**CPE646 Pattern Recognition Project proposal**

**Topic:**

recognition and classification of human emotions from speech.

**Team member:**

Yuan Tian (10422134)

Xin Li (10421856)

Linqiao Shang (10421310)

**Overview:**

Design several models to analyze human voice samples and classify to corresponding emotions.

**Model design & methodology**

Part I. voice samples collection

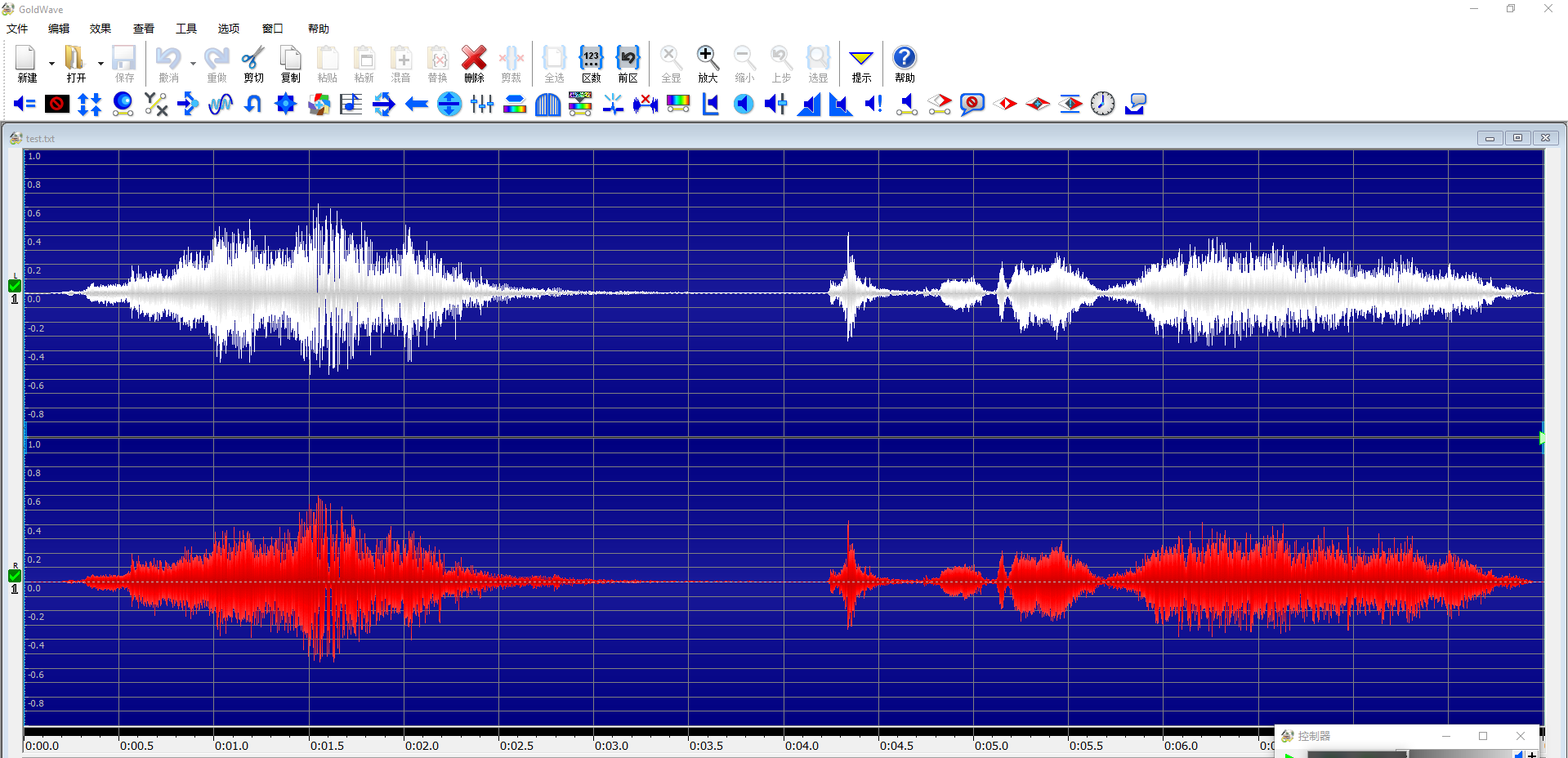
The team assumes human emotion has five categories: neutral, happy, anger, sadness, fear. In that case, for each type, this project uses 10 data. Totally the size of training data set is 50(.mp3). The ideal proportion of male and female is 1:1. Language is English. 

