Assignment 3

Singing Synthesis Using the Source-Filter Model

Yilai Chen

1. Introduction

This report presents the implementation of a singing synthesis model based on the source-filter method. The system generates voiced sounds by simulating the glottal source using harmonic sine wave summation and shaping the output using resonance filters to mimic vocal tract formants. The synthesis is controlled through pitch, formant frequencies, bandwidths, and dynamic parameters to achieve realistic vocal characteristics.

2. Methodology

2.1 Source Signal Generation

The source signal is constructed as a sum of at least 30 harmonic partials, approximating a sawtooth waveform with a spectral slope of -6 dB per octave. The fundamental frequency is modulated with vibrato:

- The fundamental frequency (f0) is set based on input parameters.
- Harmonics are generated with decreasing amplitude following the -6 dB/octave rule.
- A vibrato modulation (6 Hz) is applied to the fundamental frequency.

The source signal is generated using the following equation:

$$p(t) = \sum_{i=1}^{N} A(i^{-a_{\text{slope}}}) \sin(2\pi i f_0 t)$$

where `A` is the amplitude, `a_slope` controls the spectral slope, and `N` is the number of harmonics.

2.2 Formant Filtering

To shape the source signal into different vowel sounds, at least five second-order resonance filters are applied in series. The resonance filters are implemented as two-pole IIR filters, designed based on measured formant frequencies and bandwidths for different vowels.

Each filter is defined by:

$$a_1 = -2e^{-\alpha T}\cos(\beta T), \quad a_2 = e^{-2\alpha T}, \quad G = 1 + a_1 + a_2$$

where:

$$\beta = 2\pi f$$
$$\alpha = \frac{\beta_0}{2Q}, \quad \text{with} \quad \beta_0 = \beta\sqrt{1 + \frac{1}{4Q^2}}$$
$$Q = \frac{f}{B}$$

where `Q` is the quality factor, `f` is the formant frequency, and `B` is the bandwidth.

Each vowel has a specific set of formant frequencies and bandwidths, which are used to configure the filters dynamically.

2.3 Dynamic Control

The synthesis includes dynamic control over:

- **Sound level (dB):** Adjusting the amplitude of the output signal.
- **Spectral slope (dB/octave):** Modifying the harmonic amplitude distribution to simulate loud or soft phonation.

2.4 Melody Synthesis

A simple melody based on vowel transitions is synthesized. Each vowel is assigned a different duration and speed factor to shape the rhythm. The melody is generated by concatenating vowel-based synthesized segments.

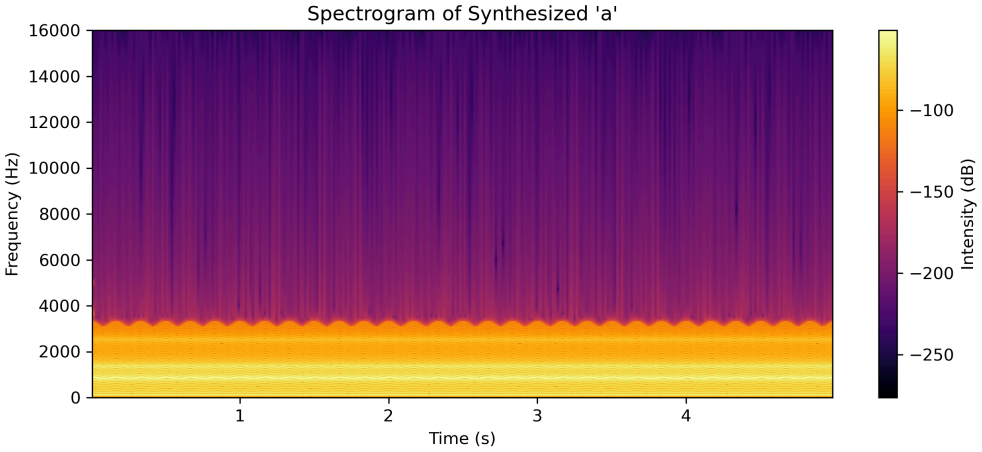
2.5 Data from Voice Lab

Vowel	f0(Hz)	F1(Hz)	F2(Hz)	F3(Hz)	F4(Hz)	F5(Hz)
a	105	840	1360	2520	3640	5000
o	103	320	760	3240	4240	7080
e	104	560	1240	2600	3400	4480
i	105	360	2200	2800	3600	4400
u	107	360	800	2160	3520	4320
v	105	320	1800	2440	3440	4440

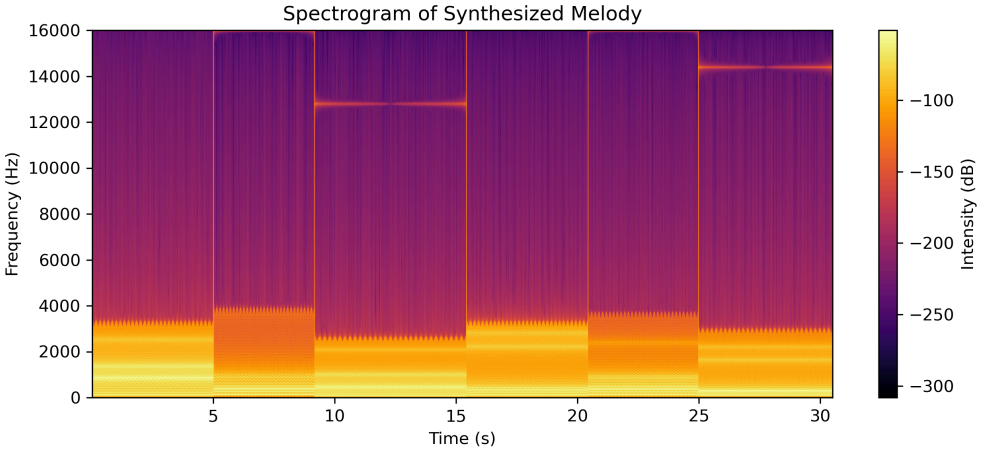
3. Results

3.1 Spectrogram Analysis

Spectrograms of individual vowels and the synthesized melody are plotted to analyze the frequency content and formant structures. The synthesized speech exhibits distinct formant patterns matching expected vowel characteristics.



The spectrogram of the synthesized vowel 'a' . The fundamental frequency and its harmonics are clearly visible, with the formant structure shaped by the applied resonance filters.



The spectrogram of the synthesized melody shows distinct formant structures at different time intervals, indicating vowel transitions. However, abrupt transitions between segments indicate a lack of smooth formant interpolation, which could be improved for a more natural singing synthesis.

3.2 Listening Test

The sound is recognizable as singing vowels but lacks natural breathiness and transient characteristics.

4. Discussion and Improvements

4.1 Issues Identified

- The model lacks noise components, making it sound too artificial.
- The formant transition is abrupt.
- Vibrato is static, not so real.

4.2 Suggested Improvements

- Add aspiration noise.
- Implement dynamic formant transitions.

5. Conclusion

This project successfully implemented a singing synthesis model using the source-filter method. The results demonstrate vowel-specific spectral characteristics and recognizable melodic patterns. Further improvements can enhance naturalness by integrating dynamic transitions and noise modeling.