University of Sheffield

COM3502-4502-6502
Speech Processing

Main Programming Assignment

Da Eun Kim

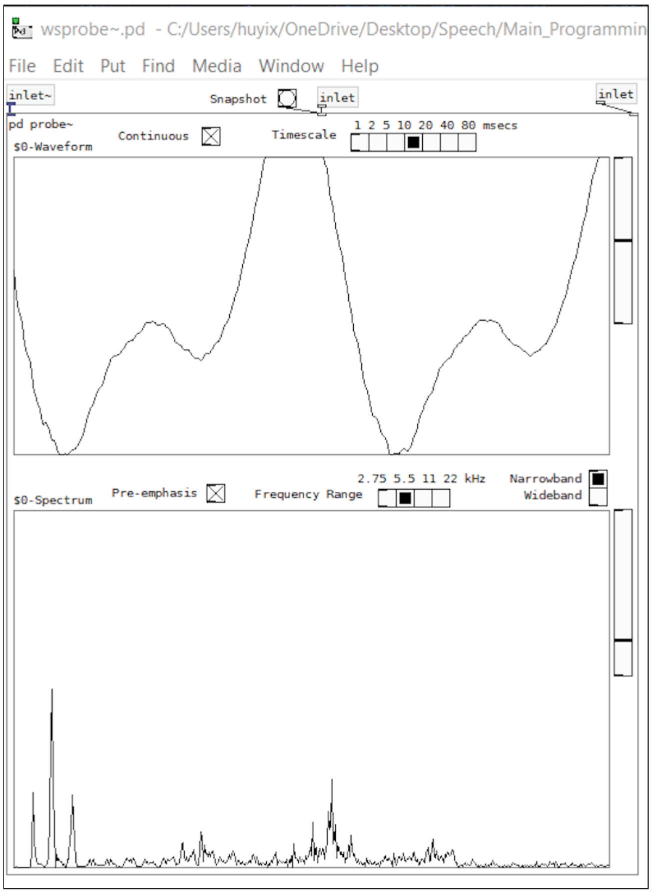
Yixiao Hu

Department of Computer Science

December 16, 2019

QUESTION 1 ***(worth up to 5 marks)***
Provide a screenshot of `[wsprobe~]` for a typical voiced sound, and explain the features in the waveform and spectrum that distinguish it from an unvoiced sound. *Hint: use the 'snapshot' feature in `[wsprobe~]` to obtain a static display.*

Voiced waveform has an impulse response. Also its wave form is periodic which means it has repeating pattern. Therefore, we can observe a fundamental frequency and harmonics in the spectrum graph. Also, in unvoiced waveform, it is aperiodic which means there is no repetitive pattern. Neither we cannot calculate a fundamental frequency in a spectrum.



QUESTION 2 ***(worth up to 5 marks)***
Which sounds are most affected when the low-pass cut-off frequency is set to around 500 Hz - vowels or consonants - and why?

Consonants will be the most affected. In waveform, there is obvious amplitude reduced both in the waveform and spectrum. In the other words, there is energy loss because of the low-pass cut-off frequency. Especially in the fricative consonants, the amplitutde diminish largely. Waveform in vowels, the amplitude is almost the same but smoothen and the spectrum has the same formant at first (below 500Hz) but above 500Hz is cut-off so it is flat and quiet.

QUESTION 3 ***(worth up to 5 marks)***
How is it that the speech is still quite intelligible when the high-pass cut-off frequency is set to 10 kHz?

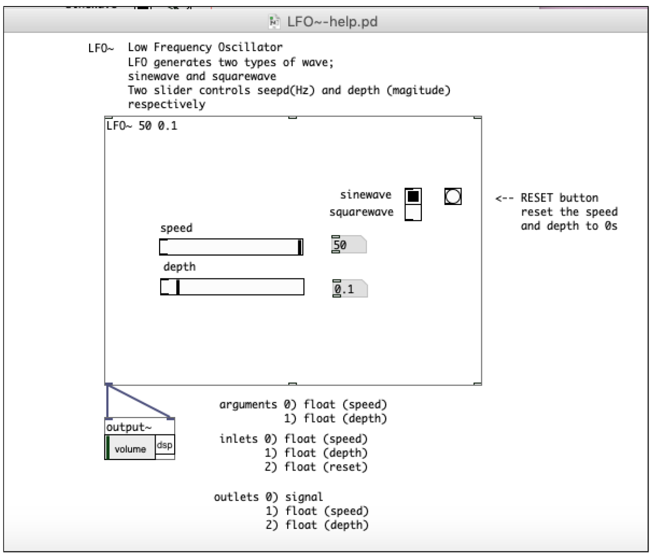
The 'tone' became different, the sound in high-pass cut-off frequency is being electrified. There is a significant difference in the amount(amplitude), which becomes less. But still, we can clearly hear the sound since humans hear up to 20 kHz. Also, less amount of amplitude can be observed through the various range of frequencies.

