CSE497-Engineering Project I

PROJECT SPESIFICATION DOCUMENT

DETERMINING OF MUSIC GENRES

ADVISOR:
Mehmet BARAN, ASSOC.PROF.

MEMBERS:
ÇAĞATAY DEMİREL / 150110054
ENES AYTEKİN / 150110010

1.Project Statement

In our project, we desire to analyze the voice. If we will be able to analyze voice, we desire to find genre of musics. When user gives voice file like .wav file, our program will be able to analyze to find its genre. If .wav file will be music as well. We will use machine learning algorithms like Hidden Markov Model to estimate and train real data. We will take same kind of inputs as much as we can and we will train them using Hidden Markov Model to find approximate model for individual kind of voice. It can be rock, rap, classical fonetics. Also it can be human voice, nature voice. Training data are can be increased and there is no limit in our project.

2. Problem Description

This project will provide the determining music genre by genre. For building this program, we will use some tools like HTK [1] with hidden markov models. In our project, HMM is the most important part. It gives us an ambition in fact it is the reason for doing this project.

With this project, we want to seperate genres one by one and understanding relationship between them. After understanding relantionship, program can be develop yourself. This will give us what people like and listen.

The most important problem is there was a limited source to develop this program. We can develop this project at some points, but after that we are planing to analyze, classify and get features from them by ourselves. As we mentioned before for choosing this project, we are intended to interest machine learning and artifical intelligence. These reasons effect on positively and we will success on this project and solve these problems.

3. Aims of the Project

There are many aims of our project and every one of them are vital. That are:

- **Understand probability from given statistics :** Its our first aim. Because, nature has too many observable inputs and outputs. Thats mean, every object has a purpose and every object effect another for purpose. If these effects will be hold as statistic, next probabilty of that effect can be estimated. And this project will give us to power of estimation from given real statistic.

- **Understand voice in nature :** Voice is a signal in nature. It moves as wave and it has weight, speed, frequency and energy. Micprophones receive them and convert continuous signals into discrete signals. Discrete means matrices that computer will understand. We will learn these intels as deeper. There are very detailed information about nature of voices [2].

- **Learning Hidden Markov Model :** Our next aim is learning hidden markov model. Because, it is the best way to use given statistic for finding probability of new observation. Also, as we researched, hidden markov model is the best way to train voice data. Voice data may be very complicated, because sound will be converted matrice that computer can understand. And these kind of matrices are very huge matrices, especially if you consider about music data, it may be giant matrices and values of them are numeric numbers. We may use gaussian mixture as observations and markov model can train gaussian mixture observations. More intels about Hidden Markov Models can be found [3].

- **Applying Hidden Markov Model to analysis of music or fonetics of human :** Wheter music of fonetics like natural language voice of humans, we will apply data to hidden markov model. Spesifically, individual data will be trained. After this training hidden markov model for individual fonetics or musical enstruments will be approximated. After this approximation, we will be able to use hidden markov models for finding genre of music or meaning of human fonetics or we may succeed about copy of human voice. At most, we desire to create virtual voice of real person. It may be our future work or the real work.

4. Related Work

Related works can be classified these categories: speech recognition, speaker recognition, music vs. speech classification, beat tracking and rhythm detection, automatic music transcription, auditory scene analysis.

Speech Recognition: Speech recognition is perhaps the most fundamental audio classification problem: giving a computer the ability to analyse and understand speech. This task is generally difficult due to the large number of ambiguities in spoken language. It only seems easy to humans because we have years of practice in understanding phonetically identical utterances by deriving clues from environmental context, information about the speaker and subject, etc. Johnathan Foote discusses speech recognition as a specific audio classification problem in [4].

Speaker recognition: Speaker recognition is of special interest for security applications where access rights are granted by analysing a voice sample. This is used to determine whether the speaker belongs to a list of authorised people.

Music vs. Speech Classification: Speech recognition systems work well on speech input, but do not provide usable results when music is fed into the system. Likewise, it does not make much sense to try and determine the music genre of a phone conversation. The ability to distinguish between music and speech is important as a front-end to a generic sound-analysis system that is able to process real-world input data, passing the signals on to the specific back-end audio signal classification program that can handle them appropriately. Recent work in that area includes [5] and [6].

