



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

РЕФЕРАТ	2
ВВЕДЕНИЕ	5
1. Классификация эмоций	6
1.1. Дискретная модель эмоциональной сферы	6
1.2. Многомерная модель эмоциональной сферы	7
1.3. Гибридная модель эмоциональной сферы.	8
2. Информативные признаки, характеризующие речь	10
2.1. Просодические характеристики	11
2.2. Спектральные характеристики	12
2.3. Динамические характеристики	13
3. Система распознавания речевых эмоций	14
3.1. Постановка задачи.	14
3.2. Предварительная обработка сигнала	14
3.3. Предобработка характерных особенностей	15
3.4. Классификация	16
3.4.1 Скрытые Марковские модели	16
3.4.2 Метод k-ближайших соседей	17
3.4.3 Модель гауссовых смесей	17
ЗАКЛЮЧЕНИЕ	19
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	22

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

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

- 1) **РЕЧЕВОЙ ЗВУК** – кратчайшая, далее неделимая единица языка.
- 2) **ФОНЕМА** – необходимый и достаточный звуковой минимум для конституирования морфемы, точнее, экспонента последней.
- 3) **АЛЛОФОН** – Вариант фонемы, обусловленный конкретным фонетическим окружением, т. е. оттеночная фонема.
- 4) **ДИФОН** – соседняя пара фонем в высказывании.
- 5) **ТРИФОН** – последовательность из трех соседних фонем.
- 6) **ДЖИТТЕР** – мера возмущений частоты основного тона, показывающая произвольные изменения в частоте смежных вибрационных циклов голосовых складок.
- 7) **ШИММЕР** – мера, характеризующая возмущениями амплитуд сигнала на смежных циклах колебаний основного тона.
- 8) **ВОКАЛИЗАЦИЯ** – отношение количества вокализованных к количеству невокализованных кадров.
- 9) **МЕЛ** – единица высоты звука, основанная на восприятии этого звука органами слуха.
- 10) **ГРЕБЕНЧАТЫЙ ФИЛЬТР, ГРЕБЕНКА ФИЛЬТРОВ** – фильтр, реализованный путем добавления версии с задержкой сигнала самому себе, вызывая конструктивную и деструктивную помеху. АЧХ гребенчатого фильтра состоит из серии регулярно расположенных меток, что создает вид гребенки.
- 11) **ОКНО** – анализируемый сегмент сигнала.

ВВЕДЕНИЕ

Идентификация эмоционального состояния на основе характеристик речевого сигнала вызывает интерес как со стороны научного сообщества, так и со стороны многих коммерческих организаций. Каждая такая система имеет некоторые задачи, которые она призвана решать, и комплекс подходов, которые применяются для решения поставленных задач.

На сегодняшний день в когнитивистике и исследованиях связанных с искусственным интеллектом происходит интенсивное развитие информационных технологий, которые вносят качественное улучшение при взаимодействии между субъектами в системах взаимодействия «человек-компьютер» и «человек-человек».

В системе «человек-компьютер» системы идентификации эмоционального состояния находят применение при коммуникации людей с голосовыми помощниками, эмоциональном окрашивании речи операторов автоматизированных колл-центров, улучшении систем виртуальной реальности и. т. д. В системе «человек-человек» появляется возможность уточнения автоматизированного перевода, улучшения систем детектирования лжи, более точной диагностики психических расстройств на основе изменения эмоционального фона за период времени, мониторинга настроения толпы.

Целью настоящей работы является классификация способов решения задачи распознавания эмоций из звучащей речи. Для достижения поставленной цели требуется решить ряд задач:

- сформировать классификацию эмоциональных состояний;
- провести обзор информативных признаков, характеризующих речь;
- описать формальную постановку задачи;
- описать существующие решения задачи.

1. КЛАССИФИКАЦИЯ ЭМОЦИЙ

Одной из главных проблем в исследованиях, связанных с определением эмоционального состояния диктора по голосу, является отсутствие четкого 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 информативных признаков, полученных из речи диктора с определенными эмоциями и аффективными состояниями.

Сегодня широко используются три подхода : дискретная, многомерная и гибридная модель.