As to neutral emotion: training samples come from News reporters.

As to other four emotions’ data: training data is voice samples that abstracted from YouTube video, with corresponding tags.

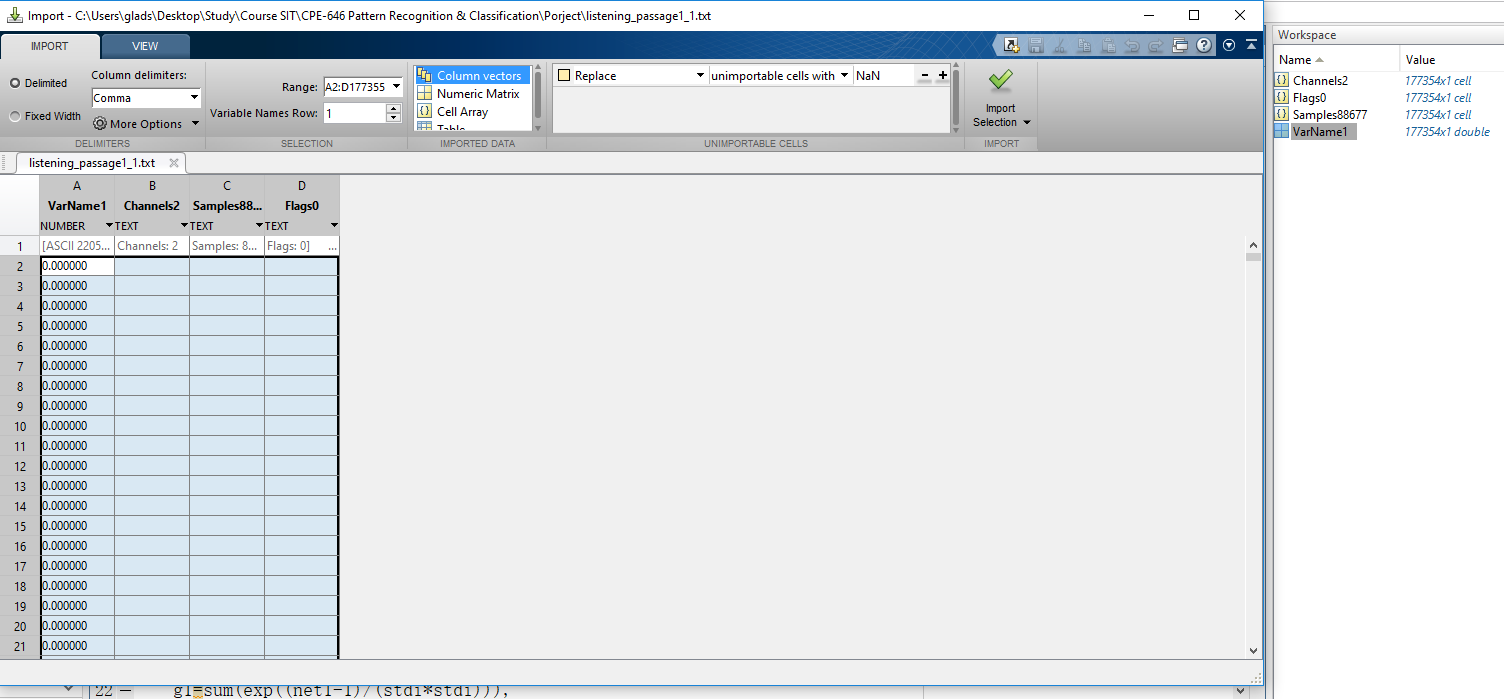
**Part II. Signal conditioning and conversion.**

First convert voice sample (.mp3) into digital signal. (using goldware)

