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 «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

«Классификация методов извлечения речевых характеристик для задачи распознавания речи у людей с дефектами речи»

Студент __ИУ7-52Б__ _____ ___И.В.Козлова___
(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель _____ ____К.А.Кивва____
(Подпись, дата) (И.О.Фамилия)

2021 г.

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)

УТВЕРЖДАЮ

Заведующий кафедрой ИУ-7
_____ И. В. Рудаков
« ____ » _____ 20 ____ г.

З А Д А Н И Е
на выполнение научно-исследовательской работы

по теме _____*Классификация методов извлечения речевых характеристик для задачи распознавания речи у людей с дефектами речи*_____

Студент группы __ИУ7-52Б____

_____Козлова Ирина Васильевна_____
(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

_____учебная_____

Источник тематики (кафедра, предприятие, НИР) ___НИР_____

График выполнения НИР: 25% к 4 нед., 50% к 7 нед., 75% к 11 нед., 100% к 14 нед.

Техническое задание *Рассмотреть классификацию систем автоматического распознавания речи, а также возможные дефекты и нарушения речи. Изучить какими бывают речевых характеристики и методы извлечения частотных характеристик. Составить классификацию методов извлечения частотных характеристик.*

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на 15-25 листах формата А4.
Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)
Презентация на 8-10 слайдах.

Дата выдачи задания « 09 » сентября 2021 г.

Руководитель НИР _____ ____К.А.Кивва____
(Подпись, дата) (И.О.Фамилия)

Студент _____ ____И.В.Козлова___
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

Содержание

Введение

В современном мире существует множество технических средств, которые могут воспринимать произносимые речевые сообщения: мобильные телефоны, автомобили, компьютеры и др. Создание приложений, с помощью которых машины могут разговаривать с человеком, особенно правильно реагируя на разговорную речь, давно начало интересовать ученых и инженеров. Однако в настоящее время такая технология оптимизирована не для всех пользователей.

Автоматическое распознавания речи – это процесс, в котором из входного речевого сигнала извлекаются необходимые признаки (речевые характеристики), затем с помощью этих признаков определяются слова/фразы, которые поступили на вход, структурная схема системы автоматического распознавания речи представлена на рисунке 1. [1] [2]

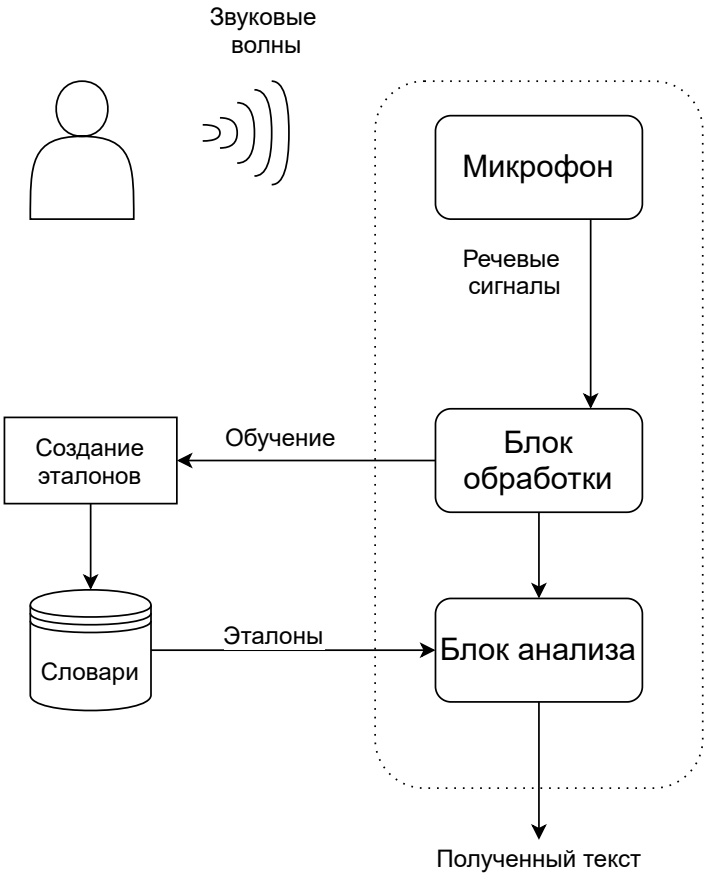

Рисунок 1 – Структурная схема системы автоматического распознавания речи

Выделение признаков (речевых характеристик) происходит из звукового сигнала, который передается машине. Такой звуковой сигнал, или речевой,

так как он передает речь человека, является нестационарным процессом, значения параметров которого непрерывно меняются. Поэтому оценки такого спектра берутся на последовательных кадрах в предположении, что на таких кадрах сигнал является стационарным и характеризуется постоянными фиксированными параметрами. [3]

Цель данной научно-исследовательской работы – провести обзор методов извлечения речевых характеристик для задачи распознавания речи у людей с дефектами речи.

Для достижения поставленной цели необходимо решить следующие задачи:

- рассмотреть классификацию систем автоматического распознавания речи;
- рассмотреть возможные дефекты и нарушения речи;
- изучить существующие речевые характеристики;
- изучить методы извлечения частотных характеристик.

