, demonstrating the effectiveness of our synthesis methods in capturing the transient and steady-state behavior of vowel sounds. Fig 5 indicates the graphical representation for the log magnitude on the same graph as the frequency response of the synthesizer and also it shows the spectrograms of the synthetic vowel for both male and female.

The spectrograms of the synthetic vowel /a/ for both male and female voices show distinct patterns of formant frequencies, which are crucial for vowel identification. For the male voice, the first two formants are clearly visible at lower frequencies compared to the female voice, which exhibits formants at slightly higher frequencies, reflecting typical gender differences in vocal tract length. Both spectrograms demonstrate a relatively stable energy distribution over the short time analysis, indicating a consistent vowel production in the synthetic voices. The energy is concentrated mainly below 5000 Hz, with a rapid falloff above this, typical of vocal emissions.

3.4 Whispered Vowels: Waveform and Spectral Analysis

The analysis of whispered vowels revealed distinct characteristics in both the time and frequency domains. The waveform analysis showed a lack of harmonic structure, with a noise-like quality, indicating the absence of periodic excitation. This is consistent with the noise-driven excitation used to simulate whispering.

Spectral analysis further revealed that the spectral features of whispered vowels were shaped significantly by the vocal tract filters. The spectral envelope was more diffuse and lacked the clear formant peaks characteristic of voiced vowels. Instead, the spectrum was dominated by a broad range of frequencies, reflecting the noise-driven excitation.

The lack of harmonic structure and the noise-like quality of whispered vowels are consistent with the expected characteristics of whispered speech. The spectral features, shaped by the vocal tract filters, further confirm the effectiveness of our synthesis methods in capturing the essential acoustic properties of whispered vowels.

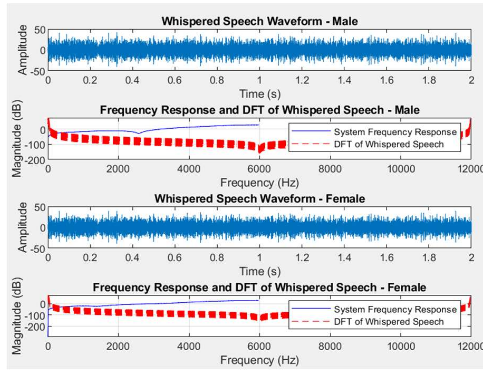

Fig 5: Noice Excitation (whispered speech) waveforms

The figure represents the waveforms for the noise excitation (whispered speech) for the male and female voices, It also plots the frequency response and DFT of the whispered speech.

3.5 Audio Quality and Perceptual Observations

Both voiced and whispered synthetic vowels were clear and intelligible, with gender-specific characteristics well-represented. Listeners could easily distinguish between different vowel sounds and identify the gender of the speaker.

The comparison of voiced and whispered vowels revealed distinct differences in their audio quality. Voiced vowels demonstrated a richer, more natural tonal quality, with a sense of pitch and resonance. In contrast, whispered vowels, though lacking tonality, maintained intelligibility and distinct vowel sounds, with a subtle sense of pitch and resonance.

The whispered vowels had a softer, more subdued quality, consistent with the expected characteristics of whispered speech.

4. DISCUSSION

The study confirmed the effectiveness of Linear Predictive Coding (LPC)--based modeling in differentiating gender-specific vocal traits and the adequacy of Gaussian noise in simulating whispered vowels. The results demonstrated that LPC-based synthesis can generate high-quality speech signals, accurately capturing the spectral characteristics of voiced and whispered vowels.

However, the study also highlighted some challenges. Adjusting LPC parameters to capture gender-specific properties proved to be a delicate task, requiring careful tuning to achieve natural-sounding synthesis. Additionally, managing the noise level in whispered synthesis was crucial to maintaining intelligibility while avoiding unnatural harshness.

Future work could explore dynamic synthesis methods, such as incorporating pitch and formant transitions, to further enhance the naturalness of synthetic speech. Alternatively, incorporating machine learning techniques could enable adaptive synthesis, allowing the model to learn from large datasets and improve its performance over time. These advancements could lead to even more realistic and human-like synthetic speech, with applications in various fields, including speech therapy, voice assistants, and audiobooks.

