Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н. Э. Баумана (национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ ИУ «Информатика и системы управления»

КАФЕДРА ИУ-7 «Программное обеспечение эвм и информационные технологии»

ОТЧЕТ

ПО НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

«Классификация задач распознавания эмоций из звучащей речи и способов их решения»

Студент ИУ7-76Б _____ Т. А. Казаева
(Подпись, дата)

Руководитель _____ Ю. В. Строганов
(Подпись, дата)

2022 г.

РЕФЕРАТ

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

1) **Речевой звук** – кратчайшая, далее неделимая единица языка.

2) **Артикулярный тракт** – совокупность огранов человеческого тела, которые используются в процессе голосоведения, и в первую очередь в процессе речепроизводства (лёгкие, гортань, глотка, ротовая и носовая полости, увула, мягкое и твердое небо, язык, зубы, губы).

3) **Голосовой источник звука** – периодическая модуляция голосовыми складками воздушного потока, подаваемого из лёгких.

4) **Шумовой систочник звука** – генерация шума турбулентными завихрениями воздушного потока в сужениях речеобразующего аппарата.

5) **Импульсный источник звука** – скачкообразное изменение давления воздуха при резком открытии смычки в артикулярном тракте.

6) **Турбулентный поток** – поток трения воздушной струи, вызванный сужением артикулирующих органов.

7) **Маскировка** – повышение порога слышимости звука (стимула) в присутствии других звуков (маскеров). Маскирующие звуки могут предшествовать (прямая), действовать одновременно (одновременная) и следовать за сигналом (обратная маскировка).

ВВЕДЕНИЕ

1. Задача распознавания эмоций по звучащей речи

2. Классификация эмоций

Одной из главных проблем в исследованиях, связанных с определением эмоционального состояния диктора по голосу, является отсутствие чёткого определения эмоции. При формализации этого понятия возникают сложности в силу многообразия психологических моделей эмоциональных процессов. Подход к классификации эмоций влияет на процесс аннотирования – разметки аудиовизуального эмоционально окрашенного контента.

Для формализации эмоциональных данных необходимо сформировать полноценную классификацию эмоциональных состояний, от которой в том числе напрямую зависит процесс аннотирования – сопоставления инофрмативных признаков, полученных из речи диктора с определенными эмоциями и аффективными состояниями.

Сегодня широко используются три подхода : дискретная и многомерная модели, а также гибридная.

2.1 Дискретная модель эмоциональной сферы

Дискретный подход основан на выделении фундаментальных (базовых) эмоций, сочетания которых порождают разнообразие эмоциональных явлений. Разные авторы называют разное число таких эмоций – от двух до десяти. П. Экман на основе изучения лицевой экспрессии выделяет пять базовых эмоций: гнев, страх, отвращение, печаль и радость. Первоначальная версия 1999 года также включала «удивление» [1; 2]. Р. Плутчик [3] выделяет восемь базисных эмоций, деля их на четыре пары, каждая из которых связана с определенным действием: страх, уныние, удивление и т. д. На рисунке 2.1 представлена схема классификации эмоций, предложенная П. Экманом.

На сегодняшний день концепция существования базовых эмоций ставится под сомнение. Теория встречает ряд концептуальных проблем, таких как, например, эмпирическое определение набора базовых эмоций или критерии синхронизации эмоциональных реакций. Однако, многие решения в области автоматического детектирования эмоций основаны на дискретной модели эмоциональной сферы. Например, решение компании Affectiva [4].

Рис. 2.1 – Атлас эмоций, предложенный П. Экманом

2.2 Многомерная модель эмоциональной сферы

Многомерная модель представляет собой эмоции в координатном многомерном пространстве. В качестве её источника рассматривают идею В. Вундта о том, что многогранность чувств человека можно описать с помощью трёх измерений: удовольствие - неудовольствие, расслабление - напряжение, возбуждение - успокоение. Вундт заключил, что эти измерения охватывают все разнообразие эмоциональных состояний [5]. Данные для этой теории были получены с помощью метода интроспекции.

Эмоциональная сфера представляется как многомерное пространство, образованное некоторым количеством осей координат. Оси задаются полюсами первичных характеристик эмоций. Отдельные эмоции – это точки, местоположение которых в «эмоциональном» пространстве определяется степенью выраженности этих параметров.

