ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА АУДИОСООБЩЕНИЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ РЕЧЕВОЙ ИНФОРМАЦИИ

Кисель Е. В.
Кафедра интеллектуальных систем, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: kisel.jenya@gmail.com

В данной работе рассматривается алгоритм распознавания аудиосообщений. Представлен прототип распознавания речи в режиме реального времени, реализованный на JavaScript. В работе рассмотрены и используются библиотеки Web Audio API и Pocketsphinx.

Введение

Речь - это последовательность звуков. Звук - это колебания частиц воздушной среды, воздействующих на слуховую систему человека и создающих слуховые ощущения. В настоящее время под понятием «распознавание речи» содержится целая сфера научной и инженерной деятельности. Каждая задача распознавания речи заключается в выделении, классификации и обработке человеческой речи из входного звукового потока. Это может быть и выполнение определенного действия на команду человека, и выделение определенного слова-маркера из большого массива телефонных переговоров, и системы для голосового ввода текста.

Основная задача системы распознавания речи - представление аудиопотока как набора символов. Как правило, распознавание речи не является абсолютно правильным. На это могут влиять такие факторы, как неточность существующих алгоритмов, неполнота используемых словарей и, в не меньшей степени, наличие шумов. В данной работе анализируется эффективность некоторых алгоритмов шумоподавления на основе результатов работы спроектированной системы по распознаванию речи в режиме реального времени.

I. Шумоподавление

Все устройства записи, как аналоговые, так и цифровые, обладают свойствами, которые делают их восприимчивыми к шуму. Шум может быть случайным и не когерентным, то есть не связанным с самим сигналом, или когерентным, вносимым устройствами записи и алгоритмами обработки [1].

В настоящее время существует много способов подавления шума. Самый простой из них - пороговый шумоподавитель. Он блокирует прохождение сигналов в паузах фонограммы, действуя как простой выключатель - либо полностью пропускает входной сигнал на выход, либо полностью его подавляет. В современных моделях задается порог срабатывания, ниже которого сигнал не проходит. Но это не всегда дает необходимый результат, так как во время звучания уровень шума все равно остается довольно высоким и заметным на слух.

Другой способ шумоподавления был распространен несколько лет назад и назывался динамическим ограничителем шума. На основе анализа высоких частот обрабатываемого сигнала происходило их ослабление в том случае, если уровень в исходном сигнале достаточно мал, и ими можно пренебречь. Для этого применялся скользящий адаптивный фильтр, который изменял полосу своего пропускания в зависимости от спектра обрабатываемого сигнала.

С развитием цифровой обработки сигналов широкое распространение получил метод спектрального вычитания. Идея метода заключается в том, что из амплитудно-частотного спектра полезного сигнала вычитается либо указанный заранее, либо выделяемый автоматически спектр чистого шума. Число частотных полос, на которые разбивается сигнал, в зависимости от реализации алгоритма может достигать нескольких тысяч, то есть ширина полосы, в которой ведется обработка, будет составлять единицы Герц. Это позволяет эффективно отфильтровывать гармоники полезного звукового сигнал от шумовых составляющих [1].

II. Алгоритм распознавания речи

В данной работе для распознавания речи создан прототип на основе готовой библиотеки pocketsphinx.js [2].

Библиотека разработана с использованием языка javascript и позволяет осуществлять распознавание речи в реальном времени прямо в браузере конечного пользователя. Принцип работы распознавания основан на использовании скрытых марковских моделей. Основные преимущества такого подхода - высокая скорость и точность распознавания.

Принцип распознавания речи при помощи библиотеки pocketsphinx.js заключается в следующем: необходимо взять аудиопоток, который должен быть предварительно обработан, разделить этот аудиопоток на высказывания молчанием, разобрать, что говорилось в каждом выска-

зывании (берутся всевозможные заданные комбинации слов, сопоставляются с аудио, выбирается наиболее подходящая комбинация) [3].