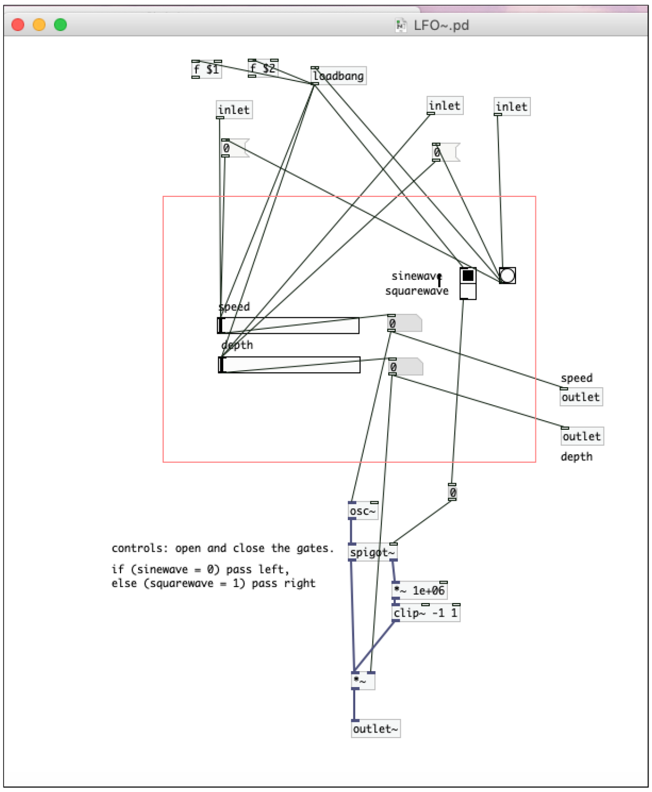
QUESTION 4 ***(worth up to 5 marks)***
COM3502-4502-6502: The `[GraphicEqualiser~]` object uses an FFT internally; what does FFT stand for and what does an FFT do?
COM4502-6502 ONLY: What is a DFT and how is it different from an FFT?

FFT stands for Fast Fourier Transform which is an effective algorithm to calculate DFT(Discrete Fourier Transform). FFT object does Fourier analyses and resynthesis of signals. It converts event-based information into frequency-based information, i.e. converts the time-domain signal into a frequency-domain signal, and whether the data is rewritten in each newly generated block. Different filter functions can be realized with FFT.

QUESTION 5 ***(worth up to 10 marks)***
With `speed = 50` and `depth = 0.5`, what are the minimum and maximum amplitudes of your LFO output, and how do they vary with changes in these two settings? Also, please provide two screenshots: (a) your `[LFO~-help]` object and (b) the internal structure of your `[LFO~]` object.

For a sine wave, the minimum amplitude was -0.49999 and the maximum was 0.5. For a square wave, the maximum and the minimum amplitude were 0.5 and -0.5 respectively. There is no difference between the amplitudes of sine and square waves.





QUESTION 6 ***(worth up to 5 marks)***
In your own words[1], why is this effect known as 'ring modulation'?

Modulation is a processing method where certain characteristics of a waveform change according to another waveform or signal. Also, its analog circuit which is used to make this effect has a ring configuration so it is named as ring modulation.

QUESTION 7 ***(worth up to 5 marks)***
Why is SSB commonly used in long-distance radio voice communications?

Because of the bandwidth of the modulation signal output by the amplitude modulation(AM) technology is twice that of the source signal, and SSB can avoid doubling the bandwidth and avoid wasting energy on the carrier, so in long-distance radio voice communication it would be more efficient. Also, it is beneficial because it increases the number of calls that could be transmitted simultaneously.

QUESTION 8 ***(worth up to 5 marks)***
COM3502-4502-6502: Why can the voice be shifted up in frequency much further than it can be shifted down in frequency before it becomes severely distorted? /emphHint: look at `[wsprobe~]`.
COM4502-6502 ONLY: Your frequency shifter changes all the frequencies present in an input signal. How might it be possible to change the pitch of a voice *without* altering the formant frequencies?

[1] I.e. do not plagiarise from Wikipedia.

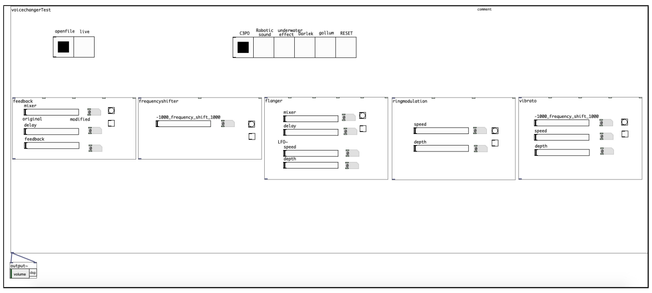
When the frequency shifted down, the fundamental frequency and harmonics are significantly reduced than when the frequency shifted up. Since the fundamental frequency is a element of deciding the tone, loss in fundamental frequency and harmonics could become more distorted.

QUESTION 9 ***(worth up to 5 marks)***
In a practical system, why is it important to keep the feedback gain less than 1?

When the feedback loop is positive and if the gain is above 1, then will occur an exponential increase in a signal. Then it will make a squeaky sound so the sound of its signal is very uncomfortable and dangerous in a practical system.

QUESTION 10 ***(worth up to 50 marks[2])***
Please provide a short[3] description of the operation of your `[VoiceChanger]` application, together with a screenshot of your final GUI.

This is the GUI of the VoiceChanger. On the top left, there are two input options; external file import and live voice. On the top right, there are five options to choose preset effects. And the last option is to reset the sound to the original sound. The user can check whether the feature is turned off or not by checking the toggle box in the object. This voice changer is comprised of 5 objects; feedback, frequency shifter, flanger, ring modulation and vibrato. The user can make their own sound by modifying each value of the sliders in these objects.



[2] 25 for functionality, 15 for design/layout, 5 for `Pd` features, 5 for innovations
[3] no more than 200 words