Meeting 5 with Prof Rubin

13 October 2017, 10:30

Bianca and Lindzi

- Change of title to "Voice conversion of a synthetically generated voice" is fine.
- Decided to take two different approaches to try map the voice and transform between the LSF values from the robot to the human voices
    - Linear Regression
    - Gaussian mixture model
- Linear regression
    - LPC coefficients model the vocal filter, and the excitation signal models the signal from the vocal cords. By changing the vocal filter, it will modulate the robotic excitation signal.
    - Difficult to transform the excitation signal, so rather will look at the LPC coefficients which can be translated more easily.
    - Played the resultant audio from linear regression audio
        - Trained using 9 voices. There's a lot of pre-processing that needs to be done.
        - The transformed voice sounds more human according to Prof. Rubin.
            - However the cadence/prosody is what makes all the difference.
- On open day we want to get some subjective feedback on the audio results.
    - Will have to put in a letter as an amendment to our ethics if we want to put those results in our reports.
    - Draft a letter and Prof Rubin will speak to the chairman about it.
    - Will be nice to have that data in the report.
- Poster for open day
    - Looked at some good examples.
    - It's a visual thing, you want more graphs/figures than words.
    - There's a general format you use, introduction, body, etc conclusion, references.
- The same principle of minimal words applies to the presentation.
    - The presentation is done together. Roughly split the time equally.
- Looking at the FFT of the original source, target and converted audio from linear regression.
    - There are some peaks added in the transformed one that kind of follow the human FFT.
    - Prof. Rubin suggested we look at the bode plots of the audio and see which frequencies need to be changed/ramped up or pushed down.
    - So try convert the bode plot to look similar.
    - We are on the way to getting the results we need.
    - Now we need to work on refining that solution
- At the next meeting will listen to the results of the GMM system.
- Having two mapping methods will be great to have the comparison and put the content in the result.

Peter and Joyce

- Graph of the MFCC results of the old and young voice, looking at what is changing with age, just to see where the biggest differences are.
    - Trying to map the changes between the two.
- Struggling with the statistics to model between the old and the young values.
    - Plotting the residuals to see the difference between the points.

- Want to see which features are normally distributed and then can do log distributions.
- Trying to think of a statistical method that they can use to map this.
    - Ideally using a machine learning algorithm would be great because you just give it the coefficients and it finds the function for you.
    - So you have a training and a testing set.
    - The problem with this is that it would only classify a voice as old or young, it wouldn't help you transform from one to the other.
- Lindzi is using the GMM approach to do something similar to what Peter and Joyce are trying to do so perhaps they can look into that.
    - Or perhaps look into other forms on non-linear regression that would maybe be able to map the coefficients sufficiently for you.
    - Prof Rubin really thinks machine learning would be great for this application.
- Joyce has done the re-synthesis of voice using the MFCCs.
    - Once they have the transformation of the MFCC coefficients this can be used to re-synthesise the transformed voice.

Next meeting at 12:00 on Monday

Prof Rubin will get in touch with the chairman for our ethics – email him to remind him.

At the end of the day both groups have proper lab projects, you've done the evaluation and assessment that's necessary.