Один из примеров описываемого подхода – модель Дж. Рассела. В ней водится двумерный базис, в котором каждая эмоция характеризуется знаком (*valence*, валентность) и интенсивностью (*arousal*, активацией). Измерение ва-

лентности отражает то, насколько хорошо человек ощущает себя на уровне субъективного переживания от максимального неудовольствия до максимального удовольствия. Измерение активации связано с субъективным чувством энергии и ранжируется в диапазоне от дремоты до бурного возбуждения. Такой подход используется, например, в датасете RECOLA [6].

Аналогично вопросу о количестве эмоций в дискретной модели, вопрос о количестве измерений остаётся открытым. Использование только двух критикуется на том основании, что они не позволяют устанавливать различия между отдельными эмоциональными состояниями (например, страх, гнев, ревность, презрение и др. имеют отрицательную валентность и высокую активацию).

2.3 Гибридная модель эмоциональной сферы

Гибридная модель представляет собой комбинацию дискретной и многомерной модели. Согласно этой классификации, в отдельной области n -мерного эмоционального пространства различия между эмоциями могут определяться в терминах измерений, имеющих отношение к этой области. Эмоции могут быть сопоставимы по измерениям внутри и вне категорий, и каждая категория может иметь свои отличительные признаки [7]. Примером такой модели являются «Песочные часы эмоций», предложенные Камбрией, Ливингстоном, Хуссейном [8]. Каждое измерение характеризуется шестью уровнями силы, с которой выражены эмоции. Данные уровни обозначаются набором из двадцати четырёх эмоций. Поэтому совершенно любая эмоция может рассматриваться как и фиксированное состояние, так и часть пространства, связанная с другими эмоциями нелинейными отношениями.

	Pleasantness	Attention	Sensitivity	Aptitude
+3	ecstasy	vigilance	rage	admiration
+2	joy	anticipation	anger	trust
+1	serenity	interest	annoyance	acceptance
0	—	—	—	—
-1	pensiveness	distraction	apprehension	boredom
-2	sadness	surprise	fear	disgust
-3	grief	amazement	terror	loathing

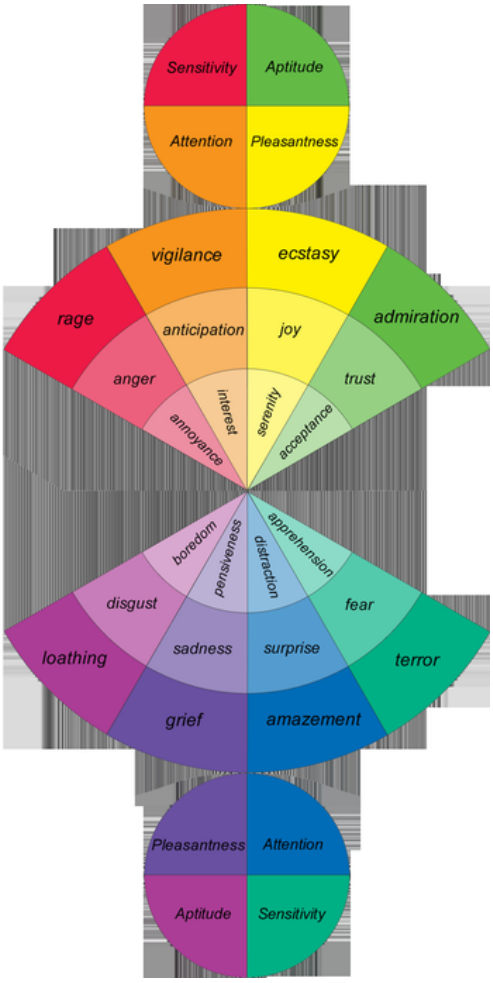

Рис. 2.2 – «Песочные часы эмоций»

Выбор подхода от цели. Многомерные модели позволяют избежать проблемы существования некоторых слов в каких-то языках, в то время как в других может не быть слов для описания этих эмоций. Это делает процесс аннотирования культурно-зависимым. Тем не менее разные аннотаторы дают разные оценки выраженности валентности или активации, поэтому в целях упрощения модели для некоторых задач больше подходит дискретная модель.

3. Автоматическое распознавание речи

3.1 Речеобразование и классификация звуков

Реечвой аппарат человека представлен на рисунке 3.1. Акустически процесс речеобразования образования состоит из нескольких независимых этапов. [9] Первый – возникновение звука в артикуляторном тракте, которое, в свою очередь, может быть реализовано либо голосовым, либо шумовым, либо импульсным источником. Второй – формирование возбуждённого звука и его излучение в пространство.

