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»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ

НА ТЕМУ:

«Метод распознавания эмоций по звучащей речи на основе скрытой марковской модели»

Студент ИУ7-76Б _____ Т. А. Казаева

(Подпись, дата)

Руководитель _____ Ю. В. Строганов

(Подпись, дата)

2023 г.

РЕФЕРАТ

Расчетно–пояснительная записка 15 с., 1 рис., 0 табл., 13 ист., 0 прил.

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

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

1) **Фреймы** – отрезки аудиосигнала длительности как правило 10-40 мс, идущие «внахлест», то есть таким образом, чтобы начало очередного фрейма пересекалось с концом предыдущего.

2) **Мел** – единица измерения частоты звука, основанная на статистической обработке большого числа данных о субъективном восприятии высоты звуковых тонов.

ВВЕДЕНИЕ

1 Аналитический раздел

1.1. Категоризация эмоциональных данных

Одной из главных проблем в исследованиях, связанных с определением эмоционального состояния диктора по голосу, является отсутствие четкого определения эмоции. Подход к классификации эмоций влияет на процесс аннотирования. Сегодня широко используются три подхода к категоризации эмоциональных данных: дискретный, многомерный и гибридный.

1.1.1. Дискретное пространство эмоций

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 сегодняшний день существование 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]

1.1.2. Многомерное пространство эмоций

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

Эмоциональная сфера представляется как многомерное пространство, об-

разованное 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 валентностью (*англ. valence*) и интенсивностью (*англ. arousal*). Измерение валентности 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 наборе данных «RECOLA» [6].

Аналогично вопросу о количестве эмоций в дискретной модели, вопрос о количестве измерений остается 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).

1.1.3. Гибридное пространство эмоций

Гибридная модель представляет собой комбинацию дискретной и многомерной модели. Примером такой модели являются «Песочные часы эмоций», предложенные Камбрией, Ливингстоном и Хуссейном. [7]

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. [8] Каждое измерение характеризуется шестью уровнями силы, с которой выражены эмоции. Данные уровни обозначаются набором из двадцати четырех эмоций. Поэтому совершенно любая эмоция может рассматриваться как и фиксированное состояние, так и часть пространства, связанная с другими эмоциями нелинейными отношениями.

1.2. Информативные признаки эмоциональных состояний

1.2.1. Классификация признаков

На эффективность классификации значительное влияние оказывает выбор набора информативных признаков, характеризующих речь. В целом, признаки можно разбить на две категории: спектральные (акустические) и лингвистические (просодические). [9]

Лингвистические признаки (мелодика речи, интенсивность, ритм, тембр, сила голоса) характеризуют содержательный аспект речи. При вычислении акустических признаков речевой поток рассматривается как некоторый квазистационарный процесс.

При вычислении спектральных признаков речевой сигнал представляется в виде дискретной последовательности цифровых значений амплитуды речевой волны, подвергается спектральному анализу. Эти характеристики могут быть условно разделены на 9 групп. [10]

1. Средние значения спектра анализируемого речевого сигнала.
2. Нормализованные средние значения спектра.
3. Относительное время пребывания сигнала в полосах спектра.
4. Нормализованное время пребывания сигнала в полосах спектра.
5. Медианные значения спектра речи в полосах.
6. Относительная мощность спектра речи в полосах.
7. Величины вариации огибающей спектра речи.
8. Нормализованные величины вариации огибающих спектра речи.
9. Значения коэффициентов кросскорреляции спектральных огибающих между полосами спектра.

Признаки 1-7 отражают особенности речевых трактов у разных лиц. Интенсивность сигнала определяют признаки 1, 2. Признаки 7, 8 связаны с динамикой перестройки артикуляционных органов речи говорящего. Признак 9 характеризует синхронность органов речи говорящего.

Выбор категории зависит от решаемой задачи. Например, в системах где речь соответствует заранее определенному сценарию (словарь голосовых команд) в наборе присутствуют в основном спектральные характеристики, в то время как при анализе слитной речи выбираются лингвистические характеристики. Для анализа эмоционального компонента речи, не касающегося её смысла, используются как просодические признаки (громкость, высота, ритм), так и спектральные.

1.2.2. Характеристика речи эмоционального состояния

1.3. Извлечение информативных признаков

Для создания и обучения модели, которая в будущем сможет предсказывать эмоции по речевому сигналу, необходимо извлечь информативные признаки из речи и перевести их в количественные показатели.

Пусть на вход модели поступает аудиосигнал. Для того, чтобы привести его в вид, который будет использован алгоритмом распознавания (оцифровать), сигнал делится на фреймы и свойства определяются в каждом фрейме. Размер фрейма выбирается от 20 до 40 мс, поскольку считается, что речевой сигнал на этом промежутке стационарный. Каждая точка, как правило, перекрывается дважды. [11] Таким образом, речевой сигнал представляется в виде 1.1:

$$x_i(n),\ 0 \leq n < N, \tag{1.1}$$

где N – размер фрейма, $x_i(n)$ – i-ый фрейм.