1 Анализ предметной области

1.1 Классификация САРР

Системы автоматического распознавания речи (САРР, англ. ASR – Auto matic Speech Recognition) помогают машинам интерпретировать устную речь и автоматизировать задачи человека, например поиск в интернете, набор текста и тд. Одним из наиболее сложных моментов в разработке таких систем является довольно широкая междисциплинарность задачи, то есть затрагиваются вопросы теории обработки сигналов, математического анализа, психологии, теории коммуникаций, а также лингвистики.

Системы автоматического распознавания речи можно классифицировать по основным аспектам [4].К таким аспектам можно отнести следующие.

1. Тип речи.
   (a) Спонтанная речь.

   Спонтанную речь можно рассматривать как речь, которая звучит естественно (с эмоциями, с внезапными паузами).

   (b) Непрерывная речь.

   Непрерывная речь – это обычная человеческая речь без пауз между словами. Этот вид значительно затрудняет машинное понимание речи.

   (c) Связные слова.

   Такие слова требуют минимальные паузы между высказываниями. Речь должна течь плавно.

   (d) Изолированная речь.

   Такие системы направлены на распознавания конкретных голосовых команд, получаемых от пользователя.

2. Размер словаря.
   (a) Маленькие словари.

   САРР с маленькими словарями (чаще до 500-1000 слов) необходимы для распознавания команд, получаемых от пользователя.

(b) Большие словари.

Такие словари чаще всего используются в САРР слитной речи. Размеры достигают до десятков тысяч слов [5].

3. Персонализация.

(a) Дикторозависимые.

К классу систем, зависящих от диктора относятся системы, которые требуют предварительного обучения и в его процессе настраиваются на определенного диктора. При смене диктора в таких системах возникает необходимость полной перенастройки.

(b) Дикторонезависимые.

К классу систем, независимых от диктора относятся системы, которые работают вне зависимости от того, кто выступает в качестве диктора. Данные системы имеют возможность распознания речи любого диктора и не нуждаются в предварительном обучении.

4. Структурные единицы.

В качестве структурных единиц могут выступать фразы, слова, фонемы. Системы, которые распознают речь, используя целые слова или фразы, называются системами распознавания речи по шаблону. Создание таких систем менее трудоемко, чем системы основанные на базе выделения лексических элементов (в таких системах структурными единицами являются фонемы).

5. Принцип выделения структурных единиц.

В современных САРР используются несколько подходов для выделения структурных единиц из потока речи.

(a) Фурье-анализ.

Данный анализ предполагает разложение исходной периодической функции в ряд, в результате чего исходная функция может быть представлена как суперпозиция синусоидальных волн различной частоты [6].

(b) Вейвлет-анализ (от англ. wavelet – «маленькая волна»).

Данный анализ раскладывает исходный сигнал в базис функций, которые характеризуют как частоту, так и время [6].

(c) Кепстральный анализ.

Данный анализ основан на выделении кепстральных коэффициентов на мел[1]-шкале, называемых мел-частотными кепстральными коэффеициентами. Кепстр – это дискретно-косинусное преобразование амплитудного спектра сигнала в логарифмическом масштабе.

6. Механизм распознавания

(a) Скрытые Марковские модели (СММ).

СММ – это модель, состоящая из N состояний, в каждом система может принимать одно из M значений какого-либо параметра. Матрицей $A = [a_{ij}]$ задаются вероятности переходов между состояниями из i состояния в j состояние. Вектором $B = [b_j(k)]$ задается вероятность выпадения какого-либо из M значений параметра в каждом из N состояний (выпадения k значения параметра в j состоянии). Вероятность того, что в начальные момент система окажется в i состоянии определяется вектором $\pi = \pi_i$. Таким образом, СММ называется тройка $\lambda = (A, B, \pi)$.

Модель называется «скрытой», потому что последовательность, в которой пребывала система неважна. Другими словами такая система выступает в роли «черного ящика», на вход которого поступает последовательность параметров, а на выход ожидаем модель, которая с максимальной вероятностью генерирует такую последовательность.

(b) Динамическое искажение времени.

Алгоритм динамического искажения времени (Dynamic Time Warping – DTW) является методикой эластичного сравнения вектора наблюдений с хранящимся шаблоном. По-другому можно сказать, что это мера подобия временных рядов, которая минимизирует эффекты временного сдвига, различного течения времени, а также обеспечивает непрерывное преобразование временных рядов

[1]Мел – единица высоты звука [7].

для того, чтобы обнаружить одинаковые формы с различными фазами.

(c) Нейронные сети.

С помощью нейронных сетей можно создавать самообучаемые и обучаемые системы распознавания речи. Некоторые факторы, которым должны отвечать такие системы: возможность контроля своих действий с последующей коррекцией, разработка системы заключается только в построении архитектуры системы.

7. Назначение

(a) Командные системы.

Такие системы используют распознавания по шаблону (фразе или слову).

(b) Системы диктовки.

Такие системы требуют более точного распознавания, то есть выделение лексических элементов. Также при интерпретации произнесенной фразы система полагается не только на то, что произносилось в данный момент, но и на фразы, сказанные ранее.

Обобщив все выше перечисленное можно представить классификацию систем распознавания речи на рисунке 1.1.

