Speech Processing Project 2 Writeup

Matt Ruffner

March 9, 2019

This project explored dynamic time warping. The algorithm was implemented in MATLAB. Figures code are in the adjoining PDF which was published from MATLAB for conveinence and readability. THis writeup will discuss spefic results and analyses of the project.

Weighting the frame to frame distance with the overall covariance matrix of the entire data set improved recognition performance by 2%. Using the built in DTW function yields an accuracy of 84%. Speaker 3 (GENERAL3 labeled data) was categorized correctly most often with a specific accuracy of 82%. Speaker 1 (GENERAL1 labeled data) was the over all least often correctly categorized label with a specific accuracy of 59%.

Normal and Fast recordings were most often correctly categorized with accuracy of 76% and 73%, respectively. Soft recordings were the least often correctly categorized enunciation, with an accuracy of 60%

Destination and zero were the most often correctly categorized words, with destination being slightly higher in accuracy. The two words which were miscategorized as each other most often were eight and eighty, with fast and soft speaking voices yielding the most confusion.

Paths were calculated and plotted for two example categorizations against the correct template. A fast and slow destination were compared along with a fast and slow zero. The paths are plotted against a straight line to emphasize deviations from a linear scaling.