Beat Tracking and Rhythm Detection: Tapping their foot along with a musical performance is an easy thing to do even for non-musicians, but has been found to be a challenging problem for automatic systems. Rhythm seems to be one of the key elements of musical information. Beat tracking systems are therefore an important part in music genre recognition. An introduction into the even more complex subject of musical rhythm can be found in [7].

Auditory Scene Analysis: At any given moment, we are surrounded by sound that generally contains more than just one auditory event. The physical signal that hits our ears does not contain any information on the individual parts that have contributed to it, but still we are able to distinguish and separate between sound events. A prediction-driven approach to ASA that is better able to cope with complex sound scenes is presented in [8].

5. Scope of the Project

Our scope is cannot be set strictly. Because, we will use hidden markov model to training data and searching from models. While, we creating hidden markov model, we will use real data from nature. Thats are like many kind of music, human fonetics(that may change for person, natural language, accent and even some emotional issues).

We will use real signals and computer receive these signals as physically. Physically means, signals transport in air and they are moving as waving. A signal has frequency, energy, weight and speed. It gives pressure to microphone, and micprophone takes difference between air pressure and signal pressure. Than it sends discrete signals to sound card.

Finally, computer has converted continuous signal into discrete signal. It means, it comes from reality. And we will approximate the reality. We desired scope is to approximate as best, spesific kind of voice. We wont be able to analyze all kind of music. Because, there are no strict limits between some genres. Some music enstruments and voices of them may will be very similar them. We will be able to find that has more strict harmonics ones.

Our final approach for this project is understand the signal of nature and understand to analyze it in computer. And if we will be able to analyze right, we will use them for individual observations.

6. Success Factors and Benefits

We can measure the achivement level of the project many ways.
Getting features and classify them the most important area in our project. We are planning to develop these algorithms and getting better results compare to early projects. For example, we assume that the genre of rock boundaries going to be more wide area, because of subset of rock. We can say the more fixed boundaries the more success.

The indicator is time for our project. The time is important because, the program should notice the changes between subsequences rapidly. If our project provides this ability, we can say it recognizes the changes quickly and with minimum execution time.We should obtain the sequences with converge time.

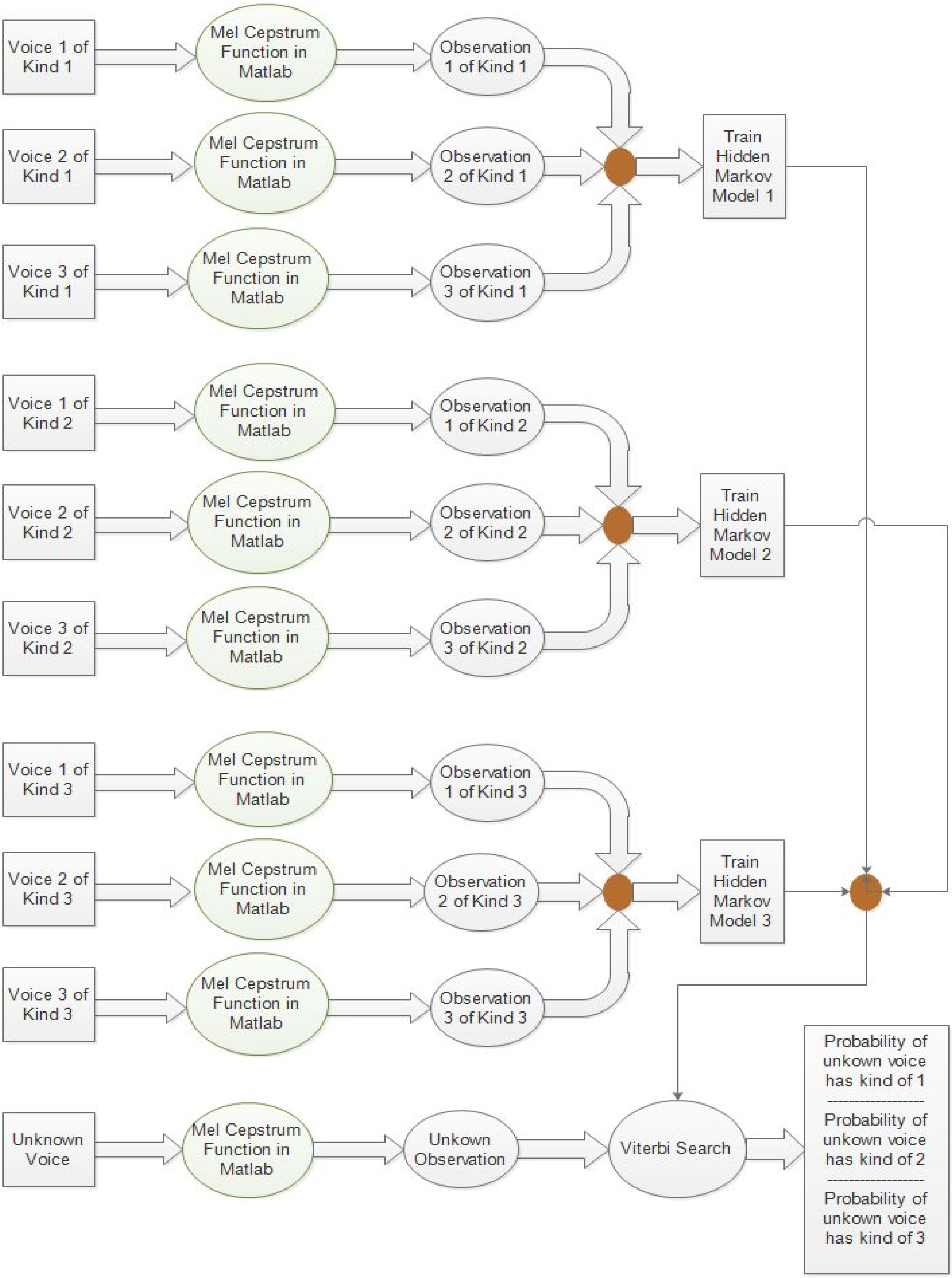
If the results of this project are correct and fast, we can save the world define music genres by manually to automatically. We and our advisor planning to specify music genres with signals. Than we are going to plan future works like determining why people choose these musics.