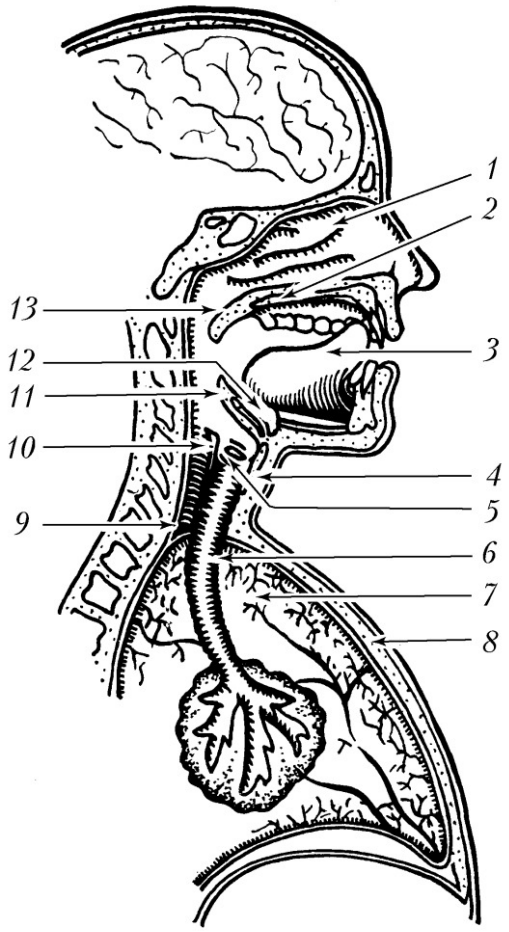

Рис. 3.1 – Речевой аппарат человека: *1 – носовая полость, 2 – твёрдое нёбо, 3 – язык, 4 – щитовидный хрящ, 5 – голосовые связки, 6 – трахея, 7 – лёгкое, 8 – грудина, 9 – пищевод, 10 – кольцеобразный хрящ, 11 – надгортанье, 12 – подъязычная кость, 13 – мягкое нёбо.*

Речевые звуки можно классифицировать в зависимости от источника воз-

никновения. Классификация приведена в таблице 3.1.

Таблица 3.1 – Классификация речевых звуков

РАЗДЕЛ		АРТИКУЛЯЦИЯ	ИСТОЧНИК	ПРИМЕР
гласные		не создаётся существенных препятствий	голосовой	[А], [У]
смычные согласные	аффрикаты	размыкание смычки происходит плавно	голосовой совместно с шумовым	[Ч], [Ц]
	взрывные	нёбная занавеска поднята, и воздух проходит в ротовую полость, а размыкание смычки происходит резко и напоминает взрыв	звонкие – голосовой с импульсным, глухие – импульсный	[П], [Т], [К] , [Б], [Д], [Г]
сонорные согласные		без участия турбулентного потока воздуха в голосовом тракте	голосовой	[Й’], [Л]
щелевые (фрикативные) согласные		артикуляторы подходят близко друг к другу, но не смыкаются полностью, в результате чего в ротовой полости происходят турбулентные колебания воздуха, создающие заметный шум	звонкие – голосовой совместно с шумовым, глухие – шумовой	[Ж], [Ж’:], [Ш], [Ш’:]

3.2 Слуховая система

Для задачи автоматического распознавания речи важно изучение слухового анализатора, поскольку неадекватная обработка сигнала может привести к потере полезных признаков.

Периферическую слуховую систему представляют как последовательно включенные наружное, среднее и внутреннее ухо (рисунок 3.2).

