

АНАЛИЗ АУДИОСИГНАЛОВ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ НА БАЗЕ МОБИЛЬНЫХ ПЛАТФОРМ

Тарасевич В.В.

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

Лихачёв Д.С. – канд. техн. наук

В данной работе описывается детектор правильности выполнения речевых заданий для систем коррекции речи, выполненный с использованием методов машинного обучения, его место в улучшении качества детектора речевых нарушений у людей с боковым амиотрофическим склерозом, а также краткий обзор используемых инструментов.

Системы обработки аудиосигналов всё чаще реализуются на основе мобильных платформ. Это системы для распознавания и поиска музыки, такие как “Яндекс.Алиса”, “Shazam”, “SoundHound”, приложения для понимания и обработки усной речи, такие как “Яндекс.Алиса”, “Google Assistant”, “Cortana” от компании Майкрософт и многие другие.

Среди таких систем есть и системы диагностики, в частности, системы диагностики Бокового Амиотрофического Склероза (БАС) [1]. БАС – неизлечимое нейродегенеративное заболевание. Основным методом диагностики данного заболевания занимает около года [1-2].

В исследовании [1] приводится описание системы автоматического детектирования речевых нарушений у пациентов с БАС при помощи методов машинного обучения. Суть метода заключается в акустическом анализе протяжного произношения звука [а], выделении ключевых выбранных признаков, их дальнейшей обработки классическими методами машинного обучения (методом к-

ближайших соседей и методом линейного дискриминантного анализа).

Методика детектирования, описанная в исследовании [1] опирается на тот факт, что произношение звука [a] является правильным, да и на то, что пациент произносит именно звук [a], а не какой-нибудь другой.

Целью данной работы является усовершенствование приложения созданного и описанного в [1-2] для автоматического детектирования речевых нарушений для диагностирования БАС при помощи оценки качества правильности произношения речевого задания (произношения звука [a]).

Для целей детектирования правильности произношения речевого задания хорошо подходят методы машинного обучения. Для распознавания речи много эффективных методов на основе машинного обучения, а распознавание протяжной буквы [a] – это по сути и есть распознавание фонемы, звука.

Используя методы машинного обучения необходимо определиться с признаками, которые будут выделяться для их последующей обработки этими методами. Признаки должны учитывать временную составляющую речи, а также частотную и амплитудную.

Для временной составляющей используется длительность фрагментов слитной речи. Также в работе используются пертурбационные признаки (амплитудные и частотные): джиттер, шиммер.

Для извлечения перечисленных признаков из аудиосигнала в данном исследовании используется проект OpenSMILE.

Проект OpenSMILE предназначен для автоматического извлечения признаков из аудиосигнала. SMILE в названии расшифровывается как “Speech and Music Interpretation by Large-space Extraction” что переводится, как “Интерпретация речи и музыки посредством многоуровневого выделения”. Детальное описание подходов используемых в OpenSMILE представлено в [3].

OpenSMILE используется для исследований и в приложениях для автоматической обработки речи и музыки в режиме реального времени.

Для обработки извлечённых признаков взят метод опорных векторов [4], поскольку он хорошо зарекомендовал себя на небольших обучающих выборках, что очень весомо для данной работы, так как имеющаяся выборка данных небольшая.

Предварительные эксперименты показали перспективность описываемого подхода.

Список использованных источников:

1. Вашкевич М.И., Рушкевич Ю.Н. Система детектирования речевых нарушений у пациентов с боковым амиотрофическим склерозом на основе теста на протяжное произнесение звука [a] // Актуальные проблемы неврологии и нейрохирургии. Рецензируемый сборник научных трудов. Выпуск 22. Под ред. д.м.н. Р.Р. Сидоровича и д.м.н., проф. С.А. Лихачева. – С. 278–290.
2. Detection of bulbar ALS using a comprehensive speech assessment battery/ Y. Yunusova [et al.] // Proceedings of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications – 2013. – P. 217-220
3. B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth, G. Rigoll. Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies.
4. Nello Cristianini, John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. — Cambridge University Press, 2000.