BIRZEIT UNIVERSITY

Faculty of Engineering and Technology

Department of Electrical and Computer Engineering

DSP Semester II 2022

Course project

Project submission deadline: 13/6/2022 23:55 on ITC
Project discussion: To be defined later

About the project:
This project must be done by teams of three students. The best arrangement is to choose a division of the project so that each of you can work on separate but interlocking parts. Teams of two or individual work will not be accepted.

Learning teamwork is also one of the more general goals of this course, so team projects will pick up points for demonstrating a successful ability to work with others.

The projects will be graded based on a project report (of around 3-4 pages) as well as in-class short presentations or discussion in my office.

Project submission must be via Moodle only, but please use PDF format and **not** Word .DOC files if at all possible, since I often have formatting problems with Word files.

Your report must have the following structure, using these section headings and using IEEE paper format [template can be found on Moodle]:

Introduction: A general description of the area of your project and why you're doing it.

Problem Specification: A clear technical description of the problem you're addressing. Formulating a general problem (e.g., transcribing music) into a well-defined technical goal (e.g., reporting a list of estimated fundamental periods at each time frame) is often the most important part of a project.

Data: What are the real-world and/or synthetic signals you are going to use to develop and evaluate your work?

Evaluation Criteria: How are you going to measure how well your project performs? The best criteria are objective, quantitative, and discriminatory. You want to be able to demonstrate and measure improvements in your system.

Approach: A description of how you went about trying to solve the problem. Sometimes you can make a nice project by contrasting two or more different approaches.

Results and Analysis: What happened when you evaluated your system using the data and criteria introduced above? What were the principal shortfalls? (This may require you to choose or synthesize data that will reveal these shortcomings.) Your analysis of what happened is one of the most important opportunities to display your command of signal processing concepts.

Development: If possible, you will come up with ideas about how to improve the shortcomings

identified in the previous section, and then implement and evaluate them. Did they, in fact, help? Were there unexpected side-effects?

Conclusions: What did you learn from doing the project? What did you demonstrate about how to solve your problem?

References: Complete list of sources you used in completing your project, with explanations of what you got from each.

The reason for this somewhat arbitrary structure is simply to help you avoid some of the more problematic weaknesses I've seen in past years. If you're having trouble fitting your work into these sections, you should probably think more carefully about your project.

Project description:

Gender Recognition of a speaker has various potential applications such as speaker identification for securing access to confidential information or virtual spaces. As a human it is easy to differentiate between male and female voices. In this project, you need to use a technique for identification of males and females (maybe further adults and children) from their speech analysis. This analysis basically aims to predict the gender of the speaker by analyzing different parameters of the voice sample. It includes simple short-time pitch frequency (or fundamental frequency F0) which can be estimated from the short frames (20-30ms) using auto-correlation method. The fundamental frequency represents the vibration rate of the vocal folds of the speaker, which is usually high for the children, female speakers, and relativity low for the male speakers.

$$R(k) = \frac{1}{N}\sum_{n=0}^{N-1} s[n]s[n-k]$$

Where, R(k) is the autocorrelation function at k. s[n] is the short frame samples with length N samples.

Clearly, R(k=0) would be equal to the average energy of the signal s[n] over the N sample frame.
If s[n] is perfectly periodic with a period of P samples, then s[n+P] = s[n]. Therfore, R(k=P)=R(k=0) = average Energy. This is not exactly true for speech signals, the autocorrelation function with k equal to the pitch would result in a large value. For various k values between 0 and P, the various terms (s[n]s[n-k]) in the autocorrelation function would tend to be mixture of positive and negative values. These would tend to cancel each other out in the autocorrelation sum to yield very low values for R(k).

This, for a given frame of N samples of Voiced speech, a plot of R(k) versus k would exhibit distinct peaks at k values of 0, P, 2P, 3P, … where P is the pitch period. The graph of R(k) would be of quite small values between these peaks. This pitch period for that frame is simply got by measuring the distance, in samples, between the peaks of the graph of the autocorrelation function.

Speech Dataset:
For this project, you need to have sufficient speech recordings for different male speakers and different female speakers. You can use any available speech dataset that meets this condition or you can collect your own dataset by recording for number of male and female speakers. Make sure that all audio files you will use, or you will record, have the same sampling frequency Fs.

Procedure:
Read an audio file from any database that you have or record your own voice.
Find its sampling frequency (Fs).
Divide the samples into short frames, e.g. 20ms with overlap 10ms.
Find autocorrelation values of each short frame, for all k values k:0→N-1.
Find the first max peak of the autocorrelation values, which is corresponding to the pitch period (in samples).

Convert the estimated pitch period P into seconds and then find the corresponding pitch frequency F0. If F0 is above a specific threshold then the voice is of a male, otherwise it is of a female. If you get incorrect results, try changing the threshold values. Threshold frequency (Fth) of a male voice lies between 85-155 Hz whereas for a female, it lies between 165 to 255 Hz.

Project deliverables by each group:

1- Mini-report as described above in IEEE template.

2- System demonstration of each part as described above.

You can use MATLAB, or Python because they have many useful functions.

Alhareth Zyoud
May. 30th 2022