1.1 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 1.1 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].

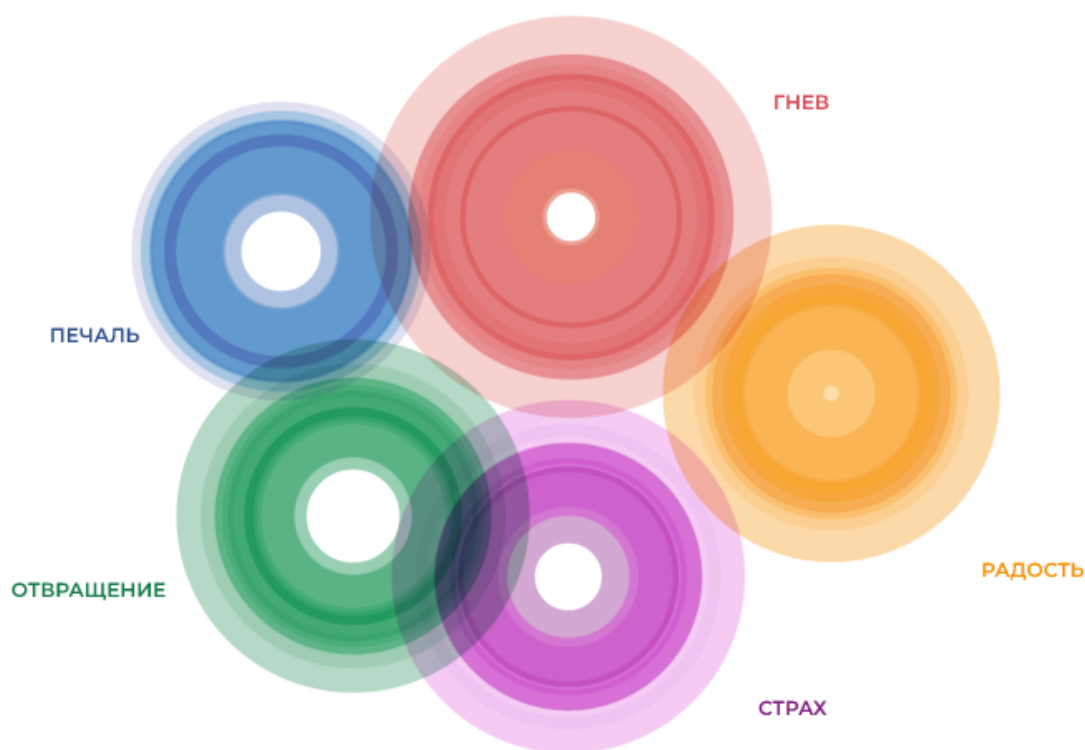


Рис. 1.1 – Атлас эмоций, предложенный П. Экманом

1.2 МНОГОМЕРНАЯ МОДЕЛЬ ЭМОЦИОНАЛЬНОЙ СФЕРЫ

Многомерная модель 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, 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 недовольства 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 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).

