




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

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

Студент: Козлова Ирина Васильевна ИУ7-52Б
Научный руководитель: Кивва Кирилл Андреевич

Цель и задачи



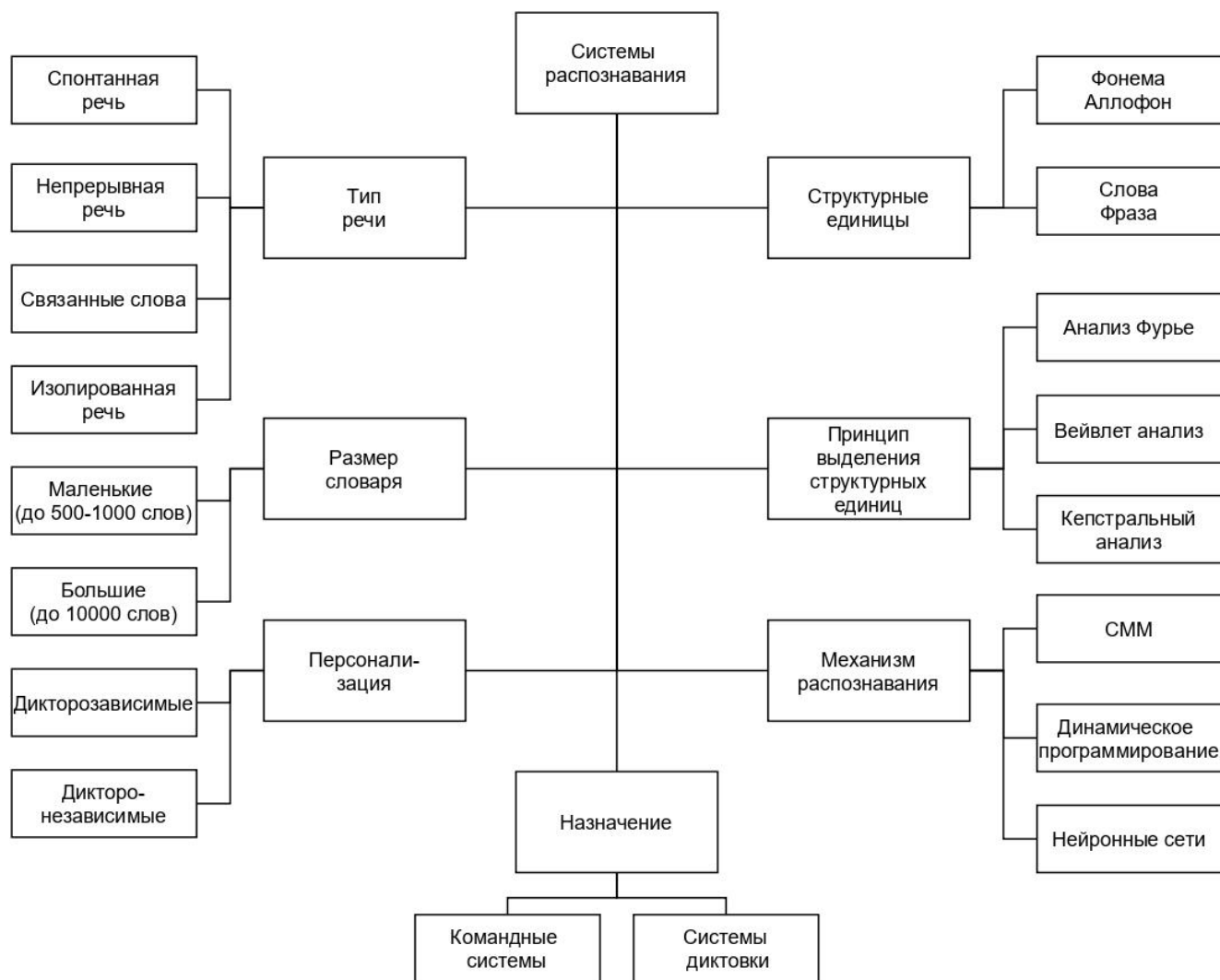
Провести классификацию методов извлечения речевых характеристик для задачи распознавания речи у людей с дефектами речи.