Для улучшения качества распознавания необходимо обработать входной сигнал. В работе использовано пороговое шумоподавление, описанное ранее.

III. Анализ совпадения речи

Определение совпадения речи состоит из трех концепций: концепция функций, концепция модели, концепция совпадений. Произнесенная речь делится на фреймы - кадры длительностью 10 миллисекунд. Для каждого фрейма определяется вектор признаков, описывающих речь. Для выделения вектора признаков речевого сигнала используется спектральное представление речи, которое можно разделить на два этапа. На первом этапе осуществляется дискретное преобразование Фурье, что позволяет получить частотный спектр речевого сигнала. На втором этапе обрабатывается, улучшается и очищается полученный спектр сигнала. При анализе речевого потока также учитываются ее изменчивость и динамические особенности речи. Используемые для этого параметры представляют собой производные по времени от основных параметров речи, таких как тембр голоса, скорость речи, изменение интонации [4].

В качестве модели речи часто используются скрытые Марковские модели (Hidden Markov Model). В этой модели процесс описывается как последовательность состояний, которые изменяют друг друга с определенной вероятностью. В любой момент поддерживаются лучшие подходящие варианты и расширяются с течением времени, создавая наиболее подходящие варианты для следующего фрейма.

Скрытая Марковская модель состоит из трех частей. Первой частью является акустическая модель, которая содержит наиболее возможные вектора признаков. Вторая часть - фонетический словарь. В нем хранятся заданные слова в виде фонем. И третья часть - языковая модель, которая определяет, какое слово может следовать за ранее распознанными словами.

Распознавание слов в работе происходит путем сопоставления входного сигнала и заданного фонетического словаря. Количество распознаваемых слов напрямую зависит от словаря, т. е. для увеличения этого самого количества необходимо расширить фонетический словарь.

IV. Результаты распознавания аудиосообщений

Результаты распознавания при различных условиях представлены в таблице 1. Ошибки в распознавании аудиосообщений обусловлены наличием аукстических помех и искажений, наличием речевых помех, разным произношением одних и тех же слов, спонтанной речи, сопровождаемой аграмматизмами и речевым мусором.

Таблица 1 – Результат распознавания

Входной сигнал	Объём выборки	Процент ошибок	Неправильно	Распознано лишнее
С фоновым шумом	50	38%	2%	36%
В тишине	50	18%	0%	18%
Фоновый шум с пороговым шумоподавлением	50	21%	1%	20%

V. Заключение

Таким образом, в работе анализируются сложности, возникающие в процессе распознавания речи, связанные с наличием различного вида шумов. Находясь в помещении без шумов распознавание является правильным. Качество распознавания в значительной степени зависит от «чистоты» входного сигнала, что обуславливает необходимость использования предварительной обработки. Рассмотренное в работе пороговое шумоподавление оказалось не очень эффективным для устранения всех шумов, что требует дальнейших исследований способов шумоподавления.

Разработанный на данный момент прототип имеет лишь ограниченное количество распознаваемых слов. Для увеличения этого количества могут применяться различные способы, в том числе использование онлайн-словаря.

Проектированием прототипа движет цель разработать устройство с уже известными технологиями распознавания без недостатков, присущим уже готовым устройствам.

1. Шкритек П. Справочное руководство по звуковой схемотехнике. Способы снижения шумов и помех. Москва: Мир, –1991. –246, 446 с.
2. Pocketsphinx [Электронный ресурс]. - Режим доступа: https://cmusphinx.github.io/wiki/ - Дата доступа: 20.04.2017.
3. Speech Analysis FAQ [Электронный ресурс] - Режим доступа: http://svr-www.eng.cam.ac.uk/ – Дата доступа: 25.03.2017.
4. Фролов А. В., Фролов Г. В. Синтез и распознавание речи. Современные решения. М. : Связь, –2003. –186 с.