



Spoken Language Processing - Spring 2022

Course Project

Submission Deadline: Saturday 25/6/2022 on ITC

Project Overview:

As you know, this course requires a comprehensive term project that will constitute 20% of your course grade. Since most of these projects require an intensive workload, you have to work on this project in groups of **three** (maximum three!) students.

Project report

Final project reports are due on Saturday 25/6/2022. They should be submitted through Moodle.

Final project report should be 2-4 pages long in the IEEE transaction letter format (including appendices, figures, references, and everything else you choose to submit). The following is a suggested structure for the final report:

1. Title, Author(s)
2. Abstract: It should not be more than 300 words
3. Introduction: this section introduces your problem, and the overall plan for approaching your problem
4. Background/Related Work: This section discusses relevant literature for your project
5. Methodology (system description): This section details the framework of your project. Be specific, which means you might want to include equations, figures, plots, etc
6. Experiments and Results: This section begins with what kind of experiments you're doing, what kind of dataset(s) you're using, and what is the way you measure or evaluate your results. It then shows in details the results of your experiments. By details, we mean both quantitative evaluations (show numbers, figures, tables, etc) as well as qualitative results (show images, example results, etc).
7. Conclusion: What have you learned? Suggest future ideas.
8. References: This is absolutely necessary.

IEEE conference paper template is found on the course page at Moodle (itc.birzeit.edu).

Project Idea:

In this project, you need to develop and evaluate a recognition system of ethnic groups from speech. The common approaches of voice and speaker recognition can be successfully used to identify the ethnicity of the speaker. The 2001 census of England and Wales identifies two main ethnic groups in the city of Birmingham, UK, namely Asian and white. These groups are well represented in Voices across Birmingham, a corpus of recordings of telephone conversational speech between individuals in the city. In this project, you need to develop a system that can identify the ethnic group of the British speaker living in Birmingham city as 'Asian' or 'White' English speaker (i.e. two classes). You can use the most common feature extraction techniques in speech processing such as Energy, Zero-crossing rate, Pitch frequency and 12 Mel-Frequency Cepstrum Coefficients (MFCCs) with their deltas and delta-delta. Two or more of the machine learning techniques, such as KNN, GMM, SVM, are used to train a model for each group, which are then used to identify the speaker ethnicity as 'Asian' or 'White'.

Some good reference papers about speaker ethnicity recognition are found on the Moodle. Training and testing speech data can be also downloaded from the course page at Moodle.

Useful tools:

- Speech Filling System (SFS): <http://www.phon.ucl.ac.uk/resource/sfs/download.htm>
- Praat software: http://www.fon.hum.uva.nl/praat/download_win.html
- Voicebox Matlab toolbox: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- Netlab toolbox (it includes MATLAB implementation of Gaussian mixture Modelling, vector quantization, Neural networks, etc): <http://www.aston.ac.uk/eas/research/groups/ncrg/resources/netlab/downloads/>
- Cambridge Hidden Markov Model Toolkit (HTK): <http://htk.eng.cam.ac.uk/download.shtml>
- Kaldi toolkit: <http://kaldi-asr.org/>
- Python Google Colab: https://colab.research.google.com/?utm_source=scs-index
- Python Kaggle: <https://www.kaggle.com/code>