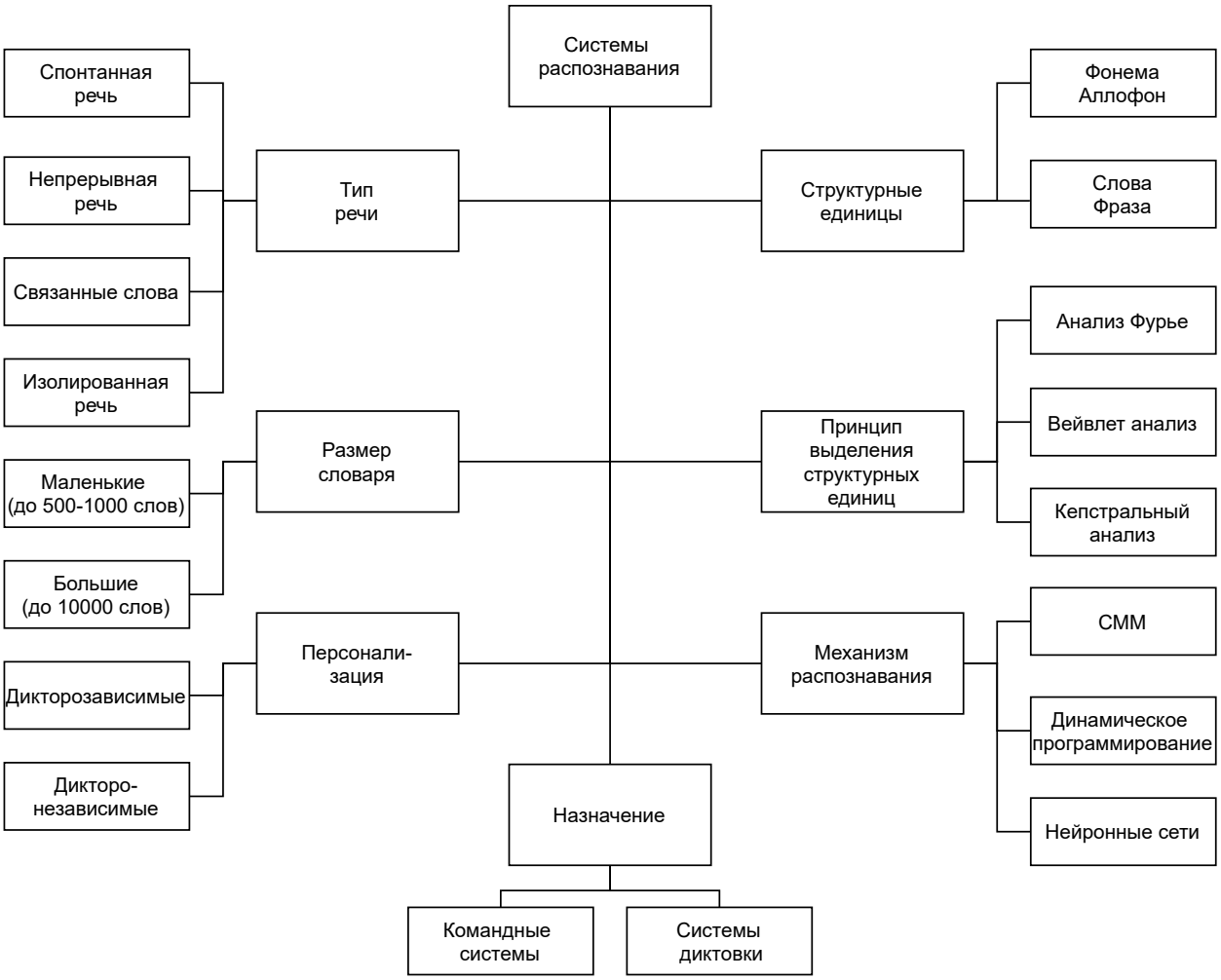

Рисунок 1.1 – Классификация систем распознавания речи

1.2 Сложности в работе САРР

Распознавание речи – это задача, усложненная тем, что речь человека характеризуется высокой степенью изменчивости [8].

Причины этого следующие:

- для одного и того же диктора произношения одних и тех же звуков (слов, фраз) будут отличаться длительностью произношения, интонацией. Часто это связано с изменением физического или эмоционального состояния человека, его настроения или условий, в которых он находится;

- произношение фонем сильно зависит от контекста, например наличие или отсутствие четкой артикуляции при разговоре;

- различные помехи (отражения звука, искажение микрофона, фоновый шум).

Отличием распознавания слитной речи от, например, отдельных команд или подготовленной речи, являются различные сбои в произношении [9] [10]. Очень сложно говорить гладко (не сбиваясь) и красиво оформлять свои мысли (четко и ровно составлять предложения), поэтому можно сказать, что основная особенность слитной речи – это сбивчивость, наличие повторений, пауз, слов в упрощенной форме (разговорный стиль) [11].

Такие особенности зачастую являются препятствием для обработки речи техническими средствами, так как уловить особенности разговорной речи человека довольно сложно машине, поэтому необходимо либо разработать метод, на основе которого машина научиться распознавать речь человека, либо составлять сверхбольшой словарь слов или звуков, что довольно затратно по памяти [11].

