Voice Control System

Conner Howell, Zhen Liu, Jia Zhang, Ziyang Zhang, Yue Zhou

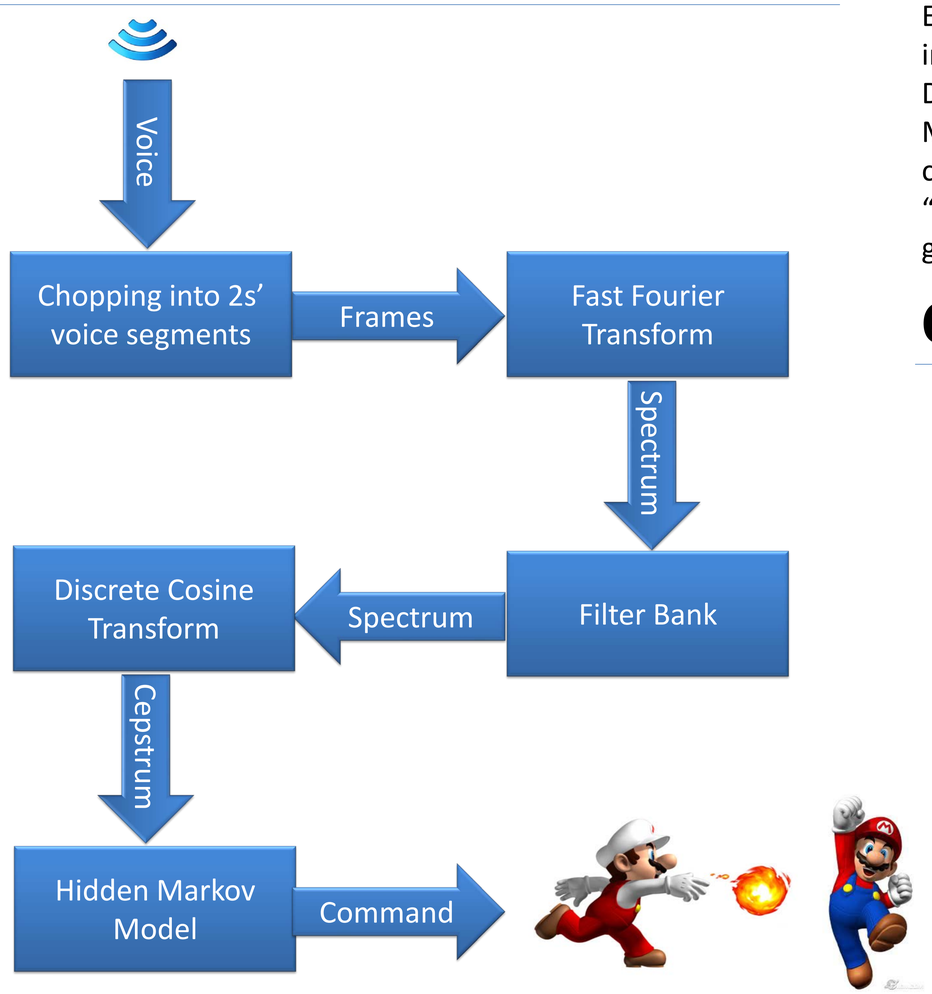
EECS 452 – Winter 2011 Design Expo

Overview

Our Voice Control System can be used to recognize a set of learned voice commands from different speakers. In this system, speakers can use the six basic voice commands: go, back, stop, jump and fireball, to control Super Mario Bro on Wii.

Our project parses the audio signal by perform MFCC (Mel-Frequency Cepstrum Coefficients) analysis on segmented audio waveform and making decisions on which command is being said based on the work of Hidden Markov Model (HMM). The whole algorithm is implemented on the C5510 DSK; control and interface with Wii is done by Spartan-3. During the execution of recognizing an operator's voice, the system chops the audio signal into frames, extracts MFCC, inputs the characteristics vector into HMM and determines the voice commands using forward algorithm.