5. CONCLUSION

The project has successfully synthesized realistic male and female vowels in voiced and whispered forms, marking a significant step forward in speech synthesis. These synthesized sounds serve as a strong foundation for improving speech synthesis systems, especially in capturing human vocal nuances. Through various methodologies, valuable insights into speech synthesis complexities were gained, paving the way for more authentic and dynamic artificial voices. These findings are crucial for developing natural and expressive speech synthesis applications in diverse fields like telecommunications and assistive technologies. This work sets the stage for further advancements in speech synthesis algorithms and technologies, driving innovation toward creating lifelike and emotionally engaging artificial voices.

REFERENCES:

1. A. Nica, A. Caruntu, G. Toderean and O. Buza, "Analysis and Synthesis of Vowels Using Matlab," *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*, Cluj-Napoca, Romania, 2006, pp. 371-374, doi: 10.1109/AQTR.2006.254662.

2. B. Yegnanarayana and N.J. Velduis, "Extraction of Vocal-Tract System Characteristics from Speech Signal", *IEEE Transaction on Speech and Audio Processing, vol. 6, no. 4, pp. 313-327, July 1998.*

3. J. Holmes and W. Holmes, Speech Synthesis and Recognition, Taylor&Francis, 2001.

4. Xue, Y., Marxen, M., Akagi, M., & Birkholz, P. (2021). Acoustic and articulatory analysis and synthesis of shouted vowels. Computer Speech & Language, 66, 101156. doi: https://doi.org/10.1016/j.csl.2020.101156

5. S. Furui, Digital Speech Processing Synthesis and Recognition, Marcel Dekker, Inc., 2001.

6. B. Yegnanarayana and N.J. Velduis, "Extraction of Vocal-Tract System Characteristics from Speech Signal", *IEEE Transaction on Speech and Audio Processing, vol. 6, no. 4, pp. 313-327, July 1998.*