1.3 Нарушения и дефекты речи

У людей с дефектами речи помимо выше описанных особенностей есть и другие, не менее важные, поэтому специальные системы распознавания речи должны также распознавать разные виды нарушения устной речи.

Виды нарушения устной речи [12].

1. Нарушения внешнего оформления устной речи.

   (a) Дисфония – отсутствие голоса или расстройство речи вследствие патологических изменений голосового аппарата: происходят различные изменения и нарушения в силе и тембре, выражающееся в охриплости, слабости голоса.

   (b) Брадилалия – медленный темп речи вследствие поражения головного мозга. Речь сильно замедляется, становится нечеткой, растягиваются гласные.

   (c) Тахилалия – быстрый темп речи, часто сопровождающийся повторением или пропуском слов, незамеченным говорящим.

(d) Заикание – нарушение речи, которое характеризуется частым повторением или пролонгацией звуков, слогов, слов, частыми остановками или нерешительностью в речи, разрывающее ее ритмическое течение.

(e) Дислалия – нарушение звукопроизношения при нормальном слухе и нормальной иннервации речевого аппарата, которое проявляется в заменах, искажениях и смешениях звуков родной речи.

(f) Ринолалия – нарушение произносительной стороны речи или тембра голоса, обусловленное анатомо-физиологическим поражением речевого аппарата: струя воздуха проходит не в ротовую, а в носовую полость, в которой происходит резонанс.

(g) Дизартрия – нарушение произносительной стороны речи вследствие поражения центральной нервной системы.

2. Нарушения структурно-семантического оформления.

(a) Алалия – полное отсутствие или недоразвитие речи у детей при нормальном слухе и первично сохранном интеллекте.

(b) Афазия – речевое расстройство уже сформировавшейся речи. Причинами могут быть перенесенные черепно-мозговые травмы, инфекционные заболевания нервной системой.

Также стоит отметить, что нарушения речи могут встречаться в комплексе, например: заикание и дизартрия.

Выше перечисленные нарушения также являются препятствием для обработки речи техническими средствами. В разделе 1.4 будет приведен эксперимент из статьи [13], в котором показано, что машина иногда не понимает, что говорит человек с нарушениями речи.

1.4 Рассмотрение эксперимента

В статье [13] проводится эксперимент работы трех платформ CAPP (Amazon, Google, IBM) для групп людей: с нейродегенеративными (медленно прогрессирующие, наследственные или приобретенные заболевания нервной

системы, ведущими к различным симптомам – к деменции [2], нарушению движения, а в следствие чего к нарушениям речи) заболеваниями и здоровых.

Записи чтения текста были расшифрованы с помощью CAPP и вручную, затем сравнены. Точность расшифровки измерялась как доля верно распознанных слов.

Результат эксперимента ожидаем: точность расшифровки CAPP для здоровых людей выше, чем для людей с заболеваниями (рассматривались 3 группы людей, с различными заболеваниями: с рассеянным склерозом и с атаксией Фридрейха [3]). При этом при увеличении продолжительности болезни, точность расшифровки CAPP снижалась.

Часть результата приведена на рисунке 1.2

Обозначения:

- Группа 1 – группа здоровых людей;
- Группа 2 – группа людей с рассеянным склерозом;
- Группа 3 – группа людей с заболеванием "Атаксия Фридрайха".

[2]Деменция — это синдром, возникающий при поражении головного мозга и характеризующийся нарушениями в когнитивной сфере (восприятие, внимание, узнавание, память, интеллект, речь).

[3]Атаксия Фридрейха — генетическое заболевание, связанное с нарушением транспорта железа из митохондрий и протекающее с преимущественным поражением клеток центральной и периферической нервной системы.

Рисунок 1.2 – Результаты эксперимента из статьи [13]

Как видно из эксперимента, системы распознавания речи работают с ошибками для людей с нейро-дегенеративными заболеваниями.

2 Классификация существующих методов

2.1 Основные характеристики речевых сигналов

Все особенности речевых сигналов можно условно разделить на то, что схоже у определенного состава людей, например, отношение к какой-либо языковой группе, и на то, что индивидуально для каждого человека – особенности, выраженные в физической индивидуальности речи, произношении, тембре голоса. [14] [15]

Для того, чтобы сформировать перечень спектральных[1] характеристик в расчет берется только первая гармоника[2].

В соответствии с измеряемой величиной основные спектральные характеристики разделены на следующие группы. [14] [15] [16]

1. Частотные характеристики.
   (a) Частота основного тона – частота первой гармоники спектра (Гц[3]).
   (b) Период основного тона (мс).
   (c) Количество побочных гармоник (Гц).
   (d) Нижняя и верхняя частота спектра (Гц).
   (e) Частотный диапазон (Гц).
   (f) Частота максимального уровня спектральной плотности (Гц).
2. Энергетические характеристики.
   (a) Нижний уровень громкости речи (дБ[4]).
   (b) Верхний уровень громкости речи (дБ).