Задачи:

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

Классификация САРР

1. Тип речи
2. Размер словаря
3. Персонализация
4. Структурные единицы
5. Принцип выделения структурных единиц
6. Механизм распознавания
7. Назначение



Нарушения и дефекты речи (1/2)

1. Дисфония – расстройство речи из-за изменений голосового аппарата.
2. Брадилалия – медленный темп речи.
3. Тахилалия – быстрый темп речи.
4. Заикание – частое повторение.
5. Дислалия – нарушение звукопроизношения.

Нарушения и дефекты речи (2/2)

6. Ринолалия – нарушение произношения или темпа голоса.
7. Дизартрия – нарушение произношения из-за поражения нервной системой.
8. Алалия – полное отсутствие или недоразвитие речи.
9. Афазия – расстройство уже сформировавшейся речи.

Основные характеристики речевых сигналов

1. Частотные характеристики

- а. частота и период основного тона
- б. нижняя и верхняя частота спектра

2. Энергетические характеристики

- а. нижний и верхний уровень громкости
- б. амплитуда основного тона

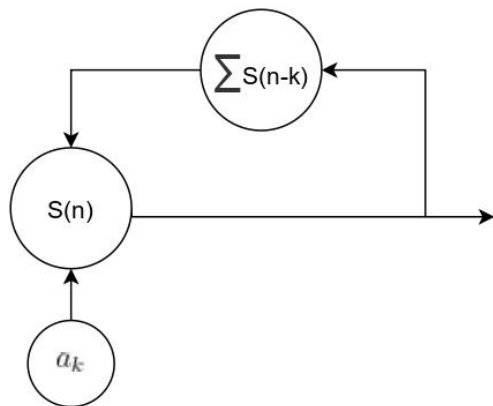
3. Временные характеристики

- а. длительность звука речи
- б. длительность пауз
- с. темп речи
- д. плотность речи

Методы извлечения речевых характеристик (1/3)

1. Линейно-предсказывающее кодирование

Участок речевого сигнала можно аппроксимировать линейной комбинацией предыдущих участков сигнала.



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

$s(n)$ – речевой сигнал

a – коэффициент предсказания

Методы извлечения речевых характеристик (2/3)

2. Мел-частотные кепстральные коэффициенты

Представление частоты речи с позиции психоакустического параметра слуха – высоты тона.

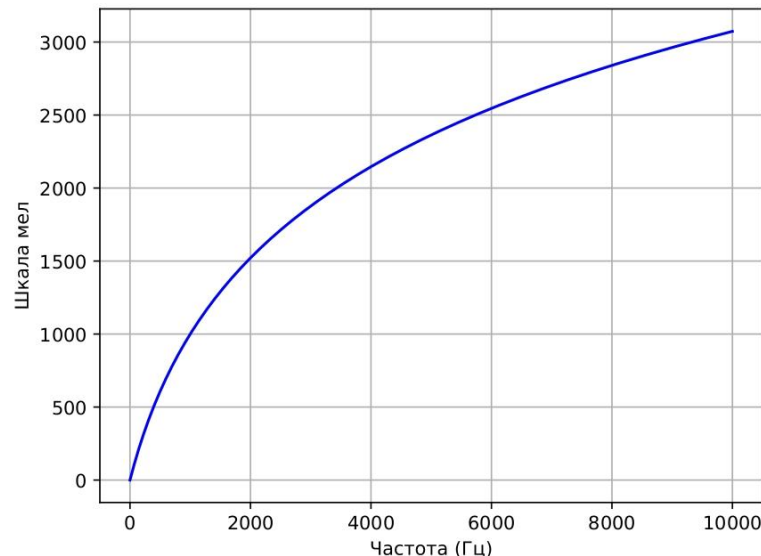
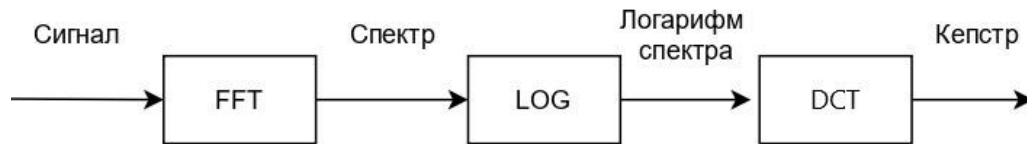


График перевода Герц в Мел

Методы извлечения речевых характеристик (2/2)

3. Кепстральные коэффициенты на основе линейного предсказания

Похож на MFCC, но он использует линейную шкалу перевода частоты звука в его высоту, воспринимаемую мозгом.



