

University of Sheffield

# COM3502-4502-6502

## Speech Processing



## Main Programming Assignment

John Ayad

Matthew Kinton

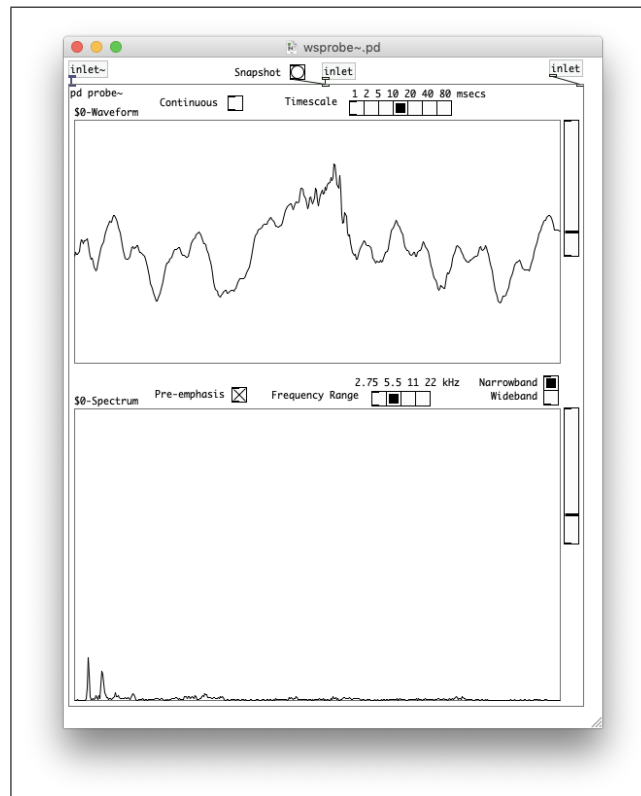
Department of Computer Science

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.*

Voiced sounds' energies are at low frequencies as seen in the spectrum. From the waveform, pulses are visible due to the glottal closure giving a more periodic waveform. On the other hand, voiceless sounds have energies at high frequencies and no pattern is visible from the waveforms.



### 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 are most affected. This is because vowels are voiced which have lower frequencies (250Hz to 1000Hz). Consonants on the other hand have higher frequencies (1500Hz to 6000Hz). The low-pass filter allows through lower frequencies (vowels) while severely damping higher frequencies (consonants). However, you could still hear dampened consonants as the filter doesn't entirely cut-off the frequencies but rather heavily dampen them at a rate of 6dB per octave.

### 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 higher the cut-off frequency, the more aggressive the dampening of the lower frequencies which results in a quieter less-noisy speech signal. A high-pass filter is mainly used for noise cancellation. It's given a roll-off frequency (Also known as the cut-off frequency). PD implements a one-pole high-pass filter with a roll-off rate of 6dB per octave.

**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?

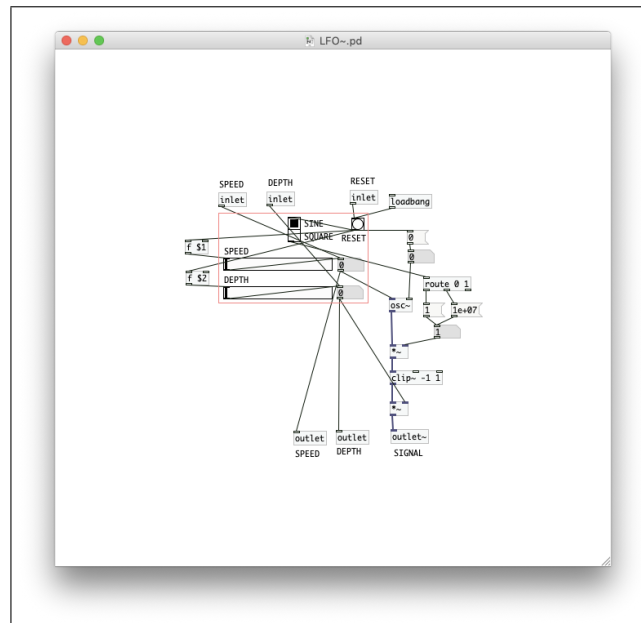
COM4502-6502 ONLY: What is a DFT and how is it different from an FFT?