[1]Спектр сигнала — это совокупность простых составляющих сигнала с определенными амплитудами, частотами и начальными фазами.[3]

[2]Гармоники — это высокочастотные сигналы, накладываемые на основную частоту, то есть частоту цепи, и которые достаточны для искажения формы волны.[3]

[3]Герц (Гц, Hz) – единица частоты периодических процессов.

[4]Децибел (дб) – относительная единица измерения, соответствующая одной десятой бел (В).

(c) Динамический диапазон (дБ).

(d) Амплитуда основного тона (дБ).

3. Временные характеристики.

(a) Длительность звука речи (мс).

(b) Длительность пауз речи между словами/фразами (мс).

(c) Скорость звуков речи (звуков/с).

(d) Темп речи (слов/мин).

(e) Плотность речи – отношение времени наличия звука к полному времени речевого сигнала (%).

2.2 Методы извлечения речевых характеристик

В системах распознавания речи одну из главных ролей играет извлечение признаков (частотной характеристики), при этом характеристики сигналов возбуждения чаще всего отбрасываются.

Извлечение признаков – это процесс удаления ненужной и избыточной информации и сохранение только полезной информации. Цель такого действия состоит в том, чтобы определить набор свойств (параметры), путем обработки формы сигнала поступившего на вход системе. Извлечение признаков включает процесс преобразования речевых сигналов в цифровую форму и измерение важных характеристик сигнала, например, энергии или частоты, и дополнение этих измерений значимыми производными измерениями. [17] [18].

Методы извлечения признаков удобно применять при обработке речи с неправильным произношением звуков, так как их можно адаптировать, извлекая нужные параметры, которые потребуются в дальнейшем.

2.2.1 Линейно-предсказывающее кодирование

Линейно-предсказывающее кодирование (англ. Linear prediction coding

(LPC)).

Принцип метода линейного предсказания состоит в том, что участок речевого сигнала можно аппроксимировать линейной комбинацией предыдущих участков сигнала. Предполагается, что речь создается возбуждением линейного изменяющегося во времени фильтра (речевого тракта) случайным шумом для невокализованных речевых сегментов или последовательностью импульсов для голосовой речи [19].

Процесс речеобразования описывается линейной системой с переменными параметрами и передаточной функцией[5] (2.6).

$$H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \tag{2.1}$$

где G – коэффициент усиления, a_k – коэффициент предсказания, p – порядок линейного предсказания.

Зависимость n-го отсчета речевого сигнала $s(n)$ от сигнала возбуждения $u(n)$ выражается в виде (2.2)

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{2.2}$$

Линейный предсказатель с коэффициентами a_k представляется в виде системы с сигналом на выходе, который рассчитывается по формуле (2.5)

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{2.3}$$

Суть данного метода заключается в нахождении *линейных коэффициентов предсказания* по речевому сигналу с минимизацией погрешности, которую можно определить по формуле (2.4)

$$e(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k) \tag{2.4}$$

Недостатком этого метода является то, что он сильно зависит от точности

[5]Передаточной функцией называется отношение изображения выходного воздействия к изображению входного при нулевых начальных условиях. [20]

произношения. Поэтому, можно сказать, что данный метод не подходит для решения задачи распознавания речи у людей с дефектами речи.

2.2.2 Мел-частотные кепстральные коэффициенты

Мел[6]-частотные кепстральные[7] коэффициенты (англ. Mel frequency Cepstral Coefficient (MFCC)).

Мел-частотный анализ представляет частоты речи с позиции психоакустического [8] параметра слуха – высоты тона. Высота тона определяет, насколько высоким или низким кажется тон слушателю. Связь между частотой звука и его высотой представлена на рисунке 2.1. [19] [22]

Перевод частоты из Герц в Мел осуществляется по формуле (2.5)

$$Mel(f) = 2595 * \log_{10} 1 + \frac{f}{700}, \tag{2.5}$$

где f – частота в Герцах, Mel – частота в мелах.

Построение признаков MFCC начинается с процедуры разбиения входного сигнала на временные окна небольшой длины.

Пусть N_{FB} – количество фильтров (обычно используют порядка 24 фильтров), (f_{low}, f_{high}) – исследуемый диапазон частот. Тогда данный диапазон переводят в шкалу мел, разбивают на N_{FB} равномерно распределенных частей и вычисляют соответствующие границы в области линейных частот.

Следующие преобразования применяются к каждому кадру.

1. Предварительная фильтрация.

   Суть данного шага – уменьшение негативных эффектов, которые возникают во время обработки звукового сигнала.

2. Примение весовой оконной функции.

   Такая функция применяется с целью уменьшения краевых эффектов, возникающих в результате разбиения сигнала на кадры.

[6]Мел – единица высоты звука.

[7]Кепстр (cepstrum) — это результат дискретного косинусного преобразования от логарифма амплитудного спектра сигнала.

