1) My thesis project is recognition of emotions in speech.

2) People express emotions not only by gestures and facial expression. Voice acoustic characteristics related with expressing emotion. And, there are many studies which describe these dependencies.

3) technology of emotions recognition has very interesting application areas. For example,

* there are situations when robot need take into account human emotions.
* Also, knowing relations between acoustic characteristics and emotions, is possible to implement **emotional speech generating**. because emotional robots are more people-friendly
* Closer to reality application domain of this technology is **estimation of customer service.** For example, call-centers store conversation records. And to estimate operator’s work, we can analyze records.

Also, it can be used in some psychological research.

4) Goals of my work. As I said there many studies about emotion recognition, and my goal is overview them and

* Distinguish differences
* Implement simple solution
* Compare accuracy

5)

6) there are many studies and their main differences in using different…

By selected emotions I mean recognizable emotions.

so let's consider each of them

7)

* Pitch or voice frequency.
* Loudness
* Formants- [acoustic](https://en.wikipedia.org/wiki/Acoustics) [resonance](https://en.wikipedia.org/wiki/Resonance) of the human vocal tract
* MFCC- is a representation of the short-term [power spectrum](https://en.wikipedia.org/wiki/Power_spectrum) of a sound

For further classification we need to extract features from these characteristics

8) Features can be divided to two groups

9) From one side classiﬁcation problem in this case very similar to other problems. and can be used general approaches. But from other side it has some specific points caused by similarity of emotions and ambiguity. That is why also is important selected emotions for recognition.

10) Accuracy of classification depends on the number of recognizable emotions. Acoustic characteristics. For example, in the table are associated characteristics and emotions. And we can see that some emotions have very similar characteristics. For example, anger-joy.

12) Implementatin can be divided to four steps. Let’s consider each of them.

13) Firstly, to avoid wrong results, sound should be filetered. Then we extract these characteristics from sound using PRAAT software package. It Provides all needed functions.

15) After first step we get out characteristics. For example in this graphic. U can see some zero intervals. It’s unvoiced intervals.

16) and these intervals allow us to divide speech to phrases.

17) So I use speaker independent features, it means that recognition is possible for any person.

18) so what is the DDS. For example, it is speech divided to phrases. And within each phrase we find local max and min values. Then difference and distance between them.

19) As training set I used EmoDB. German emotional database. It consist of more than two hundred fifty records. 70% of tem used for training and 30 for testing

21) and from training set calculated and constructed two fuzzy membership functions for each feature. The first, returns probability of that value is low, and second that value is high.

22) Then dependencies between features and emotions can be represented as in this table.

23) Eventually, our system has three main components. Extractor extracts sound characteristics from sound file. Then Analyzer compute features from sound characteristics, and classifier based on fuzzy rules gives probability of each emotion.

24) Before we consider our results, let's see how people well in emotion recognition. because there is real ambiguity between some emotions

25) we see that accuracy does not allow to use it for decision making, but it can be used as decision support. And classification without neutral emotion is more accurate.

26) also I have tested with active-passive classification. It can be used in call-center to select expressive conversations. And accuracy is acceptable