1.3 ГИБРИДНАЯ МОДЕЛЬ ЭМОЦИОНАЛЬНОЙ СФЕРЫ

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

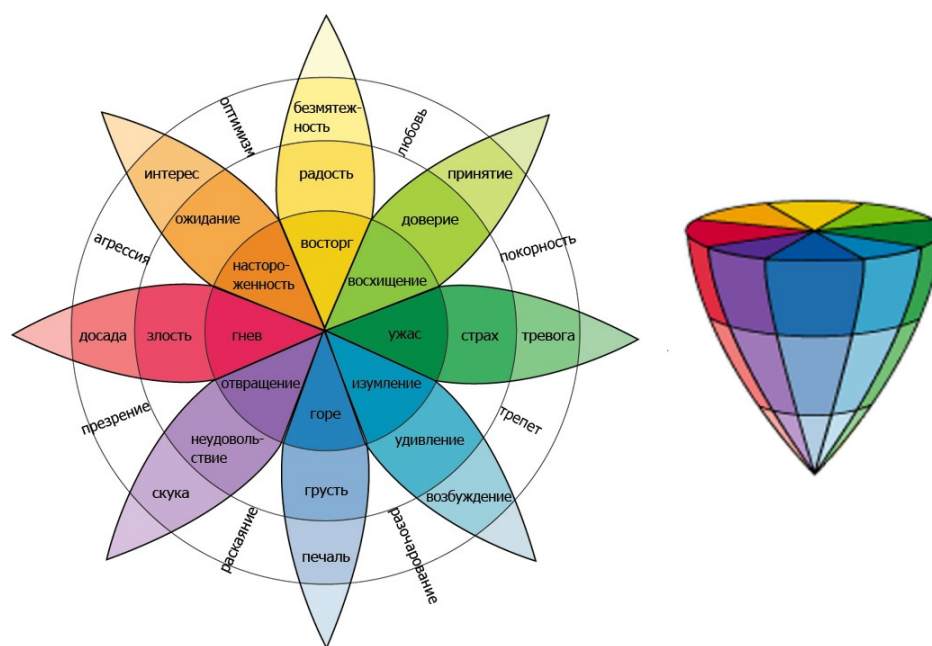


Рис. 1.2 – Трехмерная модель Роберта Плутчика, в основании которой лежим круговая траектория фундаментальных эмоций (слева)

Согласно этой классификации, в отдельной области n -мерного эмоционального пространства различия между эмоциями могут определяться в терми-

нах 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.

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.

2. ИНФОРМАТИВНЫЕ ПРИЗНАКИ, ХАРАКТЕРИЗУЮЩИЕ РЕЧЬ

Условно характеристики речи диктора можно разбить на два основных класса – *акустические* и *лингвистические*. В зависимости от решаемой задачи их относительная эффективность может быть различной. [9]

Акустические характеристики голоса могут быть условно разделены на пять категорий [10]:

- просодические;
- фонационные (отношение гармоник основного тона к шуму, джиттер, шиммер);
- динамические;
- спектральные;
- энергетические (отношение мощностей в спектральных полосах, оценка мощности сигнала).

Каждая группа показателей находит свое применение в описании отдельных аспектов голоса, являющихся показателями различных эмоциональных состояний. На ранних этапах развития систем детектирования эмоций информативность параметров определялась эвристически. В настоящем используются алгоритмы отбора информативных признаков. Существует два класса алгоритмов: «обертки» и «фильтры».

Обертка, или *wrapper based selection* использует оценку работы классификатора в качестве критерия оптимизации внутри замкнутого цикла. Примером является стратегия линейного последовательного поиска – алгоритм начинает с пустого множества и последовательно добавляет к нему наилучшие информативные признаки. [11]

«Фильтры» основаны на методах теории информации либо корреляционном анализе. Критерием оптимизации является некоторая функция, соотносящаяся с корреляциями между информативными признаками, приростом инфор-

мации при их добавлении к набору, определенными метриками в пространстве признаков и статистиками. [12]

2.1 ПРОСОДИЧЕСКИЕ ХАРАКТЕРИСТИКИ

Просодические характеристики – это совокупность темпорального, артикуляционного и интонационного ее компонентов. В таблице 2.1 представлены основные просодические параметры речевого сигнала, связанные с эмоциональным состоянием человека. [13]

Таблица 2.1 – Соответствие между просодическими особенностями речи и эмоциональным состоянием

<i>параметры</i>	<i>высокое значение</i>	<i>низкое значение</i>
изменчивость частоты основного тона	радость, гнев, страх	печаль, безразличие
уровень частоты основного тона	радость, гнев, страх, чувство приподнятости и уверенности в себе	печаль, презрение, скука, безразличие
интенсивность	радость, гнев, презрение, чувство приподнятости, уверенности в себе, силы, эмоциональное оценивание	печаль, презрение, скука, безразличие
темп	радость, гнев, страх, чувство приподнятости, уверенности в себе, безразличия	печаль, презрение, скука

Наиболее устойчивыми при создании системы распознавания эмоций являются изменчивость частоты основного тона и темп, использование параметра интенсивности связано с аппаратурными параметрами ввода сигнала в систему обработки, поэтому необходимо использовать некоторую относительную вели-

чину, например, отношение интенсивности гласных и согласных звуков. [14]

2.2 СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ

Спектральная группа характеристик основана на преобразовании Фурье. В частности, мел-частотный анализ.