[8]Психоакустика - это наука, изучающая психологические и физиологические особенности восприятия звука человеком. [21]

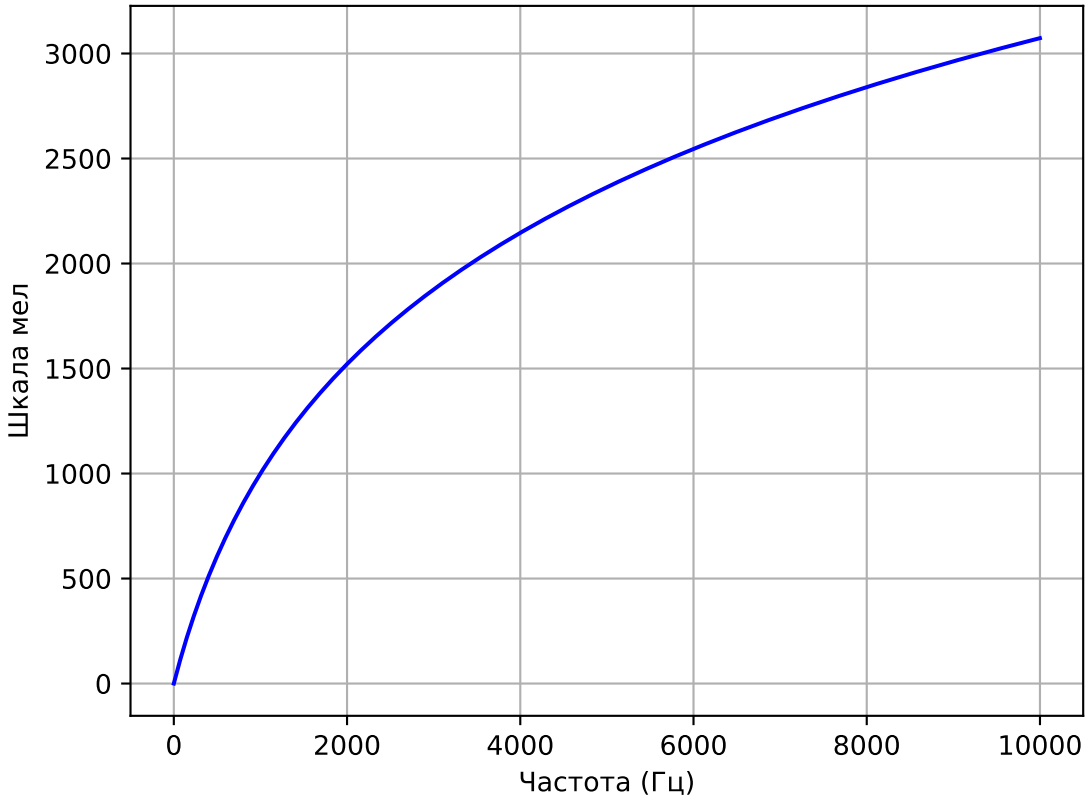

Рисунок 2.1 – Связь между частотой звука и его высотой

3. Подсчет логарифма энергии спектра для набора треугольных мел-частотных фильтров.

   Также стоит отметить, что если использовать вместо мел-шкалы линейное преобразование, то результатом вычислений будут являться так называемые линейно-частотные кепстральные коэффициенты 2.2.3.

4. В конце всех преобразований к коэффициентам приминается дискретное косинусное преобразование. В качестве итоговых значений берутся первые несколько коэффициентов дискретного косинусного преобразования.

Положительные моменты при использовании кепстральных коэффициентов:

- спектр проецируется на специальную Мел-шкалу, позволяя выделить наиболее значимые для восприятия человеком частоты;

- количество вычисляемых коэффициентов может быть ограничено любым значением.

2.2.3 Кепстральные коэффициенты на основе линейного предсказания

Кепстральные коэффициенты на основе линейного предсказания (англ. Linear prediction cepstral coefficient (LPCC)).

Метод LPCC похож на MFCC во многом, но главное отличие в том, что он использует линейную шкалу перевода частоты звука в его высоту, воспринимаемую мозгом. Этот способ хорошо работает в области низких частот, так как в этой зоне зависимость высоты звука от его частоты практически линейна. Данная особенность позволяет достичь схожих результатов при извлечении признаков в области низких частот. [19]

Суть линейного предсказания заключается в том, что линейной комбинацией некоторого количества предшествующих отсчетов можно аппроксимировать текущий отсчет, то есть

$$x(n) = \sum_{k=1}^{p} a_k x_{n-k}, \tag{2.6}$$

где a_k – коэффициент предсказания, p – порядок линейного предсказания.

На основе полученных коэффициентов линейного предсказания рассчитываются кепстральные коэффициенты по формуле (2.7). Причем таких коэффициентов может быть сгенерировано больше, чем самих коэффициентов линейного предсказания.

$$c_n = \begin{cases} a_n + \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k} & 1 \leq n \leq p \\ \sum_{k=n-p}^{n-1} \frac{k}{n} c_k a_{n-k} & n > p \end{cases} \tag{2.7}$$

Например, для сигнала может быть использовано около 12 коэффициентов линейного предсказания, из которых может быть получено порядка 18 кепстральных коэффициентов. [22]

2.2.4 Дискретное вейвлет-преобразование

Дискретное вейвлет-преобразование (Discrete Wavelet Transform (DWT)).

Для наиболее информативного анализа сложных реальных сигналов необходима обработка как по частотным, так и по временным характеристикам, а также достоверное представление уровней детализации для обнаружения закономерностей.

