



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. Однако большинство из них направлены на распознавание изолированных слов.

Для начала надо понять, что такое распознавание речи? Сначала кажется, что это довольно простое действие: человек (диктор) произнес фразу (слово, команду), а система либо понимает команду, либо отвечает, либо набирает диктуемый текст, либо делает с этой информацией что-то иное. Однако все довольно сложнее, и работы таких систем объединяет в себе множество различных технологий.

Одним из наиболее сложных моментов в разработке систем распознавания речи является довольно широкая междисциплинарность задачи, то есть затрагиваются вопросы теории обработки сигналов, математического анализа, психологии, теории коммуникаций, а также лингвистики.

Распознавание речи - это задача, которая усложнена тем, что речь человека характеризуется высокой степенью изменчивости. Причины этого следующие:

- для одного и того же диктора, произношения одних и тех же звуков (слов, фраз) будут отличаться по своей длительности произношения, интонацией, часто это связано с изменением физического или эмоционального состояния человека, его настроением или условий, в которых он находится;
- произношение фонем сильно зависит от контекста, например наличие артикуляции при разговоре;
- различные помехи (отражения звука, искажение микрофона или системы передачи, фоновый шум работы каких-либо устройств).

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 написать алгоритм, который научит компьютер (машину) понимать различные особенности слитной речи, 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.

В современном мире почти все алгоритмы распознавания речи основаны на семантико-синтаксических или стохастических ограничениях в моделях генерации фраз, но такие алгоритмы могут распознавать только идеально построенные и четко произнесенные слова или фразы, что затрудняет использование уже реализованных систем распознавания речи людям с дефектами. Если рассматривать уже существующие системы распознавания речи, то следует отметить, что используются два подхода, которые в свою очередь являются принципиально различными

- распознавание голосовых меток;
- распознавание лексических элементов.

Первый – не подходит для система распознавания речи у людей с дефектами речи, так как он основан на заранее написанному образцу и чаще используется в простых системах.

Второй подход более сложный, основанный на выделение из потока звуков (речи) отдельных лексических элементов (фонем). Такие фонемы дальше объединяются в слоги и морфемы. Этот подход более удобен в использовании, так как он не настраивается на “шаблон произношения”. Его можно применять и для распознавания речи с неправильным произношением звуков, можно создавать различные словари звуков, чтобы методом полного перебора можно было выбрать нудный и подставить его в даль-

нейшую цепочку действий. Таким образом, в сложных системах чаще всего используют именно его.