



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»

---

## Научно-исследовательская работа

Тема Выбор метода для распознавания слитной речи у людей с дефектом речи

Студент Козлова И.В.

Группа ИУ7-52Б

Научный руководитель Кивва К.А.

Москва — 2021 г.

# 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 не оптимизирована для всех пользователей.

Системы автоматического распознавания речи (САРР, англ. ASR – Automatic 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.

Распознавание речи - это задача, усложненная тем, что речь человека характеризуется высокой степенью изменчивости [1]. 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 четкой артикуляции при разговоре;
- различные помехи (отражения звука, искажение микрофона или системы передачи, фоновый шум).

Отличием распознавания слитной речи от, например, отдельных команд или подготовленной речи, являются различные сбои в произношении. Очень сложно говорить гладко (не сбиваясь) и красиво оформлять

свои мысли (не сомневаясь и не повторяя), поэтому можно сказать, что основная особенность слитной речи - это сбивчивость, наличие повторений, пауз, слов в упрощенной форме (разговорный стиль). 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 разговаривать с человеком, либо составлять сверхбольшой словарь слов или звуков, что довольно затратно по памяти.

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 неправильного произношения звуков, заикание, шепелявость, картавость и др.

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

- Слитная или раздельная речь.
- Размер словаря.
- Диктозависимость.
- 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).

- Принцип выделения структурных единиц.

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

- Фурье-анализ (Жан-Батист Жозеф Фурье – французский математик и физик). Данный анализ предполагает разложение исходной периодической функции в ряд, в результате чего исход-

ная функция может быть представлена как суперпозиция синусоидальных волн различной частоты [2].

- Вейвлет-анализ (от англ. wavelet – ”маленькая волна”). Данный анализ раскладывает исходный сигнал в базис функций, которые характеризуют как частоту, так и время [2].
- Кепстральный анализ. Данный анализ основан на выделении кепстральных коэффициентов на мел-шкале, называемых мел-частотными кепстральными коэффициентами. Кепстр – это дискретно-косинусное преобразование амплитудного спектра сигнала в логарифмическом масштабе. Мел – единица высоты звука [3].

- Алгоритм распознавания

- Скрытые Марковские модели.
- Динамическое программирование.
- Нейронные сети (обучаемые и самообучаемые).

- Назначение

- Командные системы.
- Системы диктовки.

# Литература

- [1] Taabish Gulzar Anand Singh D. K. R. N. F. 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.
- [2] А.П.Зубаков. Фурье и Вейвлет-преобразования в проблемах распознавания речи // Вестник ТГУ. 2010. т.15, вып.6. С. 1893–1899.
- [3] 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.