1.3.1. Метод главных компонент

1.3.2. Мел-кепстральные коэффициенты

Для представления огибающей спектра, которой описывается форма голосового тракта были введены мел-кепстральные коэффициенты (*англ. MFCC*). [12]

Шкала Мел (рисунок 1.1) соотносит воспринимаемую частоту или высоту чистого тона (мел) с фактической измеренной частотой (Гц). Люди гораздо лучше различают небольшие изменения высоты звука на низких частотах, чем

на высоких. [13]

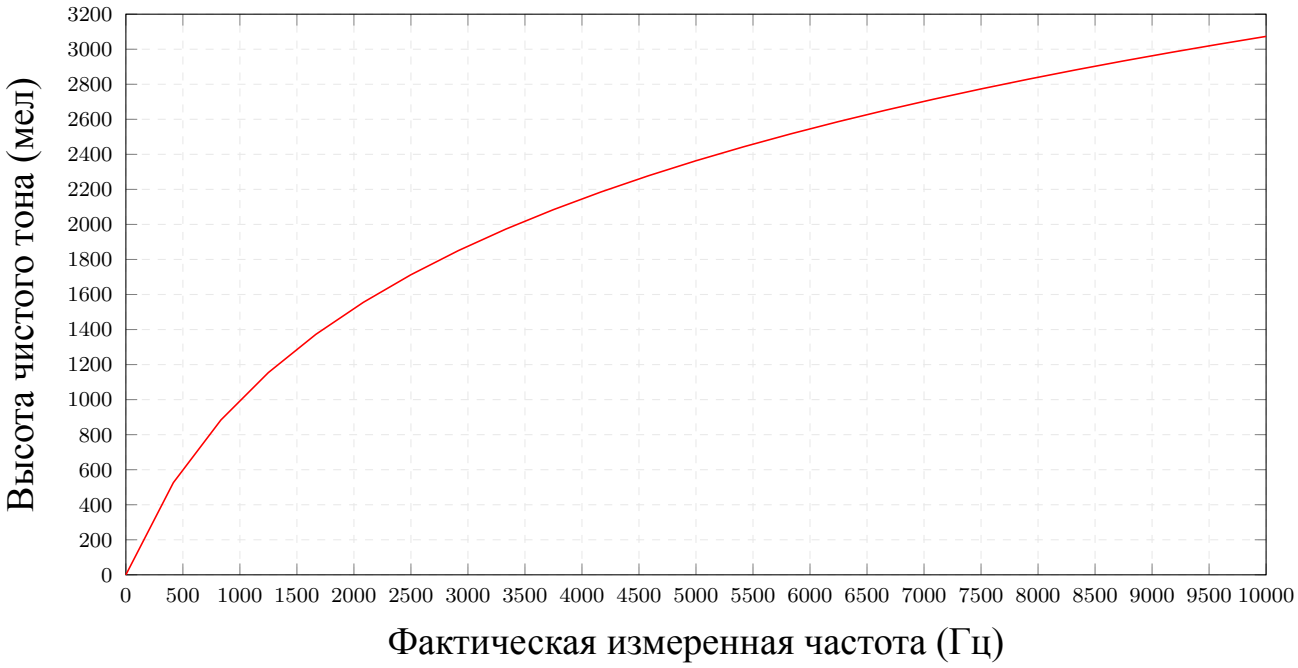

Рис. 1.1 – График зависимости частоты от мел

Формула перевода фактических частот в высоту чистого тона описывается согласно 1.2:

$$m = 1127,01048 \cdot \ln(1 + \frac{f}{700}), \tag{1.2}$$

где где f – фактическая измеренная частота (Гц), m – чистого тона (мел). Обратное преобразование из мел в фактическую частоту вычисляется согласно формуле 1.3:

$$f = 700(e^{m/1127,01048} - 1). \tag{1.3}$$

1.3.3. Кодирование с линейным прогнозированием

1.3.4. Перцептивное линейное предсказание

1.4. Существующие наборы речевых данных

1.5. Классификаторы, используемые в SER

1.6. Постановка задачи

2 Конструкторский раздел

3 Технологический раздел

4 Исследовательский раздел

ЗАКЛЮЧЕНИЕ

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. [Электронный ресурс], URL: https://www.affectiva.com/ (дата обращения: 30.9.2022).
5. *Вундт В.* Психология душевных волнений // Психология эмоций. 1984.
6. RECOLA. [Электронный ресурс], URL: https://diuf.unifr.ch/main/diva/recola/ (дата обращения: 1.10.2022).
7. Sentic Computing for social media marketing / E. Cambria, M. Grassi, A. Hussain, C. Havasi // Multimedia Tools and Applications - MTA. 2012. DOI: 10.1007/s11042-011-0815-0.
8. *Russell J.* Core Affect and the Psychological Construction of Emotion // Psychological Review. 2003.
9. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge / B. Schuller, A. Batliner, S. Steidl, D. Seppi // Speech communication. 2011. Т. 53, № 9/10. С. 1062—1087.
10. *Розалиев В.* Построение модели эмоций по речи человека // Известия Волгоградского государственного технического университета. 2007. № 9. С. 65—68.
11. Лекция 6. Признаки. Кепстральные коэффициенты. MFCC. [Электронный ресурс], URL: https://logic.pdmi.ras.ru/~sergey/oldsite/teaching/asr/notes-06-features.pdf (дата обращения: 17.2.2023).
12. Первичный анализ речевых сигналов. [Электронный ресурс], URL: https://alphacephei.com/ru/lecture1.pdf (дата обращения: 17.2.2023).
13. *Chauhan P.*, *Desai N.* Mel Frequency Cepstral Coefficients (MFCC) based speaker identification in noisy environment using wiener filter // Proceedings of International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE 2014). 2014.

ПРИЛОЖЕНИЕ А