Мел-частотный анализ представляет частоты речи с позиции психоакустического параметра слуха – высоты тона. [15] Высота тона определяет, насколько высоким или низким кажется тон слушателю. Нелинейную связь между частотой звука и его высотой отображает мел-частотная шкала (рисунок 2.1).

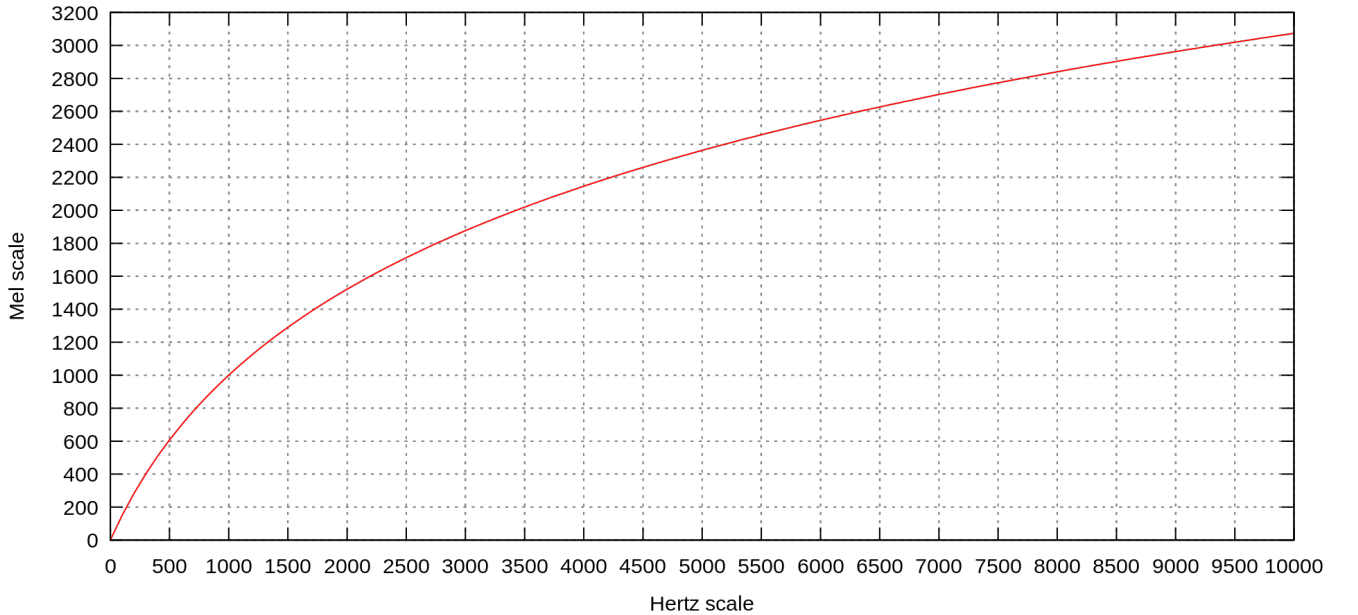


Рис. 2.1 – Мел-частотная шкала

Принято, что высота звука частотой 1000 Гц при уровне 40 дБ равна 1000 мел. [15] Перевод частоты из герц в мел осуществляется по формуле 2.1:

$$Mel(f) = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right), \quad (2.1)$$

где f – частота (Гц), Mel – частота (мел).

Гребенка фильтров для мел-частотных кесперальных коэффициентов – набор треугольных окон в мел-шкале. Поскольку высота тонов, начиная с частоты

больше 1 кГц, возрастает гораздо медленнее, высокочастотные фильтры гребенки имеют большую полосу пропускания, чем фильтры низких частот. При такой фильтрации может быть упущена значимая информация. Однако, существует комплиментарный метод, использующий обратные мел-частотные кепстральные коэффициенты.

