Description for Final Project

Team member(s)

Yicong Liu

What is the objective of the project?

To recognize a single spoken number from zero to nine. To be specific, as one speaks a number (0 - 9), the program will recognize the correct number. If possible, the project can be extended to the recognition of multiple continuous spoken numbers.

Where are you getting the data?

I get my training data from a database called FSDD, including 1501 audio files (https://github.com/Jakobovski/free-spoken-digit-dataset/tree/master/recordings). I record the voice of my own or my friends' as the test data.

Do you have ideas of how you may solve this?

This is a classification problem. I would use Mel-frequency cepstral coefficients (MFCCs) as features. At training stage, all the MFCCs of audios in the training set are calculated, and split to ten categories. At test stage, the input audio will be classified to one of the categories, and the recognition is complete.

Have you seen other attempts at this problem or related problems?

Yes, voice recognition is very common. The work I'm doing is to implement machine models or algorithms I've learned or known to solve the problem.

Do you have any other example code to start with from other groups?

No, for now. I will record and cite all the example codes during the project.