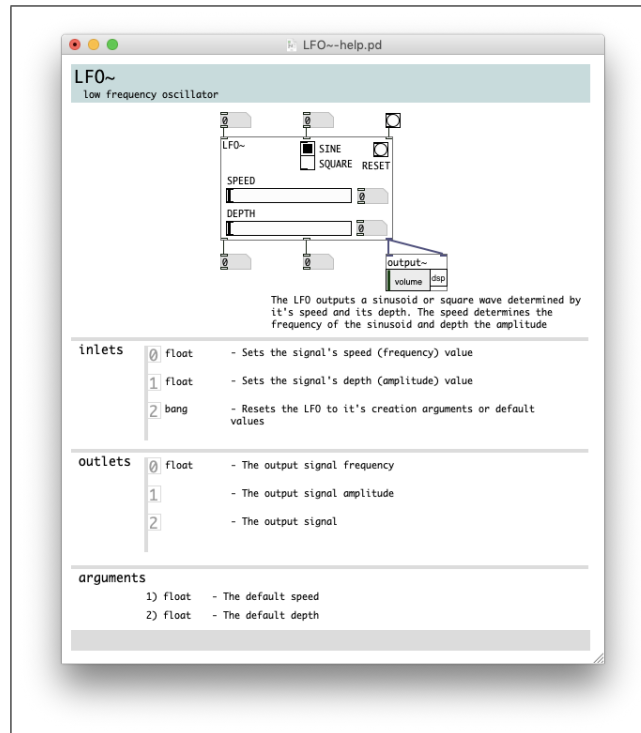
FFT stands for Fast Fourier Transform. It separates a complex sound signal into it's individual sinusoidal components each with their own amplitude and phase. FFT is an efficient variant of DFT (Discrete Fourier Transform) where it's complexity is  $n \log n$  instead of  $n^2$ .

**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 amplitude is -0.5m and the maximum is 0.5m. Varying the speed changes the frequency proportionality. Varying the depth changes the amplitude proportionality.





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

In your own words<sup>1</sup>, why is this effect known as ‘ring modulation’?

Ring modulation is a form of amplitude modulation within the audible range of frequencies. The effect is known as “Ring modulation” due to the ‘ring’ of diodes in the center of the circuit used to create this effect. The effect multiplies the two signals together which produces a signal equivalent to the sums and differences of the frequencies in both signals.

### QUESTION 7 (worth up to 5 marks)

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

Amplitude modulation on a signal results in two sidebands. One of them is higher in frequency while the other is lower in frequency. To create a SSB, one of these two sidebands (most commonly the one lower in frequency which is known as the lower sideband) along with the carrier wave are removed. This results in an improved and more efficient signal in regards to power and spectrum which is ideal for long-distance radio voice communications.

### 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

<sup>1</sup>I.e. do not plagiarise from Wikipedia.

input signal. How might it be possible to change the pitch of a voice *without* altering the formant frequencies?

The voice can be shifted up higher because the human voice ranges within lower frequencies. The human voice ranges between 85Hz to 255Hz. Shifting down the signal quickly results in an output that is outside of the human range of hearing (20Hz to 20,000Hz) whereas shifting up results in a signal that is still within the human range of hearing for much more frequencies.

### QUESTION 9 (*worth up to 5 marks*)

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

Feedback being applied is known as "Positive feedback". Such feedback results in an overall increased signal in terms of magnitude. On increasing the gain above 1, exponential growth occurs, parameters reach extreme magnitudes causes chaotic extreme behaviour. [wsprobe~] has shown that the harmonics have increased significantly while the waveform has shown a massive increase in frequency and amplitude. This is why positive feedback tends to cause system instability.

### QUESTION 10 (*worth up to 50 marks<sup>2</sup>*)

Please provide a short<sup>3</sup> description of the operation of your [VoiceChanger] application, together with a screenshot of your final GUI.

The [VoiceChanger] output is controlled by the 'MASTER VOLUME' dial. Adjusting this dial will increase and decrease the volume accordingly. Similarly, each filter and effect is controlled by their own dial. This controls the mix between the signal before the effect and after the effect. Effects are relevantly categorized and they're all connected together allowing custom combinations when used in conjunction with the effect dials. To demonstrate different settings, a number of presets have been supplied.

- Inputs
  1. Live signal (Caution: Use Headphones)
  2. File input
- Output:
  1. Processed signal
- Filters:
  1. Low-pass
  2. High-pass
  3. Band-pass
- Effects:

---

<sup>2</sup>25 for functionality, 15 for design/layout, 5 for Pd features, 5 for innovations

<sup>3</sup>no more than 200 words

1. Frequency Shifter
  2. Vibrato
  3. Tremolo
  4. Ring Modulation
  5. Delay
  6. Reverberation
  7. Flanger
- Presets:
    1. C3P0
    2. Dalek
    3. Robot
    4. Phasing (Comb filtering)
    5. Reverberant room
    6. Echo
    7. Underwater
    8. Gollum