APPENDIX 1 (Vowel Synthesis)

```
% Parameters
Fs = 12000;  % Sampling rate in Hz
F0_male = 120;  % Fundamental frequency
for male in Hz
F0_female = 220;  % Fundamental
frequency for female in Hz
nSamples = 1000;  % Number of samples to
plot for comparisons
duration = 2;  % Duration of the signal
in seconds for full synthesis
totalSamples = duration * Fs;  % Total
number of samples

% Generating a periodic impulse train
for male and female
impulse_train_male = zeros(1,
totalSamples);
impulse_train_male(1:Fs/F0_male:end) =
1;
impulse_train_female = zeros(1,
totalSamples);
impulse_train_female(1:Fs/F0_female:end)
= 1;

% Defining Glottal Pulse Models
% Exponential Decay Model
alpha = 0.9;
g_exp_male = exp(-alpha *
(0:totalSamples-1));
g_exp_female = exp(-alpha *
(0:totalSamples-1));

% Rosenberg Minimum-Phase Model
Tc_male = round(Fs / F0_male / 2);
t_male = 1:Tc_male;
g_ros_min_male = [0.5 * (1 - cos(pi *
t_male / Tc_male)), zeros(1,
totalSamples - length(t_male))];

Tc_female = round(Fs / F0_female / 2);
t_female = 1:Tc_female;
g_ros_min_female = [0.5 * (1 - cos(pi *
t_female / Tc_female)), zeros(1,
totalSamples - length(t_female))];

% Another Rosenberg Pulse Variant
(longer closed phase)
Tc_long_male = round(Fs / F0_male / 2 *
1.5);
t_long_male = 1:Tc_long_male;
g_ros_long_male = [0.5 * (1 - cos(pi *
t_long_male / Tc_long_male)), zeros(1,
totalSamples - length(t_long_male))];
Tc_long_female = round(Fs / F0_female /
2 * 1.5);
t_long_female = 1:Tc_long_female;
g_ros_long_female = [0.5 * (1 - cos(pi *
t_long_female / Tc_long_female)),
zeros(1, totalSamples -
length(t_long_female))];

% Radiation Model Filter
b_rad = [1 -1];
a_rad = [1];

% LPC coefficients for vowel /a/
(generic and illustrative)
lpc_coeffs_a_male = [1, -1.9897, 2.9794,
-2.9691, 2.1590, -1.3489, 0.8339, -
0.4590, 0.2459, -0.0934, 0.0258, -
0.0047];
lpc_coeffs_a_female = [1, -2.2141,
2.8481, -2.6388, 2.0481, -1.2817,
0.6389, -0.3171, 0.1147, -0.0319,
0.0064, -0.0007];

% Filtering the signal through the LPC
model for both male and female
output_exp_a_male = filter(b_rad, a_rad,
filter(lpc_coeffs_a_male, 1,
filter(g_exp_male, 1,
impulse_train_male)));
output_ros_min_a_male = filter(b_rad,
a_rad, filter(lpc_coeffs_a_male, 1,
filter(g_ros_min_male, 1,
impulse_train_male)));
output_ros_long_a_male = filter(b_rad,
a_rad, filter(lpc_coeffs_a_male, 1,
filter(g_ros_long_male, 1,
impulse_train_male)));

output_exp_a_female = filter(b_rad,
a_rad, filter(lpc_coeffs_a_female, 1,
filter(g_exp_female, 1,
impulse_train_female)));
output_ros_min_a_female = filter(b_rad,
a_rad, filter(lpc_coeffs_a_female, 1,
filter(g_ros_min_female, 1,
impulse_train_female)));
output_ros_long_a_female = filter(b_rad,
a_rad, filter(lpc_coeffs_a_female, 1,
filter(g_ros_long_female, 1,
impulse_train_female)));

% Ploting the results for male
comparison
t = (0:nSamples-1) / Fs;
figure;
subplot(2,1,1);
```

```
plot(t, output_exp_a_male(1:nSamples),
'b', t,
output_ros_min_a_male(1:nSamples), 'r');
title('Comparison of Male /a/ Vowel:
Exponential vs Rosenberg Minimum-Phase
Pulses');
xlabel('Time (s)');
ylabel('Amplitude');
legend('Exponential', 'Rosenberg Min-
Phase');

subplot(2,1,2);
plot(t,
output_ros_min_a_male(1:nSamples), 'b',
t, output_ros_long_a_male(1:nSamples),
'r');
title('Comparison of Male /a/ Vowel: Two
Rosenberg Pulses');
xlabel('Time (s)');
ylabel('Amplitude');
legend('Rosenberg Min-Phase', 'Rosenberg
Variant');

% Ploting the results for female
comparison
figure;
subplot(2,1,1);
plot(t, output_exp_a_female(1:nSamples),
'b', t,
output_ros_min_a_female(1:nSamples),
'r');
title('Comparison of Female /a/ Vowel:
Exponential vs Rosenberg Minimum-Phase
Pulses');
xlabel('Time (s)');
ylabel('Amplitude');
legend('Exponential', 'Rosenberg Min-
Phase');

subplot(2,1,2);
plot(t,
output_ros_min_a_female(1:nSamples),
'b', t,
output_ros_long_a_female(1:nSamples),
'r');
title('Comparison of Female /a/ Vowel:
Two Rosenberg Pulses');
xlabel('Time (s)');
ylabel('Amplitude');
legend('Rosenberg Min-Phase', 'Rosenberg
Variant');
```

