**Abstact**

В этой статье рассматриваются исследования по распознаванию эмоций в речи. Выделяются отличия некоторых подходов и раскрываются плюсы и минусы этих подходов. Затем, на основе проанализированной информации предлагается реализация упрощенной системы распознавания трех базовых эмоций: злость, радость и грусть.

**Введение**

Технология распознавания эмоций человека может быть полезна в разных областях. Наиболее очевидные из которых: построение роботов и оценка качества обслуживания клиентов. Если у говорящего робота будет способность эмоционально окрашивать некоторые выражения, это позволит ему лучше взаимодействовать с людьми. Также, чтобы робот мог хорошо взаимодействовать с людьми, ему необходимо хорошо понимать людей. Могут возникать ситуации, когда при принятии определенных решений, робот должен учитывать эмоциональное состояние человека. Например,(ПРИВЕСТИ ПРИМЕР).

Более приближенное к настоящему применение этой технологии лежит в области оценки качества обслуживания клиентов. Например, call-центры ведут записи звонков, и для оценки качества работы операторов, можно анализировать звонки и выявлять особо эмоциональные разговоры.

Существует множество исследований, раскрывающих зависимости между акустическими характеристиками и эмоциями человека. Отличаются эти исследования по различным критериям:

* *Акустические характеристики речи*
* *Признаки или свойства акустических характеристик для различных эмоций*
* *Методы классификации*
* *Распознаваемые эмоции*

Главной задачей этой работы является выяснить как влияют эти критерии на точность распознавания и предложить свою реализацию.

**2. Обзор исследований**

*2.1 Акустические характеристики*

Прежде чем приступать к рассмотрению конкретных исследований, стоит рассмотреть акустические характеристики речи:

* *Основная частота*
* *Громкость*
* *Форманты*
* *Мел-кепстральные коэффициенты (MFCC)*

*Основная частота (fundamental frequency)* – или по-другому высота тона речи. Человеческая речь по частоте примерно расположена в диапазоне от 300 до 3400 Гц. А *основная частота* речи у мужчин расположена от 85 до 185 Гц, у женщин от 165 до 255.

Наиболее часто в исследованиях используют только *высоту тона* и *громкость.* Хотя такие характеристики как *форманты* и *мел-кепстральные коэффициенты* тоже несут полезную информацию.

*Форманта- термин фонетики, обозначающий акустическую характеристику звуков речи (прежде всего гласных), связанную с уровнем частоты голосового тона и образующую тембр звука (https://ru.wikipedia.org/wiki/Форманта).*

*Мел-кепстральные коэффициенты(MFCC)- часто используются для характеристики речевых сигналов. Чаще используются для распознавания речи.*

*2.3 Признаки или свойства акустических характеристик*

Для классификации эмоций по акустическим характеристикам необходимо выделить некоторые признаки для определенных эмоций. И эти эталонные признаки рассчитываются на основе обучающей выборки. Таким образом, при классификации речи происходит следующее:

1. Рассчитываются признаки для заданной речи
2. Признаки заданной речи сравниваются с эталонными признаками

По тому как рассчитываются эталонные признаки их можно разбить на две группы: speaker dependent и speaker independent.

*Speaker dependent*

В этом случае нам необходима обучающая выборка для каждого человека, речь которого будет оцениваться. Этот подход не универсален, и может быть применен в узкой сфере. Хотя и может оказаться, что такие признаки могут оказаться более точными. Например: средняя частота(тон), средняя громкость, диапазон тона и громкости. Такие признаки обычно основаны поиске средних значений акустических характеристик.

