University of Sheffield

COM3502-4502-6502
Speech Processing

Main Programming Assignment

Uddhav Agarwal

Malvin Todorov

Department of Computer Science

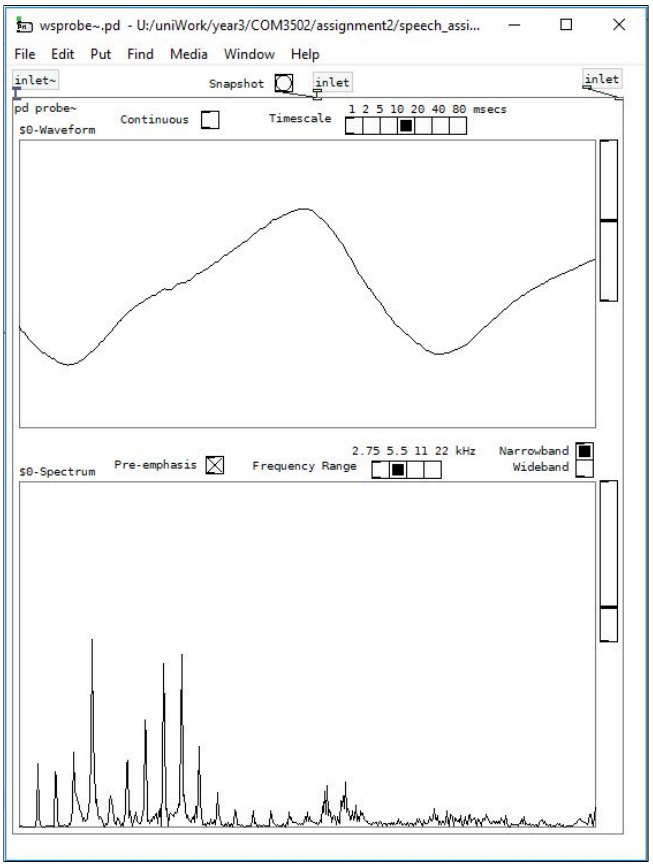
December 7, 2018

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.*

In the following screenshot the waveform is depicting the voiced sound when the speaker is saying the word 'blood'. It is noted that a voiced sound has a repeating pattern and is repeated at regular intervals as can be seen in the waveform, therefore the voiced sounds is periodic. The waveform for an unvoiced sound on the other hand is observed to have a random pattern and is repeated at irregular intervals, hence unvoiced sounds are aperiodic.

The spectrum for the voiced sound shows the troughs between the formant peaks. This is due to some frequencies getting absorbed in our vocal tract. On the other hand an unvoiced sound such as a fricative consonant produces stronger formant peaks with weaker troughs between them because of the signals are stochastic or white noise.



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?

The consonants are most affected when the low-pass cut-off frequency is set to around

500 Hz and to be precise the voiceless fricative consonants are affected. This is due to the fact that for fricative consonants acoustic energy occurs at higher frequencies and when the cut-off frequency is set to the low-end of around 500 Hz the fricative consonants are affected. This is apparent in the waveform displayed in the [wsprobe]. When the cut-off frequency is reduced the curves for the fricative sounds are smoother as compared to when the cut-off frequency is higher

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?

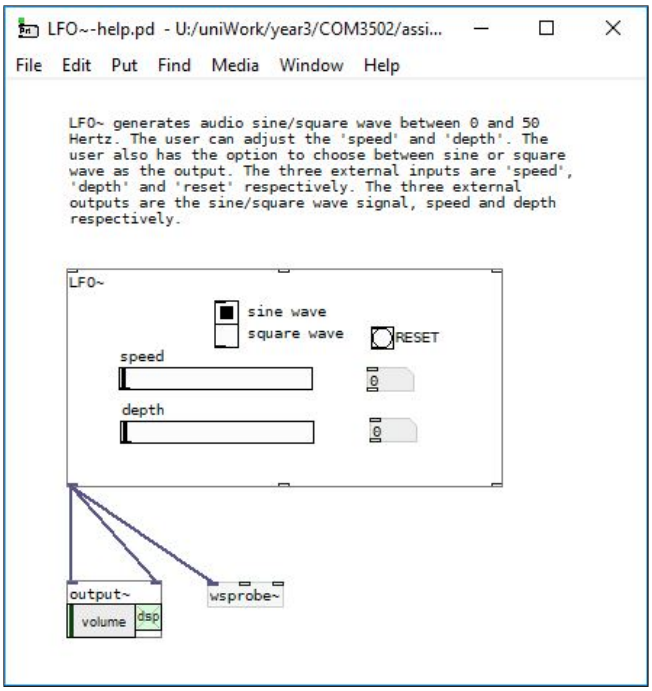
Despite the cut-off frequency being set to 10 kHz the speech is audible because some of the formants at a lower frequency range are still permitted. So we can hear the fundamentals and the subsequent harmonics with frequency less than 10kHz.