Также набор треугольных фильтров используется в нахождении линейно-частотных кепстральных коэффициентов (*linear frequency cepstral coefficients*), но фильтры расположены равномерно по линейной полосе частот. Актуальность данного метода объясняется тем, что по теории речеобразования строение речевого тракта, и в частности его длина, отображается в высокочастотной области спектра, которой мало внимания уделяется при использовании мел-частотных кепстральных коэффициентов. [16]

2.3 ДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ

Среди динамических характеристик, применяющихся в современных системах распознавания речи можно назвать звонкость и сонорность.

Характеристика звонкости. Использование характеристики звонкости может привести к улучшению результатов распознавания за счет лучшей дифференциации фонем. Динамическая характеристика звонкости (основного тона) – величина, выражающая насколько периодическим является речевой сигнал в момент времени t . Для измерения периодичности используют автокорреляционную функцию. Автокорреляция $R^t(\tau)$ выражает схожесть между временным окном и его копией, смещенной на τ .

Характеристика сонорности. Сонорность фонемы можно определить как степень участия ее шумовых составляющих. Для получения значения динамического коррелянта предлагается [17] использовать производную спектра в частотной области (2.2):

$$S_t^{(i)} = \log \left(\sum_{n=0}^{N/2} |a_t^{(i)}[n]| \right), \quad (2.2)$$

где $a_t^{(i)}[n]$ – производная спектра i -го порядка нормализованного спектра $\hat{X}_t[n]$.

3. СИСТЕМА РАСПОЗНАВАНИЯ РЕЧЕВЫХ ЭМОЦИЙ

3.1 ПОСТАНОВКА ЗАДАЧИ

Задача заключается в выявлении эмоциональной составляющей речи по акустическим и лингвистическим признакам и распознавание смысла эмоции индивида. Эти эмоции, являясь непроизвольной реакцией индивида на попытку исследователя тем или иным образом прояснить исследуемую ситуацию, должны были бы позволить сделать заключения в отношении сказанного испытуемым и прояснить как позицию испытуемого в ситуации дознания.

Система распознавания эмоций на основе речи соотносит исходные данные на входе – речевой сигнал, к определённом классу на выходе – виду эмоции, с помощью выделения существенных признаков – речевых особенностей, характеризующих эти данные.

В работе системы можно выделить четыре основных этапа: предварительная обработка речевого сигнала, выделение его характерных особенностей, предварительная обработка особенностей речевого сигнала перед классификацией и классификация.

3.2 ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА СИГНАЛА

Предварительная обработка включает в себя аналоговоцифровое преобразование (АЦП) входного сигнала для получения его спектральных составляющих; цифровую фильтрацию сигнала, удаление неречевых компонентов сигнала и. т. д. [18]

Под термином *цифровая фильтрация* обычно понимают локальную цифровую обработку сигнала скользящим окном. При этом полагают, что размер окна много меньше размера выборки обрабатываемого фрагмента сигнала. Для каждого положения окна, за исключением, возможно, небольшого числа крайних точек выборки, выполняются однотипные действия, которые определяют отклик, или выход фильтра. Различают линейную и нелинейную цифровую фильтрацию. [19]

Линейная цифровая система описывается уравнением свертки (3.1):

$$y[n] = \sum_{l=-\inf}^{\inf} h_l x[n-l], \quad (3.1)$$

где $x[n]$ – входная выборка, $y[n]$ – выходная выборка, h_l – импульсная характеристика системы.

Примером нелинейной цифровой фильтрации являются семейства порядковых фильтров. Они широко используются в задачах цифровой обработки сигналов и изображений, в частности для обнаружения объектов, выделения границ, подавления импульсных помех. Отклик порядкового p -фильтра определяется как p -я порядковая статистика [20, 24, 25], т. е. элемент под номером p , где p – одно из чисел $\{0, 1, \dots, N-1\}$, где N – размер апертюры фильтра в вариационном ряду, полученном из выборки исходных данных.

Оцифрованные образцы речи сначала нормализуются. Далее образцы сегментируются на кадры продолжительностью 30 мс с перекрытием 10 мс с использованием окна Хэмминга, которое имеет следующий вид (3.2):

$$w_n = 0,54 - 0,46 \cdot \cos\left(2\pi \frac{n}{N-1}\right), \quad n = 0, \dots, N-1, \quad (3.2)$$

где N – длина окна, выраженная в отсчетах.