При обработке данных в современных пакетах математической обработки данных может выполняться дискретизированная[9] версия непрерывного вейвлет-преобразования с заданием дискретных значений параметров (a, b) вейвлетов[10] с произвольным шагом Δa и Δb.

Результатом является избыточное количество коэффициентов, которое намного превосходит число коэффициентов, которые могут использоваться для выявления тонких особенностей исследуемых сигналов.

Дискретное вейвлет-преобразование оперирует с дискретными значениями параметров a и b, которые задаются, как правило, в виде степенных функций (2.8)

$$a = a_0^{(-m)}, b = k * a_0^{(-m)} \tag{2.8}$$

где m – параметр масштаба, k – параметр сдвига.

Число использованных вейвлетов по масштабному коэффициенту m задает уровень декомпозиции сигнала, при этом за нулевой уровень $(m = 0)$ обычно принимается сам сигнал.

Недостатком вейвлет-преобразований можно считать их относительную сложность расчетов.

Достоинством можно считать детальный разбор поступающего сигнала, что позволяет выявить множество тонких моментов. Именно поэтому данный метод подходит для решения задачи распознавания речи у людей с дефектами речи больше, чем методы, описанные в разделах 2.2.2 и 2.2.3.

На рисунке 2.2 приведена классификация ранее рассмотренных методов извлечения частотных характеристик.

[9] Дискретизация — в общем случае — представление непрерывной функции дискретной совокупностью ее значений при разных наборах аргументов. [23]

[10] «Вейвлет» (wavelet) в переводе с английского означает «маленькая (короткая) волна».

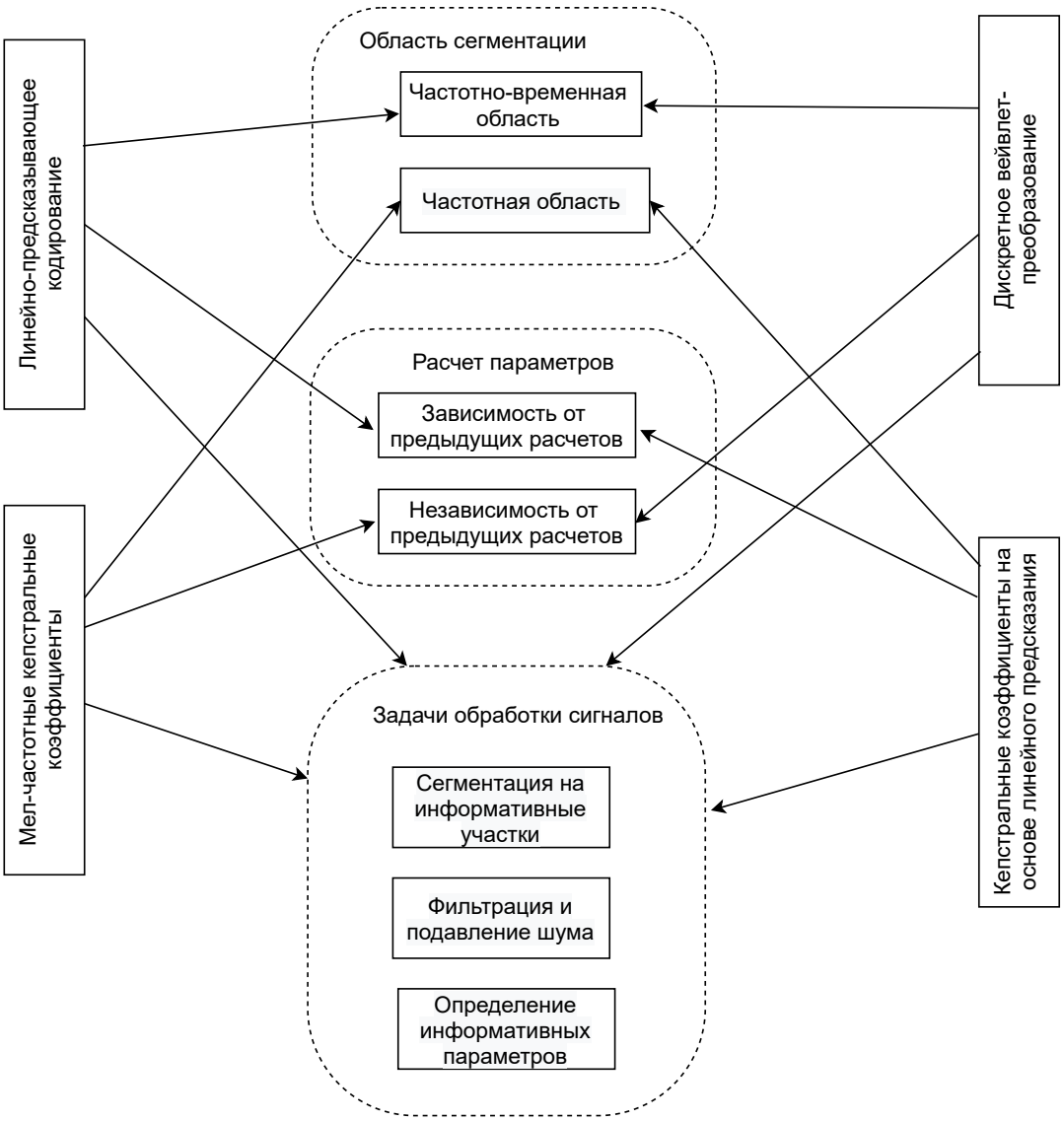

Рисунок 2.2 – Классификация методов извлечения частотных характеристик

Заключение

В ходе выполнения данной работы были рассмотрены классификации систем автоматического распознавания речи, возможные дефекты и нарушения речи, изучены существующие речевые характеристики, методы извлечения частотных характеристик.