7. Methodology and Technical Approach

In our project firstly, we will find observations about spesific kind of voices. That is why firstly we will specify kind of signals in nature. It will be human fonetics and genre harmonics of musics. After observing data from nature, we will save it to simple database. These simple database contains .wav files as audio files.

During our project implementation, using matlab buildin function melcepstrum will help us to create observation from .wav files. After create observations, we will estimate and train our Hidden Markov Model. When we train enough, model will be able to used for viterbi searching to new observation. There will be detailed information about melcepstrum [9].

Finally, using viterbi search for each Hidden Markov Model, each one of them will create probability as output. We will choose the highest probability Markov Model. A Markov Model that is chosen will show us the kind of observation.

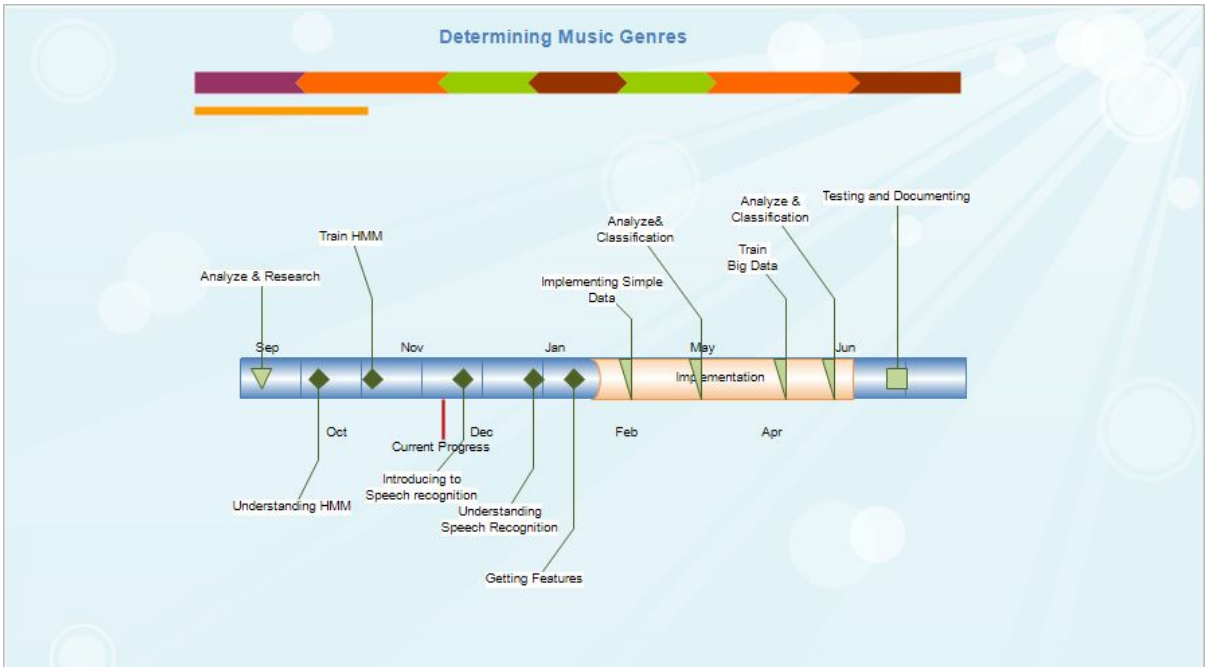


Voice 1 of Kind 1
Mel Cepstrum Function in Matlab
Observation 1 of Kind 1
Voice 2 of Kind 1
Mel Cepstrum Function in Matlab
Observation 2 of Kind 1
Train Hidden Markov Model 1
Voice 3 of Kind 1
Mel Cepstrum Function in Matlab
Observation 3 of Kind 1
Voice 1 of Kind 2
Mel Cepstrum Function in Matlab
Observation 1 of Kind 2
Voice 2 of Kind 2
Mel Cepstrum Function in Matlab
Observation 2 of Kind 2
Train Hidden Markov Model 2
Voice 3 of Kind 2
Mel Cepstrum Function in Matlab
Observation 3 of Kind 2
Voice 1 of Kind 3
Mel Cepstrum Function in Matlab
Observation 1 of Kind 3
Voice 2 of Kind 3
Mel Cepstrum Function in Matlab
Observation 2 of Kind 3
Train Hidden Markov Model 3
Voice 3 of Kind 3
Mel Cepstrum Function in Matlab
Observation 3 of Kind 3
Unknown Voice
Mel Cepstrum Function in Matlab
Unkown Observation
Viterbi Search
Probability of unkown voice has kind of 1
Probability of unkown voice has kind of 2
Probability of unkown voice has kind of 3

8. Management Plan

There will be several phases during the project. These will be in outline as follows:

Phase 1: Research about speech recognition and get informaton every phase.
Phase 2: Understanding of hidden markov models (aka.HMM) behaviour. Train and manipulating HMM. Understanding also Baum-Welch algorithm and Viterbi.