3.3 ПРЕДОБРАБОТКА ХАРАКТЕРНЫХ ОСОБЕННОСТЕЙ

Векторы признаков, полученные на предыдущем этапе, перед подачей их на вход выбранному классификатору нормализуются; содержащие более 2–10% нулевых значений, отбрасываются. [20] Так как на большинство классификаторов негативно влияет избыточность, то для уменьшения размерности итогового вектора признаков используются различные алгоритмы выбора признаков (прямой выбор признаков, генетический алгоритм, последовательный прямой плавающий поиск).

3.4 КЛАССИФИКАЦИЯ

Распознавание голоса отличается от многих систем распознавания тем, что в данном случае предметом является процесс, а не статическое изображение, как в случае с распознаванием, например, изображений. Поэтому чаще всего образец голоса представляется не в виде единого вектора признаков, а в виде последовательности векторов признаков, каждый из которых описывает характеристики небольшого участка речевого сигнала. Последовательность векторов, полученная после этапа обработки сигнала, используется для построения шаблона диктора или для осуществления сравнения с уже построенными шаблонами. Точность классификации в значительной мере зависит от выбранного типа классификатора.

3.4.1 СКРЫТЫЕ МАРКОВСКИЕ МОДЕЛИ

Основой для скрытой марковской модели является марковская цепь. Марковская цепь задаётся начальным распределением вероятностей $p^0(x)$ и вероятностями перехода $T(x', x)$

$T(x', x)$ – это распределение следующего элемента цепи в зависимости от следующего; распределение на $(t + 1)$ -м шаге описывается согласно 3.3:

$$p^{t+1}(x') = \int T(x'; x)p^t(x)dx. \quad (3.3)$$

В дискретном случае $T(x', x)$ – это матрица вероятностей $p(x' = i \mid x = j)$.

Марковская модель – статистическая модель, имитирующая работу процесса, похожего на марковский процесс с неизвестными параметрами, и задачей ставится определение неизвестных параметров на основе наблюдаемых.

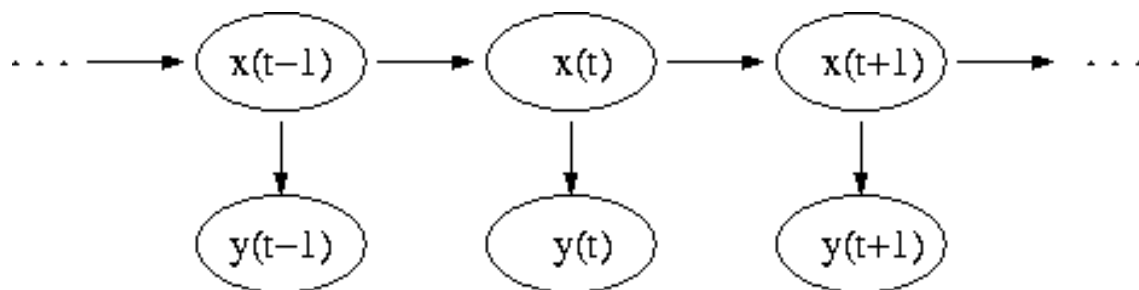


Рис. 3.1 – Скрытая марковская модель

На рисунке 3.1 $x(t)$ — сам процесс (модель), а $y(t)$ — наблюдаемые неизвестные параметры.

Главное свойство скрытой марковской модели — следующее состояние зависит только от предыдущего состояния 3.4:

$$p(x(t) = x_j \mid x(t-1) = x_{j_{t-1}}, \dots, x(1) = x_{j_1}) = p(x(t) = x_j \mid x(t-1) = x_{j_{t-1}}). \quad (3.4)$$

Здесь вероятность $a_{ij} = p(x(t) = x_j \mid x(t-1) = x_i)$ не зависит от времени t . Вероятности a_{ij} составляют матрицу перехода $A = (a_{ij})$.

При распознавании речи наблюдается не $x(t)$, т. е. реальные состояния модели, а $y(t)$ — некоторая функция от них.