Среди методов извлечения было уделено внимание коэффициентам линейного предсказания, MFCC-коэффициентам и LPCC-коэффициентам, а также вейвлет-преобразованию. Мел-частотные кепстральные коэффициенты в последнее время приобрели высокую популярность и достаточно эффективно используются в задаче автоматического распознавания речи. Дискретное вейвлет-преобразование же дает возможность расчета большего количества признаков, что позволит выделить из сигнала достаточно информации для дальнейшей работы.

На основе проделанной работы можно сделать вывод о том, что наиболее подходящий метод извлечения признаков для решения задачи распознавания речи у людей с дефектами речи является дискретное вейвлет-преобразование, так как данный метод позволяет рассчитать большое количество коэффициентов поступившего на вход сигнала для определения более полной информации о нем.

Литература

1. Bhuvaneshwari Jolad D. R. K. Different feature extraction techniques for automatic speech recognition // International Journal of Engineering Sciences, Research Technology. 2018. February. P. 181–188.

2. Shreya Narang M. D. G. Speech Feature Extraction Techniques // International Journal of Computer Science and Mobile Computing. 2015. March. P. 107 – 114.

3. Б.П. Бойко В.А. Тюрин. Спектр сигнала. Казанский (приволжский) федеральный университет институт физики, 2014. с. 38.

4. Федосин С.А. Еремин А. Ю. Классификация систем распознавания речи // ГОУВПО «Мордовский государственный университет им. Н. П. Огарева». 2018. С. 1–4.

5. Д.Н. Бабин И.Л. Мазуренко А.Б. Холоденко. О перспективах создания системы автоматического распознавания слитной устной русской речи. С. 10–15.

6. А.П.Зубаков. Фурье и Вейвлет-преобразования в проблемах распознавания речи // Вестник ТГУ. 2010. т.15, вып.6. С. 1893–1899.

7. А.К.Алимурадов А.Ю.Тычков А.П.Зарецкий А.П.Кулешов. Способ определения кепстральных маркеров речевых сигналов при психогенных расстройствах // Труды МФТИ. 2017. Том 9, №4. С. 201–214.

8. Taabish G. A. S. A Systematic Analysis of Automatic Speech Recognition: An Overview // International Journal of Current Engineering and Technology. 2014. E-ISSN 2277 – 4106, P-ISSN 2347 - 5161. P. 1664–1675.

9. R. P. K. R. Continuous Speech Recognition // Asian Journal of Computer Science And Information Technology. 2014. 62 - 66. P. 62–66.

10. В.О. Верходанова А.А. Карпов. Моделирование речевых сбоев в системах автоматического распознавания речи. С. 10–15.

11. А.А. Леонович. Проблемы распознавания слитной речи // Цифровая Обработка Сигналов. 2007. No4. С. 1664–1675.

12. А.Н. Смолина. Дефекты речи // Эффективное речевое общение (Базовые компетенции). 2014. С. 121–122.

13. S. B. G. V. Automatic speech recognition in neurodegenerative disease // International Journal of Speech Technology. 2021. May. P. 771–780.

14. Столбов М. Б. Основы анализа и обработки речевых сигналов. Университет ИТМО, 2021. с. 106.

15. Поддубный С.А. Иванушкин М.И. Основные спектральные характеристики речевых сигналов и их определение // КВВУ им. генерала армии С. М. Штеменко. 2020. С. 1–9.

16. Х.М.Ахмад В.Ф. Жирков. Введение в цифровую обработку речевых сигналов. Федеральное агентство по образованию Государственное образовательное учреждение высшего профессионального образования Владимирский государственный университет, 2007. с. 190.

17. Praphulla A. Sawakare R. R. Speech Recognition Techniques // International Journal of Scientific, Engineering Research. 2015. Volume 6, Issue 8, August-. P. 1664–1675.

18. Kishori R. R. R. Feature Extraction Techniques for Speech Recognition // International Journal of Scientific, Engineering Research. 2015. Volume 6, Issue 5, Ma. P. 143–148.

19. А.В.Судьенкова. Обзор методов извлечения акустических признаков речи в задаче распознавания диктора // Сборник научных трудов НГТУ. 2019. № 3–4 (96). С. 139–164.

20. Усынин Ю.С. Теория автоматического управления. Челябинск Издательский центр ЮУрГУ, 2010. с. 176.

21. ВОЛОГДИН Э.И. Слух и восприятие звка. Санкт-Петербург, 2012. с. 36.

22. Первушин Е.А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование. 2011. вып. 24. С. 41–54.

23. Ястребов И.П. Дискретизация непрерывных сигналов во времени. Теорема Котельникова. Нижегородский государственный университет им. Н.И. Лобачевского, 2012. с. 31.