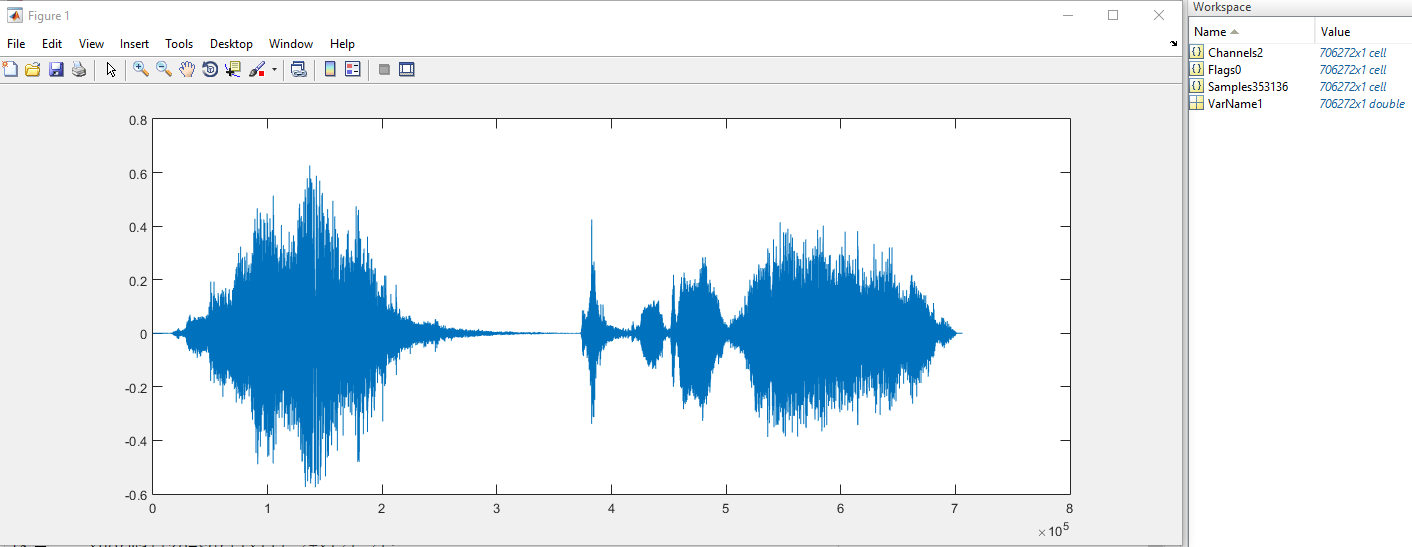


Then covert (.mp3) file into (.txt)file and import data into MATLAB.

https://image.prntscr.com/image/r1On8CS_QfKW9yBaQPaEiQ.png



Therefore, the digital data can be preserved and used as training and test samples in MATLAB.



Distinguish the and separate those signals. Use the following eigenvalues as input of our model.

Eigenvalues are

1. Maximum, Minimum, Mean, Median and Variance of:

* Short-term energy and differential signal
* The 1st, 2nd order of pitches’ derivative
* Range of Pitch
* Derivative of 1st, 2nd, 3rd formants

1. Ratios

* Harmonic noise ratio(HNR)
* Silence, speech region numbers.
* Frames number of silence, speech.
* Spectral energy distribution

**Part III. Classification and Model Design**

Recently there are some popular methods that used in Emotion Recognition:

* Decision tree
* SVM
* Random forest

“From the evaluations, it is concluded that the proposed DC based emotions recognition scheme achieves high accuracy with RF classifier at each stage.” 3

Also,

* RNN

LSTM/ BLSTM network (By one-label and CTC approach)4

The project will try to repeat the experiment result of those two papers’ model and make a comparison between Random Forest(the optimal method and modified LSTM/BLSTM methods).

**References:**

1. Huang Cheng-wei, Zhao Yan, Jin Yun, Yu Yin-hua and Zhao Li. A Study on Feature Analysis and Recognition of Practical Speech Emotion. Journal of Electronics and Information Technology, China, 2008, 33(1), 112-116.
2. Lin Yi-lin, Wei Gang. Speech Emotion Recognition Based on HMM and SVM. Proceeding of the Fourth International Conference on Machine Learning and Cybernetics , Guangzhou, China, 2005, August, 18-21.
3. Badshah, A. M., Ahmad, J., Lee, M. Y., & Baik, S. W. (2016). Divide-and-Conquer based Ensemble to Spot Emotions in Speech using MFCC and Random Forest. *Divide-and-Conquer based Ensemble to Spot Emotions in Speech* using *MFCC and Random Forest,* 4-5. Retrieved November 16, 2017, from https://arxiv.org/abs/1610.01382.
4. Chernykh, V., Sterling, G., & Prihodko, P. (2017). Emotion Recognition From Speech With Recurrent Neural Networks. *Emotion Recognition From Speech With Recurrent Neural Networks*. Retrieved November 16, 2017, from <https://arxiv.org/abs/1701.08071>.
5. The Interactive Emotional Dyadic Motion Capture (IEMOCAP) Database [Advertisement]. (2004). Retrieved December 10, 2017, from http://sail.usc.edu/iemocap/