3.4.2 МЕТОД К-БЛИЖАЙШИХ СОСЕДЕЙ

В качестве шаблона диктора в данном методе используется полный набор векторов обучающей последовательности. Сравнение образца с таким шаблоном происходит следующим образом: каждый вектор тестовой последовательности сравнивается с каждым вектором шаблона для определения минимального расстояния. Полученные расстояния усредняются для формирования итоговой оценки (3.5):

$$d_k = \sum_{i=1}^L \min_{x_j \in S_{p_k}} d(x_i, x_j) \quad (3.5)$$

Для снижения вычислительной трудоёмкости используют различные методы сокращения шаблона либо методы сохранения данных для ускорения поиска, такие, как, например, kd-дерево или другие методы. [21]

3.4.3 МОДЕЛЬ ГАУССОВЫХ СМЕСЕЙ

Модель представляет собой взвешенную сумму Гауссиан (3.6):

$$p(x|\lambda) = \sum_{i=1}^M w_i p_i(x), \quad (3.6)$$

где λ – модель диктора, M – количество компонентов модели, w_i – веса компонентов, такие, что (3.7):

$$\sum_{i=1}^M w_i = 1. \quad (3.7)$$

Функция плотности вероятности каждого компонента вычисляется следующим образом:

$$p_i(x) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left(-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right), \quad (3.8)$$

где D – размерность пространства признаков, μ_i – вектор математического ожидания, Σ – матрица ковариации.

ЗАКЛЮЧЕНИЕ

Распознавание эмоций по звучащей речи – алгоритмически сложная задача, включающая в себя ряд подзадач, классификация которых была проведена в представленной работе.

На рис. 3.2 приведены основные признаки, по которым можно классифицировать системы распознавания.