System Diagram

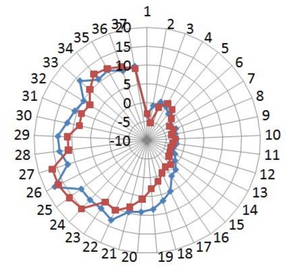


Voice Command Recognition

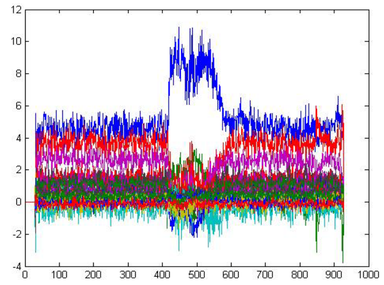
We applied MFCC (Mel-frequency cepstral coefficient) model for voice feature extraction, which is widely used in speech detection and recognition. The mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound. And Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. The difference between the cepstrum and the mel-frequency cepstrum is that MFCC approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum.

Power Cepstrum of Signal = $\mathcal{F}\left\{\log(|\mathcal{F}\{f(t)\}|^2)\right\}|^2$

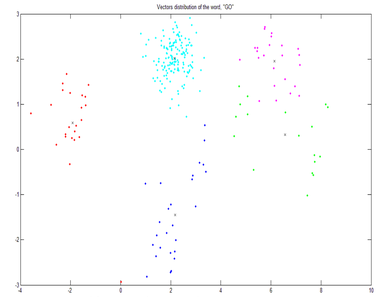
Mel-Scale Frequency Conversion $m = 2595\log_{10}\left(\frac{f}{700}+1\right)$



Similarity measurement

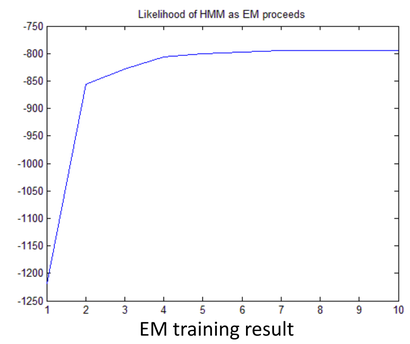


MFC coefficients



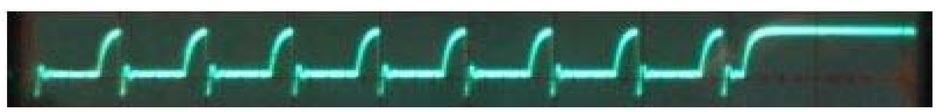
Clustering using K-mean

HMM is trained before being tested. We constructed several HMMs, each corresponding to a command word. During the training of HMM, we adopt Expectation-Maximization (EM) algorithm to optimize parameters in HMM to increase the likelihood of the certain HMM given the corresponding command. Due to the complex computation of EM, HMM training session is performed using MATLAB beforehand. The construction of HMM: Each HMM has 5 states each containing one Gaussian Distribution. K-means algorithm is adopted to get a "educated guess" for HMM parameters such as mean, variance, weight when given set of vectors.



EM training result

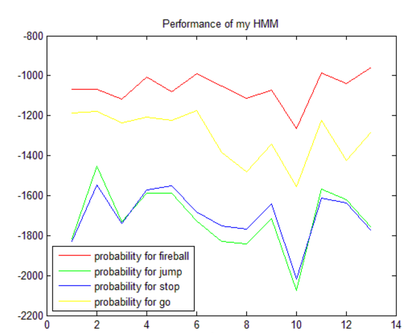
Control and Interface with Wii



Typical Serial Signal when waiting for instruction

The Mimicked Signal When A,B is being pressed

LEDs on S3SB (Spartan-3 Starter Board) is used to give information to speaker. By switching the switches, one person's set of HMM parameters is being chosen. The Wii's protocol with controller is mimicked using Spartan-3. Patterns of serial signal, each corresponding to a certain command are being generated when the voice decision is made.



HMM performance