University of Sheffield

COM3502-4502-6502 Speech Processing

Main Programming Assignment

Jakub Wielgorski

Hubert Baran

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.*

The waveform in voiced sounds, has repeated patterns as it is demonstrated on the Figure 1. In the spectrum it is shown as multiple harmonics.

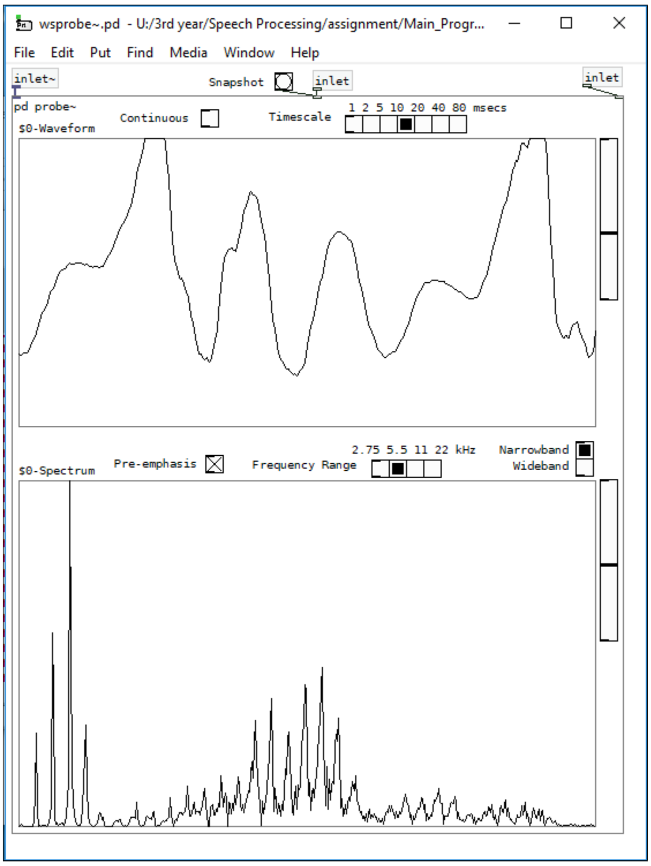

Figure 1: Waveform and spectrum for voiced sound.

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?

Sounds that are most affected when the low-pass cut-off frequency is set to around 500Hz are consonants. It is because all consonant sounds fall in the range of high frequency (certainly higher than 500 Hz) so they are all affected by the filter. Vowels on the other hand are only partially affected due to the fact that they are in the low frequency range.

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?

When the high-pass cut-off frequency is set to 10 kHz, the speech is still intelligible due to the fact that energy below the cut-off frequency is not removed completly but only reduced.

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. It is a algorithmic method that is used for computing the spectrum (for exapmle the energy at different frequencies) by performing a Fourier analysis on a time series. One of the biggest advantages of FFT is its efficiency, reducing the complexity linked to Discrete Fourier Transform and performing computations in quciker O(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.

With the speed set to 50 and the depth equal 0.5 the minimum amplitude of our LFO output is -0.5 and maximum 0.5. The change of speed does not affect the amplitude, however the change of depth has an influence on the amplitude. With the rise of the value of this parameter both maximum and minimum values rise as well, and vice versa.

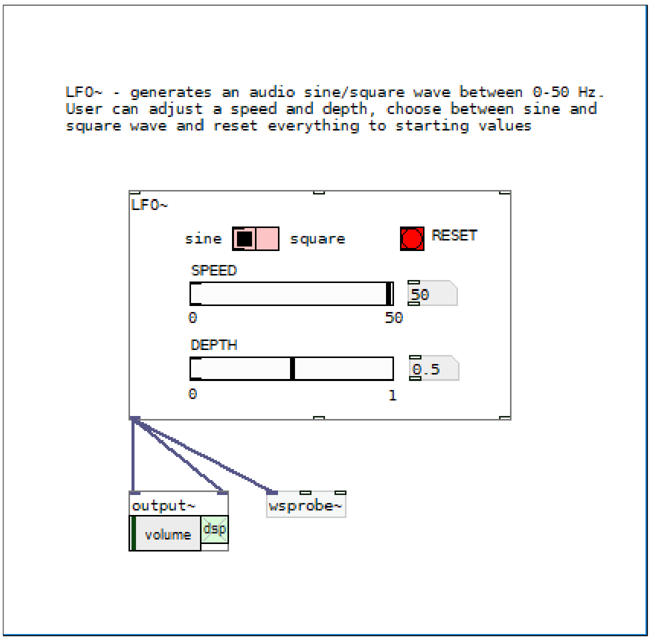

Figure 2: LFO∼-help object.