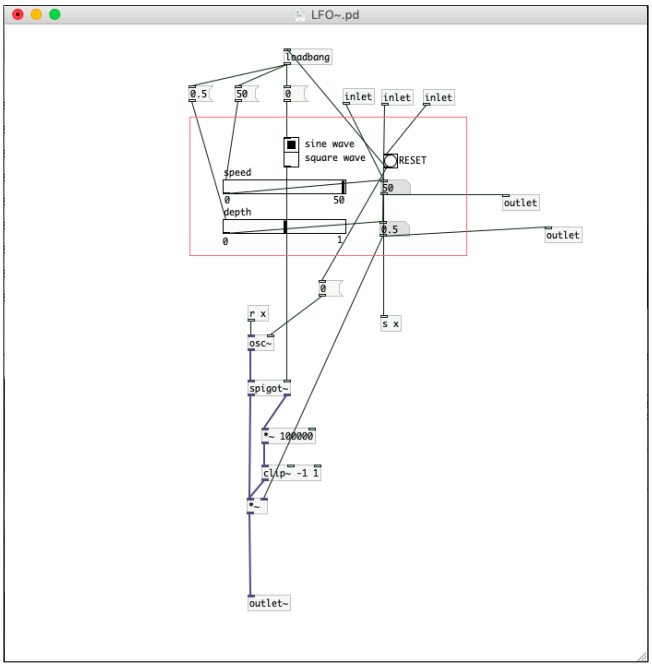
QUESTION 4 ***(worth up to 5 marks)***
COM3502-4502-6502: The `[GraphicEqualiser~]` object uses an FFT internally; what does FFT stand for and what does an FFT do?
COM4502-6502 ONLY: What is a DFT and how is it different from an FFT?

FFT stands for the fast fourier transform.Fast fourier transform is used to compute the discrete fourier transform (DFT). As compared to DFT, the FFT computes the transformations almost 100x faster because it reduces the time complexity to (n log n).

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.

The minimum and maximum anolitude of the LFO output is (with the error correction of +- 5 dB) +90 dB and -90 dB respectively. The amplitude is unaffected but he change in speed but is proprtional to the depth.





QUESTION 6 ***(worth up to 5 marks)***

In your own words[1], why is this effect known as 'ring modulation'?

Ring modulation is a signal processing function which similar to tremolo in the sense that they both modulate the amplitude. Ring modulation multiplies two input frequencies which results in a sum and difference of the frequencies in both the signals.

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

Historically due to the cyclic structure formed by diodes in the analog circuit, this was named 'ring-modulation'.

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

Despite the simplicity, the transmission of a standard amplitude modulated signal wastes a lot of transmitter power and when compared to single-sideband modulation, AM is twice the bandwidth of the audio signal to be transmitted. It is used in long-distance radio communications because as only one sideband is transmitted, resulting in a lower bandwidth which improves the signal to noise ratio by a factor of two. In addition because the transmitter power is reduced as the carrier is not transmitted it is used by rookie radio equipment.

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?

As noticed in the wsprobe when shifting the frequency down the the fundamental frequency disappears and as fundamental frequency carries a lot of information the sound signal becomes distorted when frequency is shifted down.

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

If feedback is greater than 1, the resulting gain is negative which leads to instability. When feedback is less than 1, the gain is positive but for feedback which is 1, the gain can go up to infinite values and keeps getting bigger causing the signal to become polluted.

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 a basic voice modulator. The user has to option to choose between three inputs: pre-recorded sound, live voice or upload a file. In this project several algorithms were applied to create a voice modulation patch in Pure Data. The different effects that the user can apply to the sound signal are: vibrato, flanger, feedback, delay, ring-modulation, frequency shift and vibrato. Each of these effects were achieved

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

by shifting pitch-sensitive fundamental waves of the sound signal by manipulating different components of a sound wave life the frequency, amplitude etcetera. Each of the effect can be fine tuned to user preferences and the effects can be added together. The user can also choose from a range of presets to mimic the voices of some of the pop-culture icons such a C3PO an Gollum.