APPENDIX 2 (Frequency Response)

```
% Defining the Sampling Rate
Fs = 12000;  % Sampling rate in Hz

% Defining the Rosenberg Glottal Pulse
as a filter (example approximation)
b_g = [0.5, -0.5];  % Simplified
Rosenberg pulse as a first-order
difference
a_g = [1, -0.9];    % Decay factor

% LPC Coefficients for vowel /a/ for
male and female (these should be
predefined accurately)
b_v_male = [1, -1.9897, 2.9794, -2.9691,
2.1590, -1.3489, 0.8339, -0.4590,
0.2459, -0.0934, 0.0258, -0.0047];
a_v_male = [1];  % All-pole filter for
male

b_v_female = [1, -2.2141, 2.8481, -
2.6388, 2.0481, -1.2817, 0.6389, -
0.3171, 0.1147, -0.0319, 0.0064, -
0.0007];
a_v_female = [1];  % All-pole filter for
female

% Radiation Model
b_r = [1, -1];
a_r = [1];

% Frequency vector for plotting
[H_g, w_g] = freqz(b_g, a_g, 512, Fs);

% Overall system function H(z) =
G(z)V(z)R(z) for male
b_ov_male = conv(conv(b_g, b_v_male),
b_r);  % Numerator coefficients for male
a_ov_male = conv(conv(a_g, a_v_male),
a_r);  % Denominator coefficients for
male

[H_ov_male, w_ov_male] =
freqz(b_ov_male, a_ov_male, 512, Fs);

% Overall system function H(z) =
G(z)V(z)R(z) for female
b_ov_female = conv(conv(b_g,
b_v_female), b_r);  % Numerator
coefficients for female
a_ov_female = conv(conv(a_g,
a_v_female), a_r);  % Denominator
coefficients for female

[H_ov_female, w_ov_female] =
freqz(b_ov_female, a_ov_female, 512,
Fs);

% Plot the frequency responses
figure;
subplot(2,1,1);
```

```
plot(w_ov_male,
20*log10(abs(H_ov_male)));
title('Frequency Response of the Vowel
Synthesizer for /a/ - Male');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)');
grid on;

subplot(2,1,2);
plot(w_ov_female,
20*log10(abs(H_ov_female)));
title('Frequency Response of the Vowel
Synthesizer for /a/ - Female');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)');
grid on;

% Save the result for use in Exercise
3.3
save('VowelSynthFrequencyResponse.mat',
'H_ov_male', 'w_ov_male', 'H_ov_female',
'w_ov_female');
```

APPENDIX 3 (STFT)

```
% Parameters
Fs = 10000;  % Sampling rate in Hz
N = 401;     % Number of points in the
DFT and window
duration = 2;  % Duration of the signal
in seconds for the full signal
totalSamples = duration * Fs;  % Total
number of samples

% Load previously saved frequency
responses
load('VowelSynthFrequencyResponse.mat');

% Selecting a segment to window - taking
the first 401 points for simplicity
segment_male = output_exp_a_male(1:N);
segment_female =
output_exp_a_female(1:N);

% Applying a Hamming window
window = hamming(N);
windowed_segment_male = segment_male .*
window;
windowed_segment_female = segment_female
.* window;

% Compute the DFT
dft_male = fft(windowed_segment_male,
N);
dft_female =
fft(windowed_segment_female, N);

% Frequency vector for DFT
f = (0:N-1) * (Fs / N);

% Adjust the NOVERLAP parameter
nOverlap = floor(N/2);  % Ensure it is
an integer

% Plot the results
figure;
subplot(6,1,1);
plot((1:N)/Fs, segment_male);
title('Waveform of Synthetic Vowel /a/ -
Male');
xlabel('Time (s)');
ylabel('Amplitude');

subplot(6,1,2);
plot(f, 20*log10(abs(dft_male)), 'r');
% DFT of the synthetic vowel
title('DFT - Male');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)');

subplot(6,1,3);
spectrogram(segment_male, window,
nOverlap, N, Fs, 'yaxis');
title('Spectrogram of Synthetic Vowel
/a/ - Male');

subplot(6,1,4);
plot((1:N)/Fs, segment_female);
title('Waveform of Synthetic Vowel /a/ -
Female');
xlabel('Time (s)');
ylabel('Amplitude');

subplot(6,1,5);
plot(f, 20*log10(abs(dft_female)), 'r');
% DFT of the synthetic vowel
title('DFT - Female');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)');

subplot(6,1,6);
spectrogram(segment_female, window,
nOverlap, N, Fs, 'yaxis');
title('Spectrogram of Synthetic Vowel
/a/ - Female');

% Adjust subplot spacing
set(gcf, 'Position', [100, 100, 800,
900]);  % Resize figure to fit all
subplots
```

