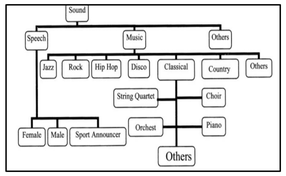
Music and Speech Discrimination

Afagh Farhadi, Jiayi Ren, Thomas Culeton
University of Rochester, New York

Introduction

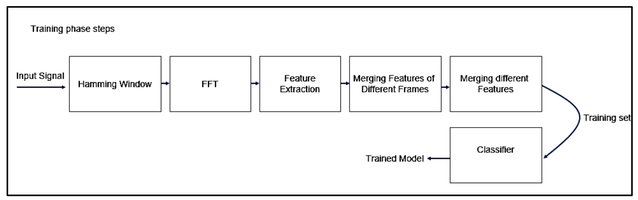
- Discrimination between speech and music has been a challenge in the field of audio signal processing.
- Using a variety of characteristics, input signal will be fed to MATLAB program and it will predict which class of sound the sample recording is.

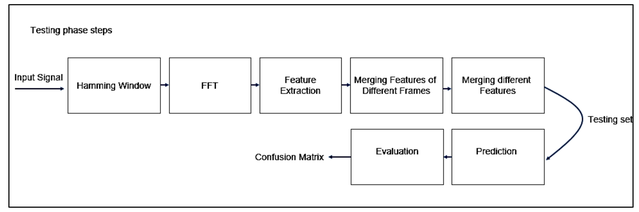


Applications:

- Cocktail party problem: difficulty realizing speech in noise and music background.
- Multimedia domain and automatic speech recognition (ASR).
- Audio Coding of the signals :different efficient methods for encoding music or speech.

Method





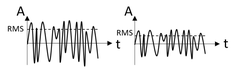
Dataset

- GTZAN dataset is used for training and testing.
- Half of the data is for training and the other half for testing
- No overlap between training and testing data

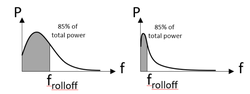
	Training Data Set	Testing Data Set
Number of the files	32	32
Duration	30 s	30 s
Sampling Rate	22 KHz	22 KHz

Feature Extraction

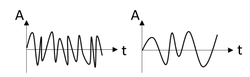
1- Root Mean Square (RMS) : $\mathrm{RMS} \triangleq \sqrt{\sum_{n=1}^{N} x^2(n)}$



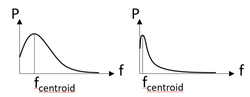
2- Spectrum roll-off: Frequency that most signal power is below

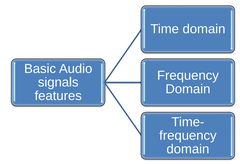


3- Zero Crossing Rate: How often a signal crosses the x-axis

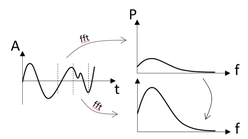


4- Spectral Centroid: "Center of Frequency" where power is concentrated



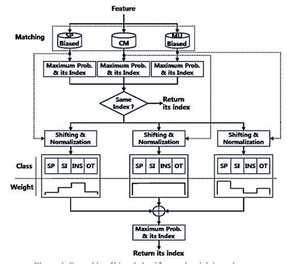


5- Spectral Flux: Spectral differences between frames



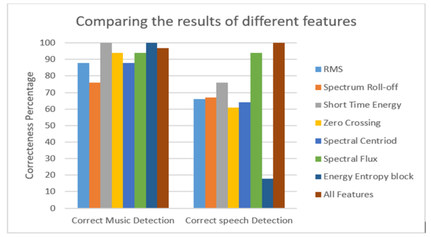
6- Short Time Energy (STE) : In speech there are voiced and unvoiced segments of the signal in which the amplitude changes to lower values for unvoiced segments and increases in voiced segments. However for music there is no unvoiced segment so the STE is larger in music signal than speech.

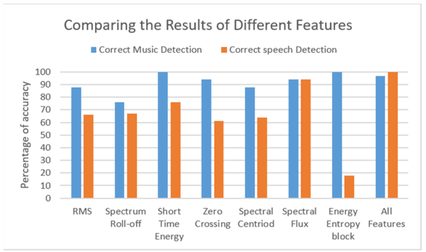
Classifier



- Building off the Gaussian Mixed Model classifier from assignment 5
- Classified by indexing (labelling music or speech) the data chunks that are extracted
- Training separate classifiers with different chunk sizes
- Take a majority vote of multiple classifiers to determine classification of samples

Results





Live Testing

Our trained model is capable of taking in live audio and processing it, classifying it into its proper class.

Conclusions and Future Work

As it stands currently, our project accurately discriminates between speech and music samples. It rarely mistakes a music sample for a speech sample, and always classifies a speech samples correctly. While some characteristics, such as zero-crossing and spectrum roll-off, are less accurate than others, the combination of all these characteristics provides a highly accurate way of discrimination between music and speech samples. We can also see that some characteristics are more biased towards one class, so combining all the features grants the best results, through its use of a larger pool of information. The next logical step for this project would be to expand it to discriminate between additional types of data, such as singing. Another step could be to perform source separation for samples containing multiple classes. We could also improve our model by training and testing it with samples containing a higher noise floor.

References

- Carey, Michael J, et al. A Comparison of Features for Speech, Music Discrimination. Ensigma Ltd., 1999, pp. 149–152, A Comparison of Features for Speech, Music Discrimination.
- "Zero Crossing Rate." Zero Crossing Rate - an Overview | ScienceDirect Topics, www.sciencedirect.com/topics/engineering/zero-crossing-rate.
- Kim, Kibeom, et al. *Speech Music Discrimination Using an Ensemble of Biased Classifiers*. Audio Engineering Society, 2015.
- Al-Shoshan, Abdullah I. "Speech and music classification and separation: a review." Journal of King Saud University-Engineering Sciences 19.1 (2006): 95-132.
- Some Basic Audio Features, Theodoros Giannakopoulos, https://www.mathworks.com/matlabcentral/fileexchange/19236-some-basic-audio-features
- Scheirer, Eric, and Malcolm Slaney. "Construction and evaluation of a robust multifeature speech/music discriminator." 1997 IEEE international conference on acoustics, speech, and signal processing. Vol. 2. IEEE, 1997.
- El-Maleh, Khaled, et al. "Speech/music discrimination for multimedia applications." 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100). Vol. 4. IEEE, 2000.