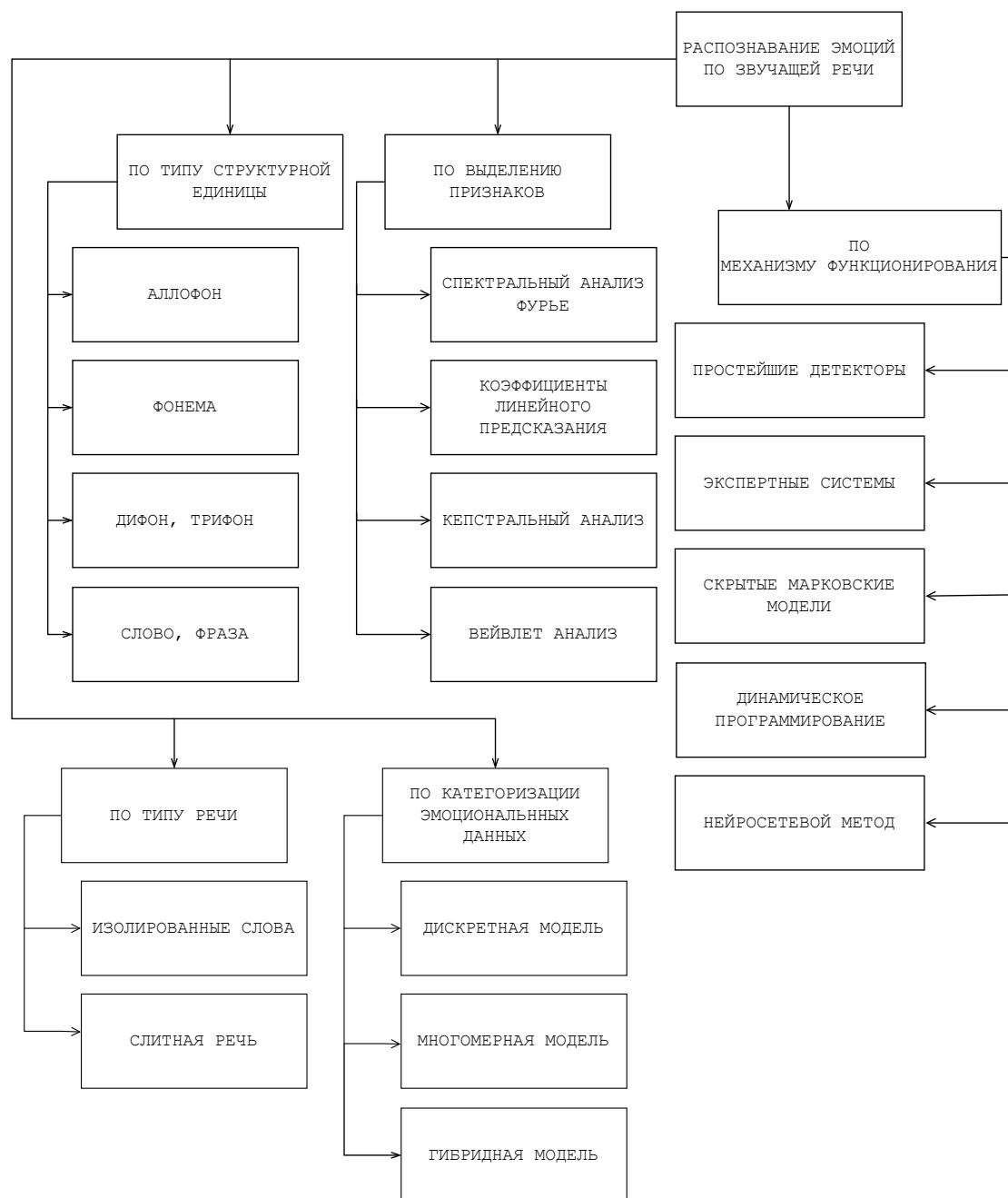


Рис. 3.2 – Классификация систем распознавания речи

Предложена структура системы распознавания эмоций по звучащей речи.

1. Проводится предобработка поступившего на вход аудиофайла.
2. Извлекаются значимые признаки речевого сигнала.
3. Проводится предобработка извлеченных признаков.
4. Классификатор принимает решение о виде эмоции на основе полученного вектора признаков.

На точность работы систем распознавания оказывает влияние множество факторов. Прежде всего, влияние окружающей среды, а также изменение условий записи.

Корпуса, которые используются для экспериментальной оценки, не всегда способны смоделировать перечисленные ситуации. Поэтому результат существенно зависит от того, насколько представительна база. Результат также зависит от продолжительности материала, используемого в каждом тесте и для создания моделей, и от количества пользователей в корпусе.

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. 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, In Press. — 2011.
10. *Ayadi E., Kamel M., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern Recognition. — 2011.
11. *Pudil P., Novovicova J. and Kittler J.* Floating search methods in feature selection // Pattern Recognition. — 1994.
12. *Hall M. A.* Correlation-based feature selection for machine learning // Ph. D. Thesis, Hamilton, NZ: Waikato University, Department of Computer Science. — 1998.

13. *Дорохина Г. В.* Методы пофонемного распознавания, использующие свойства языка и речи // Искусственный интеллект №4 – – Донецк: ИПШ «Наука і освіта». — 2008.
14. *Фёдоров В., Юрков П.* Распознавание эмоционального состояния человека по акустическим параметрам речи // Третий междисциплинарный семинар «Анализ разговорной русской речи». — 2009.
15. *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.
16. Linear versus mel frequency cepstral coefficients for speaker recognition. / X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, S. Shamma // Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2011). — 2011.
17. *Комаров Д.* Использование фонетических характеристик для автоматического распознавания речи // Вестник Санкт-Петербургского университета. Язык и литература. — 2007. — № 3—I. — С. 48—55.
18. Acoustic Features Extraction for Emotion Recognition / J. Rong, Y.-P. P. Chen, M. Chowdhury, G. Li. — 2007. — Янв. — DOI: 10.1109/ICIS.2007.48.
19. *Ахмад Х. М., Жирков В. Ф.* Введение в цифровую обработку речевых сигналов: учебное пособие. — 2007.
20. *Брановицкий А., Борисов Д.* Система автоматического распознавания эмоций по речевому сигналу. — 2019.
21. The DET curve in assessment of detection task performance : тех. отч. / A. Martin, G. Doddington, T. Kamm, M. Ordowski, M. Przybocki ; National Inst of Standards ; Technology Gaithersburg MD. — 1997.