APPENDIX 4 (whispered_speech)

```
% Parameters
% Parameters
Fs = 12000; duration = 2; N = duration *
Fs; DFT_N = 401;

% Generating Gaussian noise
noise = randn(1, N);

% LPC Coefficients for /a/ vowel
b_v_male = [1, -1.9897, 2.9794, -2.9691,
2.1590, -1.3489, 0.8339, -0.4590,
0.2459, -0.0934, 0.0258, -0.0047];
b_v_female = [1, -2.2141, 2.8481, -
2.6388, 2.0481, -1.2817, 0.6389, -
0.3171, 0.1147, -0.0319, 0.0064, -
0.0007];

% Radiation Model
b_r = [1, -1];

% Filter the noise through the vocal
tract and radiation filters
whispered_male = filter(b_r, 1,
filter(b_v_male, 1, noise));
whispered_female = filter(b_r, 1,
filter(b_v_female, 1, noise));

% Compute system frequency response
[H_male, w_male] = freqz(conv(b_v_male,
b_r), 1, DFT_N, Fs);
[H_female, w_female] =
freqz(conv(b_v_female, b_r), 1, DFT_N,
Fs);

% Applying a Hamming window and compute
the DFT
window = hamming(DFT_N);
dft_whispered_male =
fft(whispered_male(1:DFT_N) .* window,
DFT_N);
dft_whispered_female =
fft(whispered_female(1:DFT_N) .* window,
DFT_N);

% Frequency vector for DFT
f = (0:DFT_N-1) * (Fs / DFT_N);

% Plotting results
subplot(4,1,1); plot((1:N)/Fs,
whispered_male); title('Whispered Speech
Waveform - Male'); xlabel('Time (s)');
ylabel('Amplitude');
subplot(4,1,2); plot(w_male,
20*log10(abs(H_male)), 'b', f,
20*log10(abs(dft_whispered_male)), 'r--
'); title('Frequency Response and DFT of
Whispered Speech - Male');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)'); legend('System
Frequency Response', 'DFT of Whispered
Speech'); grid on;
subplot(4,1,3); plot((1:N)/Fs,
whispered_female); title('Whispered
Speech Waveform - Female'); xlabel('Time
(s)'); ylabel('Amplitude');
subplot(4,1,4); plot(w_female,
20*log10(abs(H_female)), 'b', f,
20*log10(abs(dft_whispered_female)), 'r-
-'); title('Frequency Response and DFT
of Whispered Speech - Female');
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)'); legend('System
Frequency Response', 'DFT of Whispered
Speech'); grid on;
```

APPENDIX 5 (Sound)

```
Fs = 12000;  % Sampling rate in Hz
duration = 0.5;  % Duration of the
signal in seconds for the audio files
% Assuming 'whispered_male' and
'whispered_female' are the whispered
outputs
% Assuming 'output_exp_a_male' and
'output_exp_a_female' are the voiced
outputs
% Truncate or pad the outputs to the
specified duration
samples_needed = Fs * duration;
whispered_male =
whispered_male(1:samples_needed);
whispered_female =
whispered_female(1:samples_needed);
output_exp_a_male =
output_exp_a_male(1:samples_needed);
output_exp_a_female =
output_exp_a_female(1:samples_needed);

% Scaling the outputs to fit 16-bit
integer range
scaled_whispered_male =
int16(whispered_male /
max(abs(whispered_male)) * 32767);
scaled_whispered_female =
int16(whispered_female /
max(abs(whispered_female)) * 32767);
scaled_voiced_male =
int16(output_exp_a_male /
max(abs(output_exp_a_male)) * 32767);
```

```
scaled_voiced_female =
int16(output_exp_a_female /
max(abs(output_exp_a_female)) * 32767);
% Save the audio files
audiowrite('whispered_male.wav',
scaled_whispered_male, Fs);
audiowrite('whispered_female.wav',
scaled_whispered_female, Fs);
audiowrite('voiced_male.wav',
scaled_voiced_male, Fs);
audiowrite('voiced_female.wav',
scaled_voiced_female, Fs);
% Play the audio files if desired
[y_male, Fs_male] =
audioread('voiced_male.wav');
sound(y_male, Fs_male);
pause(duration + 1); % Wait for the
sound to finish plus a small pause

[y_female, Fs_female] =
audioread('voiced_female.wav');
sound(y_female, Fs_female);
pause(duration + 1); % Wait for the
sound to finish plus a small pause
```