Phase 3: Project Specification Document: there will be preparing of this document with our advisor.
Phase 4: Getting some input documents for our project contains examples of speech recognition.
Phase 5: We will begin to implementation of speech recognition.
Phase 6: This phase will be testing phase for the measure the success of implementation.
Phase 7: There will be preparing for the representations such as create slides and documents for this progress. There will be also Progress Report in this phase.
Phase 8: Getting input documents contains musical signals. That are generated from us and our advisor.
Phase 9: In this phase, we will try to implementation of the real data of musics. We will work machine learning getting some features and classify them.
Phase 10: There will be testing of the implementation.
Phase 11: Getting output file contains classified music genres that are simple.
Phase 12: We will modify the implementation for the wide spectrum of genres.
Phase 13: Again testing phase of implementation for big data.
Phase 14: We will create an interface for simple using.
Phase 15: Testing phase the final version of implementation with our advisor.



9. References

[1] http://htk.eng.cam.ac.uk/ (November 5, 2015)

[2] http://www.newworldencyclopedia.org/entry/Harmonic (November 5, 2015)

[3] https://en.wikipedia.org/wiki/Hidden_Markov_model (November 5, 2015)

[4] Jonathan Foote. An overview of audio information retrieval. Multimedia Systems, 7(1):2–10, 1999.

[5] T. Zhang and C. Kuo. Content-based classification and retrieval of audio. In SPIE's 43rd Annual Meeting - Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations VIII, San Diego, July 1998.

[6] Jonathan Foote. A similarity measure for automatic audio classification. In Proc. AAAI 1997 Spring Symposium on Intelligent Integration and Use of Text, Image, Video, and Audio Corpora, March 1997.

[7] Jeff Bilmes. A model for musical rhythm. In ICMC Proceedings, pages 207–210. Computer Music Association, 1992.

[8] Daniel P. W. Ellis. Prediction-driven computational auditory scene analysis for dense sound mixtures. Technical report, International Computer Science Institute, Berkeley CA, 1996.

[9] William Brent. Department of Music and Center for Research in Computing and the Arts, UCSD.