

ДЕТЕКТОР ЗАИКАНИЯ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Лукашик Р.В.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Лихачёв Д.С. – доцент, канд. тех. наук

Распознавание и классификация речевых дефектов, таких как заикание, является сложным процессом. Данная работа описывает процесс обработки речевого сигнала и выявления акустических параметров для распознавания и классификации заикания. Были сделаны выводы об использовании комбинации временных и частотных акустических параметров речи.

В настоящее время существует большое количество исследовательских работ, направленных на автоматическое распознавание речевых дефектов, таких как заикание. Большинство из них базируется на одном из трех классификаторов: искусственные нейронные сети, скрытые марковские модели и метод опорных векторов. Точность определения заикания, используя различные классификаторы, отличается у различных исследований из-за объема базы данных заикающейся речи, использования искусственного или естественного заикания и др. параметров. Используя скрытые марковские модели исследователи добились 96% точности определения заикания [1].

При распознавании заикания стояла задача выбора акустических параметров речи, анализ которых должен быть дать наилучшие результаты в определении заикания.

Иногда при заикании возникают спонтанные мышечные спазмы речевого аппарата, которые блокируют произношение определенных звукосочетаний. Блокировку можно легко наблюдать, анализируя речевой сигнал во временной области, получая данные для статистического анализа. Анализируя речевой сигнал по временным параметрам, можно определить длительность сигнала, т.е. продолжительность звука.

Однако, анализируя речь только лишь по временным характеристикам, нельзя добиться высокой точности распознавания заикания. В этом случае может быть полезен анализ речевого сигнала в частотной области. Расчет мел-кепстральных характеристик (MFCC) может помочь в определении гласных звуков и обнаружении речевых нарушений, связанных с повторением гласных, а также их удлинением. Удлинение гласных является одной из значимых особенностей, характеризующих заикание. Из-за относительно большой продолжительности артикуляции гласных звуков их легче анализировать с использованием кепстрального анализа.

Повторение гласного или всего слога анализировать сложнее, чем продление. Это связано с тем, что продолжительность их артикуляции значительно короче, и по этой причине объем информации, получаемой из анализа отдельных фреймов значительно меньше. Однако эта проблема решается частотой встречаемости повторений, которая значительно выше, чем в случаях продления.

Исходя из этого были сделаны выводы о необходимости использования комбинации как временных, так и частотных акустических параметров речи для возможности более точного

определения повторяющихся фрагментов сигнала, характерных для заикания. Среди параметров речи были выбраны временные признаки: длительность сегмента и высота сегмента, а среди частотных признаков: частота основного тона, интенсивность, амплитуда, энергия.

В качестве базы для разрабатываемого инструментария извлечения акустических параметров была выбрана система OpenSMILE, в которой реализована функциональность первичной обработки речевого сигнала, т.е. погашение шумов за счет применения частотных фильтров, сегментация речевого сигнала с перекрытием, применение преобразования Фурье (FFT512) и извлечения MFCC.

Для анализа речевого сигнала были выбраны следующие параметры: длина фрейма – 1024 отсчета; число отсчетов в преобразовании Фурье – 512; окно Хэмминга; порядок сглаживания в косинус-преобразовании был установлен равным 12.

Важной и наиболее трудозатратой частью разработки детектора заикания является анализ речевых баз аудиозаписей людей с заиканием с целью выявления акустических параметров и обучения нейронной сети. Необходимо наработать базу аудиозаписей людей с заиканием различного возраста, пола, скорости речи и типа заикания: это может быть повторение отдельного слога, целого слова, запинки и другого рода заикания. База аудиозаписей должна быть достаточно обширной для возможности обучения нейронной сети, проверки корректности ее обучения и повышения точности распознавания речевых дефектов.

Список использованных источников:

1. Overview of Automatic Stuttering Recognition System, Lim Sin Chee, Ooi Chiam Sazali Yaacob, Universiti Malaisia Perlis, 02600 Jejawi Perlis, 2009
2. Jayanthi Sasisekaran, Effects of Delayed Auditory Feedback on Speech Kinematics in Fluent Speakers, University of Minnesota, 2012, PMID:23409597.
3. Marilia, Effect of delayed auditory feedback on stuttering with and without central auditory processing disorders, Brasil, 2017, PMID:29236907.