

EE513 Project 4: Whisper and Pitch Transformations

Matt Ruffner
matthew.ruffner@uky.edu

ABSTRACT

In this project, recordings of speech were processed in an attempt to replace the voiced speech (that in which the vocal folds are active) with unvoiced speech (a whisper). Linear Predictive Coefficients (LPCs) were employed to determine if the a particular segment is voiced. White noise was used to emulate the whisper, colored by the pitch of the spoken voice it was replacing.

1. APPROACH

For this project, 40ms of audio were read in at a time with a 50% overlap with the previous segment. The overlap allowed for lossless reconstruction when a triangular tapering window was applied to minimize transition noise.

Main program execution occurs in the file `p4.m`, with auxiliary helper functions in the files `isvoiced.m` and `whisperfy.m`.

For every 40ms segment of speech, the LPCs are found and passed to one if not both of the auxiliary functions. If the first formant frequency is below 2kHz, it is considered to be a segment of voiced speech. This is a very crude method and could be improved upon greatly. It preforms relatively well, if overly ambitious, on both male and female test recordings.

If the segment is considered to be voiced by the previous judgment, the second function, `whisperfy.m` is called. This function generates white noise the same length as the segment, filtering it with the LPCs to modulate it at that particular pitch.

Finally, the created whisper segment is tapered so it will add in to the output signal with minimal 'popping' where signal discontinuities occur.

2. RESULTS

The program was tested on `man1_take1.wav` and `woman1_take1.wav`. The voiced speech decision algorithm is overly ambitious at determining a segment is voiced, however the level of intelligibility is not overly affected.

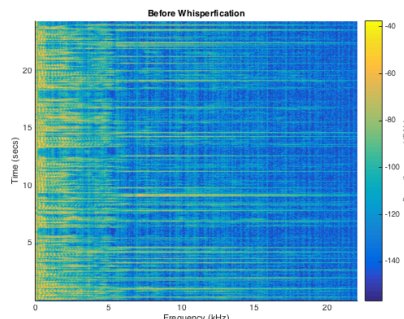


Figure 1: Woman recording take 1 before modifications.

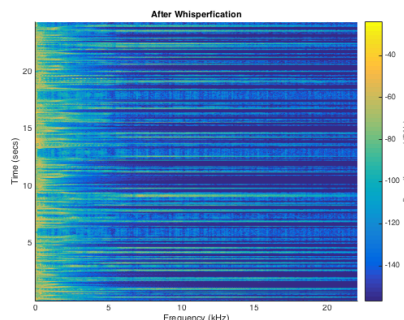


Figure 2: Woman recording take 1 after whisperification.

Spectrograms of before and after are shown in Figures 1 and 2. Sections where the threshold was crossed rapidly are darkened on the after spectrogram, my guess is when normalizing the colored noise there was low signal energy to begin with, resulting in somewhat of a dead space.

3. CONCLUSIONS

A more vigorous voiced speech determination system as well as whisper generator would have created a better whisper synthesizer, however I must say I am impressed at the level of intelligibility that remains after the voiced speech has been so heavily modified.

Also, individual testing mechanisms for each part outlined in the project assignment were not left in the code, as functionality was verified along the development path.