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 – 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 [1].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 исходная 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 коэффеициентами. Кепрст – это дискретно-косинусное преобразование амплитудного спектра сигнала в логарифмическом масштабе. Мел – единица высоты звука [3].

- Назначение
  - Командные системы.
  - Системы диктовки.

Распознавание речи - это задача, усложненная тем, что речь человека характеризуется высокой степенью изменчивости [4]. 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, фоновый шум).

Отличием распознавания слитной речи от, например, отдельных команд или подготовленной речи, являются различные сбои в произношении. [5] [6] 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) [7]. 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 [7].

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. Все эти особенности необходимо учитывать при разработке САРР.

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. [8] [9].

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

Различные методы извлечения признаков включают [8] [9]:

- Линейно-предсказывающее кодирование (англ. Linear prediction coding (LPC))

  Линейное предсказание уже продолжительное время остается одним

из основных подходов к задачам цифровой обработки речи. Принцип метода линейного предсказания состоит в том, что участок речевого сигнала можно аппроксимировать линейной комбинацией предыдущих участков сигнала.

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

- Мэл-частотные кепстральные коэффициенты (англ. Mel frequency Cepstral Coefficient (MFCC))

  Мел-частотный анализ представляет частоты речи с позиции психоакустического параметра слуха – высоты тона. Высота тона определяет, насколько высоким или низким кажется тон слушателю. [10] [11]

- Кепстральные коэффициенты на основе линейного предсказани (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. [10]

- 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. Идея применения вейвлетов состоит в многомасштабной обработке сигнала, т. е. в анализе сигнала в разном увеличении с разной степенью детализации. [10] [11]

  Недостатком вейвлет преобразований можно считать их относительную сложность расчетов.

В статье [12] проводится эксперимент работы трех платформ CAPP

(Amazon, Google, IBM) для двух групп людей: с нейродегенеративными (медленно прогрессирующие, наследственные или приобретенные заболевания нервной системы, ведущими к различным симптомам – к деменции, 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 CAPP для здоровых людей выше, чем для людей с заболеваниями (рассматривались 2 группы людей, с различными заболеваниями: с рассеянным склерозом и с атаксией Фридрейха). При этом при увеличении продолжительности болезни, точность CAPP снижалась.

Часть результата приведена на рисунке 1

Обозначения:

- Accuracy – точность распознавания слов
- One Word, Two Words, Three Word – точность распознавания до одного слова, двух слов, трех слов
- Female, Male – женщины, мужчины
- Healthy Control – группа здоровых людей
- Multiple Scierosis – группа людей с рассеянным склерозом
- Friedreich's Ataxia – группа людей с заболеванием "Атаксия Фридрайха"

Как видно из эксперимента, системы распознавания речи работают с ошибками для людей с нейродегенеративными заболеваниями.

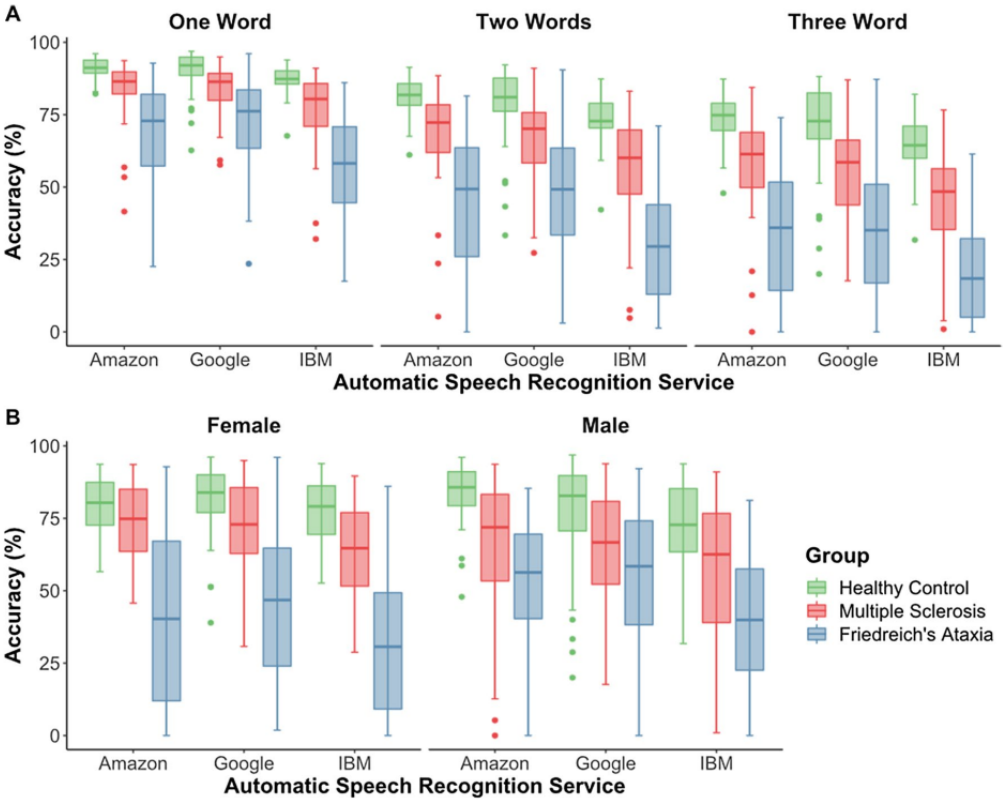

Рисунок 1 – Результат эксперимента

Литература

[1] Федосин С.А. Еремин А. Ю. Классификация систем распознавания речи.

[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.

[4] 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.

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

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

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

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

[9] 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.

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

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

[12] S. B. G. V. Feature Extraction Techniques for Speech Recognition // International Journal of Speech Technology. 2021. 24:771–779. P. 771–780.