**1. Goals of the Project**

Research dependence between emotions and acoustic characteristics of speaker. Particularly, between average pitch, loudness and temp of speech. Actually, some emotions can be recognized using acoustic characteristics. In addition, there are few researches where these dependencies are described. In my thesis, I am trying to use these researches and extend them. Particularly, split speech to phrases, and analyze each phrase and estimate whole speech by summing phrase estimations. Also, measure temp of speech to estimate speech confidence. Moreover, at the end- try test this solution to estimate call center’s conversation records. Because, this technology can be used to improving of quality of call-center service. If select expressive conversations from all conversations one can analyze them.

**2. Overview of the System Specification**

In total system should provide an estimation of conversation.

For this, in first, is necessary extracting acoustic characteristic. It should be able read anyone sound files in any format and bitrate.

In second, do normalization and filtering unnecessary noises. It should be provide from all sound information select only voiced regions. There is problem, because each sound file can has different levels of noise and loudness of voice. And to split noise and voice is necessary set optimal threshold.

Software specification: Windows + .NET

**3. Background Theory**

Theoretical basis of my thesis can be divided to 3 parts:

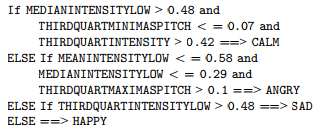
1. How from bytes of sound file get acoustic characteristics of speech.
2. Normalization and filtering techniques
3. Researches about dependencies between emotions and acoustic characteristics and applying them.

First part is solved using Fast Fourier Transform (FFT). FFT provide retrieving pitch and intensity of signal for each time moment. However to apply FFT, in first we need to get sound information from file. Using C# it can be done by some libraries: [NAudio](https://naudio.codeplex.com/), [PRAAT](http://www.praat.org).

In second part [min-max normalization](http://en.wikipedia.org/wiki/Min-maxing) used for intensity normalization. But to filter noises is necessary to set optimal threshold and it will be able only after experiments with real conversation files.

Third part is the most interesting part. In researches described: how emotions influences to pitch and intensity of voice. There are used average pitch, average intensity, and information about how pitch and intensity are changed during conversation. And provided some tables where showed relations between emotions and these characteristics.

For example, these rules from one research:



These simple rules work but it may be not enough. In my thesis, I will try to apply these techniques. In addition, I will bring something new, because I will analyze speech phrase by phrase and analyze temp of speech.

**4. Overview of Task Specification and Project Schedule**

1. Implement module which provide getting acoustic characteristics(pitch, loudness) –Complete

I used NAudio library for sound file reading and implemented FFT algorithm.

This solution represented as SoundParametersExtractor class

I tried to use PRAAT library. It doesn`t require implementing FFT algorithm. It just retrieve sound file and return sound parameters. In addition, PRAAT library provide normalization and filtering functions. And now I am learning how to write scripts in PRAAT scripting language. Because with it we can solve pre-processing problems very easy.

1. Normalizing
   1. Implement intensity normalizing –Complete

I implemented min-max normalization for intensity.

Also, here can be used PRAAT library normalizing functions.

* 1. Find speech pauses and split speech to phrases –on-going

1. (Research)

Base papers:

* <http://www.pyoudeyer.com/emotionsIJHCS.pdf>
* [**http://web.media.mit.edu/~mehoque/Publications/IVA-Hoque-Yeasin-Louwerse.pdf**](http://web.media.mit.edu/~mehoque/Publications/IVA-Hoque-Yeasin-Louwerse.pdf)
* [**http://speetech.by/press/analytics/6**](http://speetech.by/press/analytics/6)
* [**http://www.scholarpedia.org/article/Speech\_emotion\_analysis**](http://www.scholarpedia.org/article/Speech_emotion_analysis)