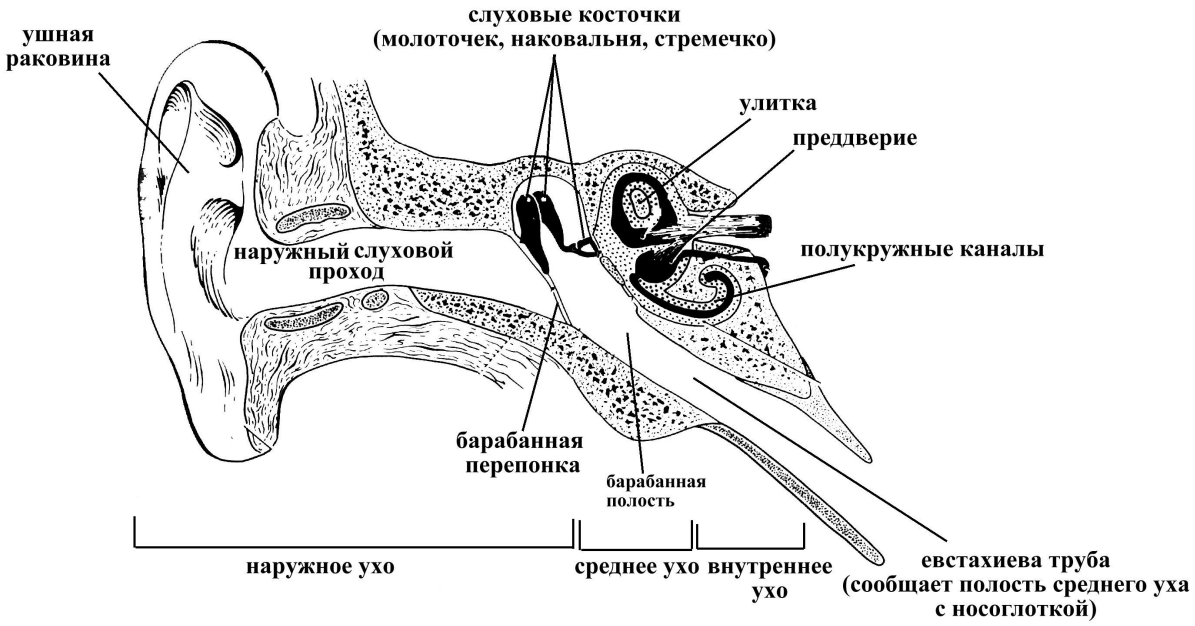

Рис. 3.2 – Слуховая система человека

Система каналов улитки является акустическим фильтром низких частот. Таким образом, высокочастотные составляющие слышимого звука, а также продуктов отоакустической эмиссии затухают в вестибулярном и тимпанальном каналах и не достигают круглого окна. Частота запирания ниже 14 кГц[10].

3.2.1 Восприятие высоты звука

О восприятии звука на разных частотах даёт представление следующий эксперимент [11]. Слушателю предъявляют гармонический сигнал и просят выставить частоту второго сигнала так, чтобы на его субъективный взгляд она была в два раза выше или ниже, чем частота предъявленного. Восприятие высоты звука зависит от его интенсивности, поэтому, для получения однозначных результатов, интенсивность стимулов во втором эксперименте фиксируют в 40 дБ. Из полученных результатов следует, что слуховая система склонна рассматривать низкочастотные компоненты речи более подробно, чем высокочастотные – начиная с 1000 Гц, шкалы можно считать близкими к логарифмическим.

3.2.2 Восприятие громкости звука

Громкость – это субъективная оценка интенсивности звука. Ощущение громкости зависит не только от частоты, но и от длительности звукового сти-

мула. Для количественной оценки абсолютной громкости была принята специальная единица сон. Громкость в 1 сон – это громкость синусоидального звука с частотой 1000 Гц и уровнем 40 дБ относительно звукового давления $2 \cdot 10^{-5}$ Па.

Количественно зависимость воспринимаемой громкости звука (в сонах) и его звукового давления может быть представлена в виде 3.1:

$$S = Cp^{0.6} \tag{3.1}$$

4. Информативные признаки, характеризующие речь

Выделение информативного набора признаков, коррелированных с эмоциональным состоянием, оказывает значительное влияние на эффективность автоматического детектирования эмоций.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Ekman P.* Universals and Cultural Differences in Facial Expression of Emotion // Nebraska Symposium on Motivation. Vol. 19. — 1972.
2. *Ekman P.* An Argument for Basic Emotions // Cognition and Emotion. — 1992.
3. *Plutchik R.*, *Kellerman H.* Emotion: Theory, Research, and Experience: Vol. 1. Theories of Emotion. // London: Academic Press. — 1980.
4. Affectica. — URL: `https://www.affectiva.com/` (дата обращения: 30.9.2022).
5. *Вундт В.* Психология душевных волнений // Психология эмоций. — 1984.
6. RECOLA. — URL: `https://diuf.unifr.ch/main/diva/recola/` (дата обращения: 1.10.2022).
7. *Russell J.* Core Affect and the Psychological Construction of Emotion // Psychological Review. — 2003.
8. Sentic Computing for social media marketing / E. Cambria [и др.] // Multimedia Tools and Applications - MTA. — 2012. — DOI: `10.1007/s11042-011-0815-0`.
9. *Фант Г.* Акустическая теория речеобразования. — 1964.
10. *Дидковский В.С. Лунева С.А. К. С.* Передаточная функция улитки внутреннего уха человека. Часть 2 // Биомедицинские приборы и системы. — 2014.
11. *Алдошина И.* Основы психоакустики. // Звукорежисер. — 1999.