

COM3502-4502-6502

Speech Processing

Programming Assignment #1

(worth 25%)

Introduction

By now you should have completed the introductory Pure Data (Pd) programming exercise, and been signed-off by a post-graduate demonstrator. If not, then you must complete this step before proceeding to the following assignment.

This programming assignment is one of two for **COM3502-4502-6502 Speech Processing**. It draws on material that has been covered in Lecture #1 to Lecture #9, and it is worth **25%** of the overall course mark.

Note: *You may work in pairs on this assignment. Please team up with someone else within your COM3502 or 4502-6502 cohort. If you cannot find a partner, please contact the lab demonstrators.*

You are permitted to re-use any of the Pd examples provided during the course.

Logistics

You are free to complete this assignment in your own time. However, drop-in sessions are available in the Computer Science Lewin Laboratory (*Regent Court G12-Blue*) on Mondays from 13:00 to 13:50 during which it will be possible to gain advice and guidance from a post-graduate demonstrator.

The deadline for handing-in Programming Assignment #1 (via MOLE) is ...

Midnight Friday 10th November (week 7)

Overall Objective

The aim of this assignment is to investigate various aspects of your own voice by implementing some basic speech analysis algorithms in Pd. In particular, you are to create a Pd program to measure (and display) the means, standard deviations and distributions for the following speech parameters ...

- voiced (V) segment durations
- unvoiced (UV) segment durations
- first formant (F1) frequencies
- second formant (F2) frequencies

Note: *In addition to the above, students registered for COM4502 or COM6502 are required to estimate the length of their vocal tract.*

Specific Requirements

Your Pd program should have the following features ...

- an easy-to-use GUI
- an ability to select between processing a pre-recorded audio file or 'live' input from a microphone
- a real-time display of the means and standard deviations for (i) V segment durations, (ii) UV segment durations, (iii) F1 frequency values and (iv) F2 frequency values.
- a real-time display of four histograms showing the distributions for these four features.

The meaning of 'real-time' is that all measures should be calculated on-the-fly as the signal arrives - not after it has finished (they will, of-course, stabilise, as the utterance proceeds). Also, take care not to include in your calculations data from when the subject is not speaking. This means that the 'live' microphone input will require some method to start and stop the processing (e.g. start/stop buttons or code that implements beginning/end of utterance detection). For each histogram, it will be necessary to decide on an appropriate 'bin size'. In developing your code, you may wish to use the 'speech.wav' sample provided with the example Pd patches.

Evaluation

Once you have met the implementation requirements specified above, you should use your program to answer the following questions ...

1. How much speech (in seconds) do you need in order to obtain stable estimates of the parameters?
2. What are the average durations of voiced (V) and unvoiced (UV) segments in the utterance:
"She had your dark suit in greasy wash water all year"?
3. What are the average F1 and F2 values for the four 'cardinal' vowels [i], [a], [ɔ] and [u]?
4. **COM4502-6502 only:** Estimate the length of your vocal tract based on the formant frequencies for your neutral vowel 'schwa' [ə]. Compare the length of your vocal tract with at least five other speakers (*using an equal number of male and female voices*).

Hand-In Procedure

One member of each pair should submit a '**Zip**' file (via MOLE) containing a **short report in pdf format** and your Pd source file(s). Your zip filename should be of the form 'COM3502-PA1-RogerMoore-ac1rkm.zip' (**with your COMcode, name and id substituted for mine**).

The 'report.pdf' file should contain ...

- the names of the two team members and the proportion of effort by each (e.g. 50:50)
- the principles you used to estimate the different parameters
- the design of your code (i.e. screenshots)
- the results of your evaluation (i.e. the answers to Qs 1-3/4 above)

Assessment criteria: The marking scheme will take into account ...

- the report contents/layout – 10%
- the functionality of the Pd code (*including any discovered bugs*) – 30%
- the design/layout of the Pd code (*including comments*) – 20%
- the use of Pd features (e.g. *abstractions, graphics, help functions*) – 15%
- the results of the evaluation – 25%

Standard departmental penalties apply for late hand-in¹ and plagiarism².

Feedback (*including provisional marks*) will be provided via MOLE within three weeks of the hand-in deadline.

The deadline for handing-in Programming Assignment #1 (via MOLE) is ...

Midnight Friday 10th November (week 7)

¹ <http://www.dcs.shef.ac.uk/intranet/teaching/public/assessment/latehandin.html>

² <http://www.dcs.shef.ac.uk/intranet/teaching/public/assessment/plagiarism.html>