FFT – быстрое преобразование Фурье

LOG – логарифм

DCT – дискретное косинусное преобразование

Методы извлечения речевых характеристик (2/2)

4. Дискретное вейвлет-преобразование

Обработка по временным характеристикам, анализ в разном увеличении с разной степенью детализации.

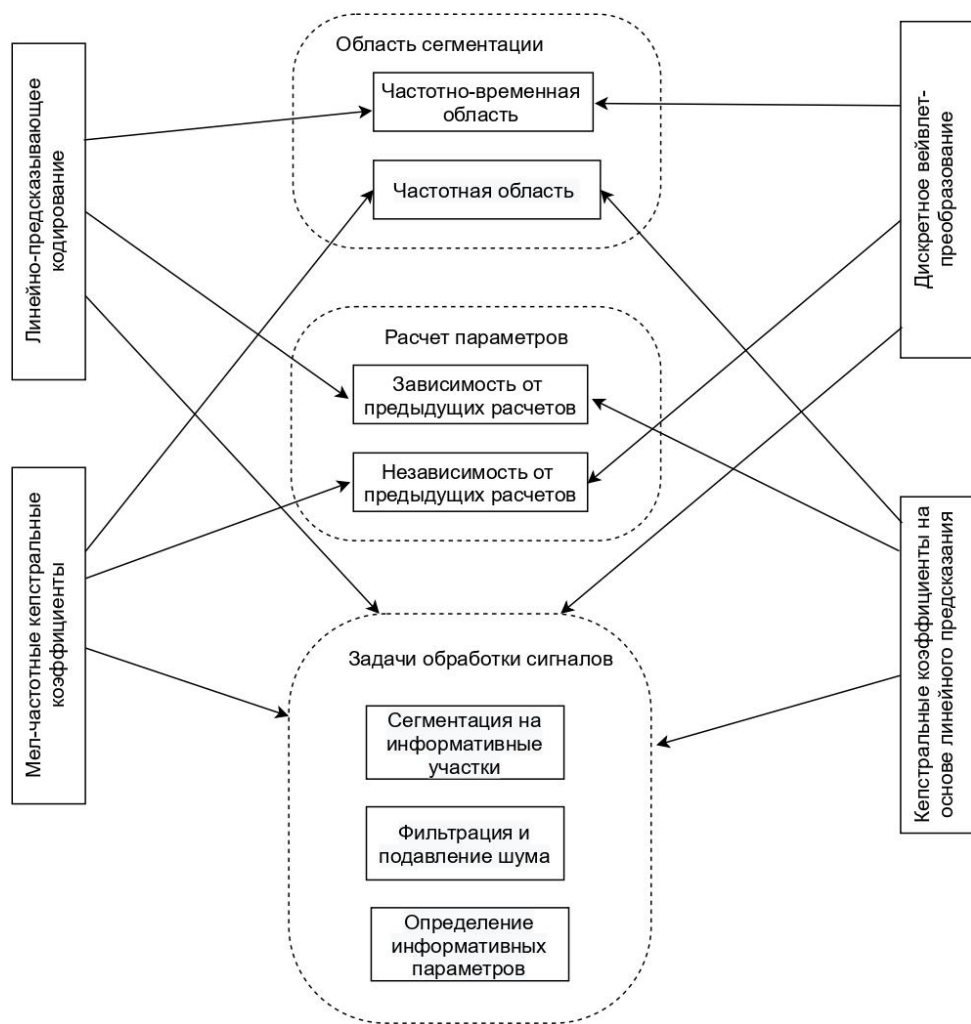
$$\gamma(\tau, s) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t - \tau}{s} \right) dt$$

Вейвлет-преобразование – это преобразование, отображающее функцию $x(t)$, определенную на временной оси переменной t , в функцию двух переменных s и τ .

τ – представляет параллельный перенос

s – масштаб, $\psi(t)$ – материнский вейвлет

Классификация методов извлечения частотных характеристик



Заключение

- рассмотрена классификация систем автоматического распознавания речи;
- рассмотрены возможные дефекты и нарушения речи;
- изучены существующие речевые характеристики;
- изучены методы извлечения частотных характеристик.