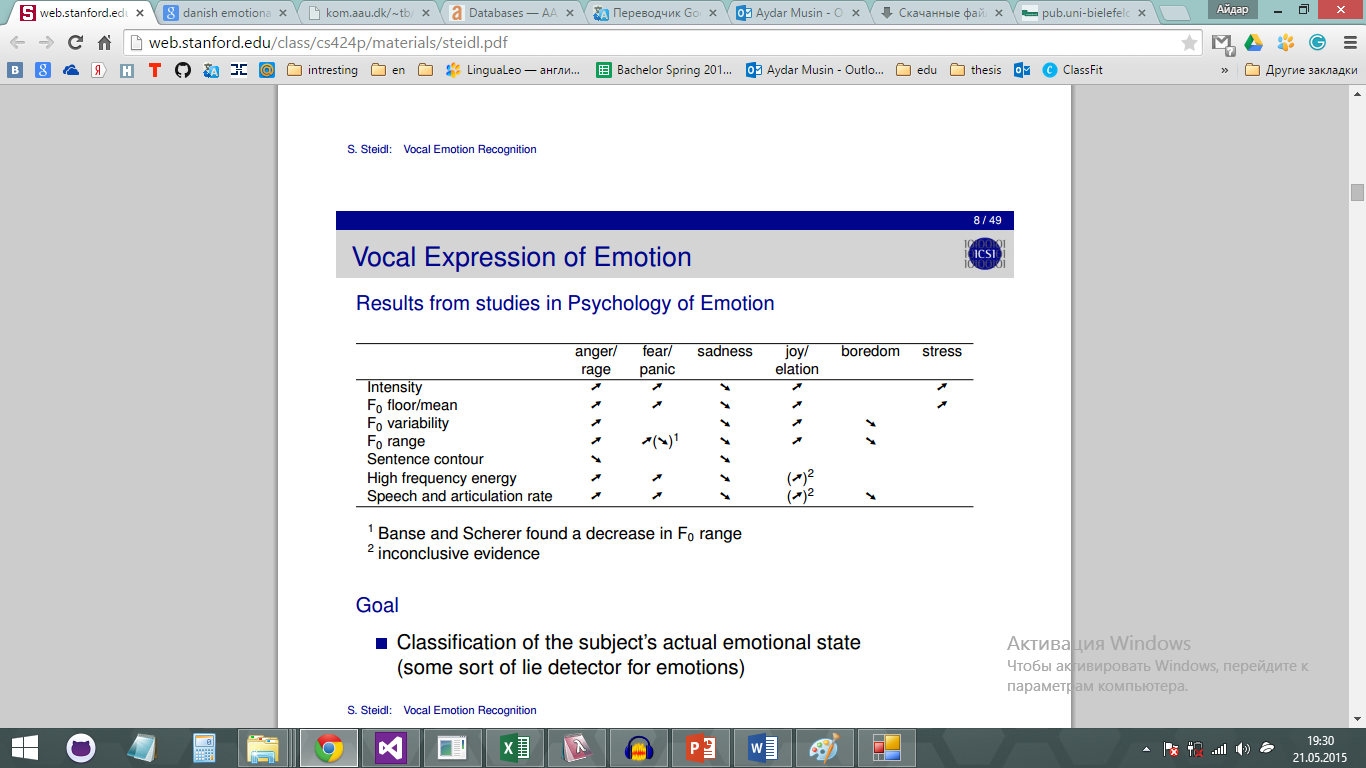
*Speaker independent*

Универсальный подход, потому что все признаки рассчитаны для различных людей. И для точной работы необходима большая обучающая выборка с большим количеством людей. Например, Pitch DDS, Formant DDS, длительность фраз, длительность пауз, DDS (разница-расстояние-наклон) используется для оценки динамики речевого сигнала. Эти признаки более подробно описываются в части:????????.

**2.4. Методы классификации**

**2.5 Распознаваемые эмоции**

Точность распознавания зависит не только от выделяемых признаков и методов классификации. Но и зависит от выбранных распознаваемых эмоций. Вызвано это тем, что некоторые эмоции очень похожи. И чтобы правильно классифицировать похожие эмоции необходимо либо вводить новые признаки, либо улучшать методы классификации. Поэтому это компромисс между количеством распознаваемых эмоций и точностью. И необходимое количество эмоций зависит от требований конкретной задачи.



Например, эта таблица из статьи [Stefan Steidl; October 26, 2010;” Vocal Emotion Recognition”; ICSI].

На этой таблице видно, что по многим признаками такие эмоции как злость(anger) и радость(joy) очень похожи. Поэтому довольно важно определить набор

В большинстве исследований выделяют две основные характеристики речи: высота тона (частота звука) и громкость. И при оценке эмоционального состояния сравнивают текущие характеристики с средними. Такой подход не универсален, ведь такие сложности как: разный голос у людей, разное качество записи, не позволяют рассчитать среднее значение этих характеристик для всех людей. И получается, чтобы наиболее точно определять эмоции определенного человека, необходимо несколько его нейтральных записей речи.

Также существуют исследования, в которых предлагается работать с другими

Но люди, даже если слышат голос впервые, могут определять базовые эмоции, такие как: злость, радость, грусть. Поэтому в этом исследовании мы попытаемся собрать лучшие техники, привнести свои и разработать систему, способную распознавать эмоции любого человека без статистических данных на него.

**Проблема**

Задачу распознавания эмоций человека по записи его речи можно разбить две на части:

* Считывание данных с записи, избавляясь от лишних шумов, нормализуя громкость.
* Связывание всех меняющихся характеристик речи с базовыми эмоциями.

Поскольку в записи часто присутствуют различные шумы, которые могут повлиять на получение правильных результатов, необходимо к записи предварительно применить несколько фильтров(4.2). Также запись может быть очень разбросана по громкости, поэтому необходимо нормализовать громкость.

Как уже говорилось в части 1, многие исследования используют такие характеристики как средняя высота тона и громкости. Но все люди разные и у всех разные тон и громкость. При этом, эти показатели могут еще зависеть от качества записи. Например, если микрофон стоял далеко, то громкость будет маленькой и правила не сработают.

А тон, по-другому частота звука, у каждого индивидуальная. И невозможно подобрать какое-то глобальные средние значения тона и громкости, которые использовались бы при оценке эмоционального состояния человека.

В этой статье мы будем рассматривать только три базовые эмоции: злость, радость, грусть. В то время как в других исследованиях еще присутствуют такие эмоции как: отвращение, страх, удивление и т.д. Это обусловлено тем, что эти базовые эмоции наиболее ярко выражены и легко распознаются человеком.

Выделим некоторые характеристики, которые описаны в исследованиях (напоминаю, мы работаем только с тремя эмоциями, поэтому характеристики приводятся только для них):

*- радость: тон выше среднего, громкость выше среднего, быстрый ритм.*

*- злость: тон выше среднего, громкость намного выше среднего*

*- грусть: тон ниже среднего, громкость ниже среднего*

Первое, что бросается в глаза, так это слово *среднее.* Везде для определения эмоции требуется среднее значения для сравнения. Если перед нами стоит задача определения эмоций одного человека, тогда мы можем собрать множество его нейтральных записей речи и подсчитать эти средние значения для этого человека.

Но наша задача определять эмоции без предварительного сбора информации для каждого человека. Поэтому необходимо выделить признаки, позволяющие определять эмоции без средних значений.

**Идея**