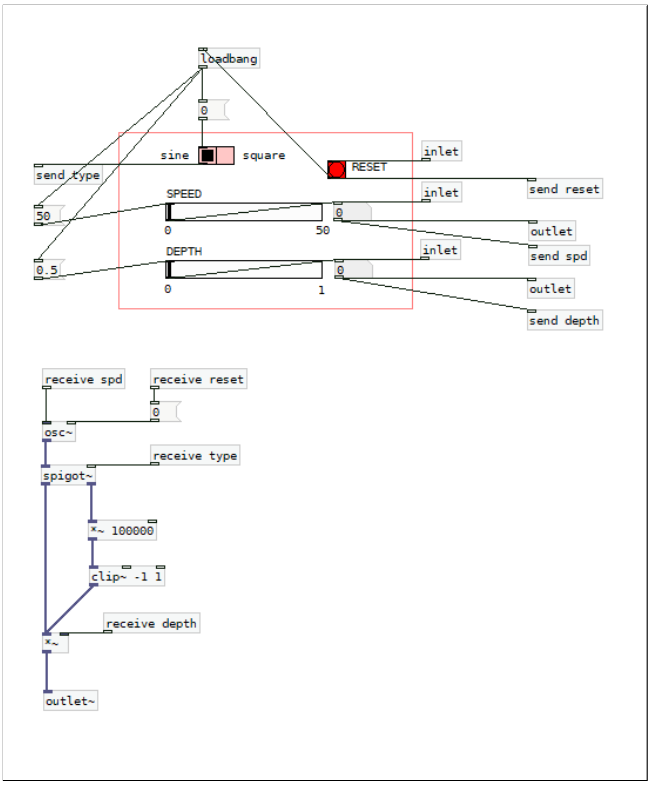

Figure 3: The internal structure of the LFO∼ object.

QUESTION 6 *(worth up to 5 marks)*

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

This effect is known as ‘ring modulation’ due to its original implementation as an analog circuit where diodes used in this implementation were connected in a cyclic structure.

QUESTION 7 *(worth up to 5 marks)*

Why is SSB commonly used in long-distance radio voice communications?

SBB is commonly used in long-distance radio voice communications because it improves the efficiency of the transmission over amplitude modulation. It does that by removing one of the sidebands as they carry the same information because they are mirror images, one generated by adding the frequency form low frequency oscillator in our example, another by computing the difference in frequency of LFO and signal. The carrier is also removed but it can be revoke in the receiver. Comparing to AM transmission, this procedure allows to decrease level of transmitter power by half and only one sideband allows for even bigger reduction. Moreover, the bandwidth of the receiver can be shortened in half, leading to an improvement of signal to noise ratio by the factor of two.

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

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? *Hint: 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 observed in wsprobe∼, shifting down the frequency causes first formant (the one which is the furthest to the left in the spectrum) which contains the most of the informations to be moved out of the natural frequency range of the human speech (the range is usually between 1 kHz and 4 kHz). This leads to the distortion. On the other shifting up the frequency causes the formant to be pushed to the right and still be in the mentioned range.

QUESTION 9 *(worth up to 5 marks)*

In a practical system, why is it important to keep the feedback gain less than 1?

The feedback sends the proportion of the output signal back to the input. For feedback gain set to low values this signal will eventually fade away. However setting feedback gain to value equal to 1 would cause to send the whole signal back to input leading to the infinite loop. In practical systems, this can easily damage the system's components.

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.

The voice changer consists of components developed in previous part of the assignment as well some additional sound effects created by our team from scratch. Firstly, after opening the program user has to choose the source of the sound. The app allows 'live' speech input in addition to the ability to select a particular pre-recorded file. Once the source is chosen, the sound wave of the input is displayed, which can indicate any modulation effect applied by user to the sound. Another way to follow the change of the sound is to use the button 'wsrobe ' which provides the waveform and the spectrum of the input sound. Each sound effect component in the graphical user interface has a switch, which can turn on and off the effect, and the 'reset' button that can set all settings to default. Each effect is controlled by appropriate to its features sliders (all properly labelled). Sound modulation effect implemented in the app are: high-pass and low-pass cut-off frequency filters, band-pass filter, reverb, vibrato, ring modulator, tremolo and mixer.

The application should be run on pd-extended form main.pd file.

[2] 25 for functionality, 15 for design/layout, 5 for `Pd` features, 5 for innovations

[3] no more than 500 words

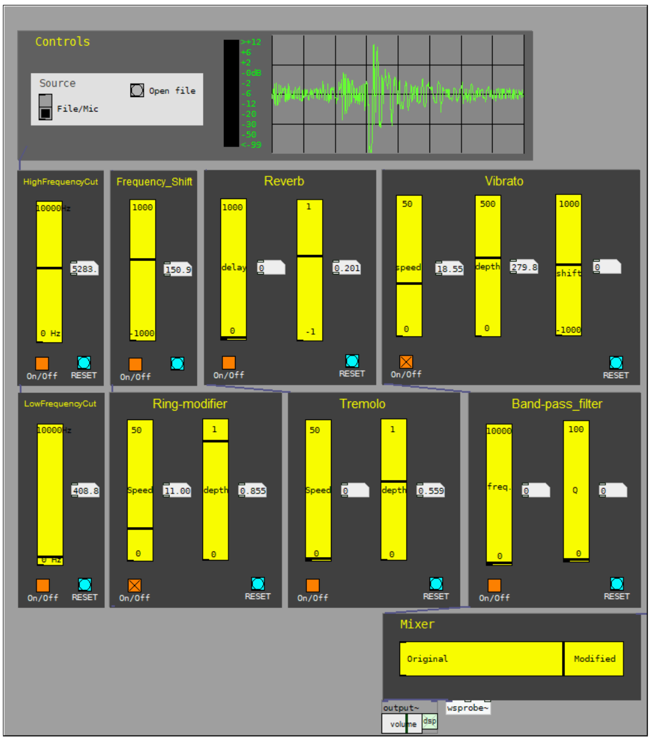

Figure 4: Waveform and spectrum for voiced sound.