

ОТЗЫВ

**на диссертационную работу Пикалёва Ярослава Сергеевича
на тему «Совершенствование методов и программных средств
распознавания слитной русской речи», представленную на соискание ученой
степени кандидата технических наук
по специальности 05.13.01 – Системный анализ, управление и
обработка информации (по отраслям) (технические науки)**

В период работы над диссертацией Пикалёв Ярослав Сергеевич проявил себя ответственным, исполнительным и творческим молодым учёным, имеющим ярко выраженные аналитические способности. Его компетентность в области нейросетевых технологий и знание новейших методов обработки речевых и текстовых данных позволила соискателю проводить анализ современных научных работ отечественных и зарубежных ведущих исследователей в области распознавания речи, самостоятельно ставить серьезные научно-исследовательские задачи, определять эффективные пути их решения, получать достоверные теоретические и практические результаты, качественно анализировать и интерпретировать их, формулировать выводы. Его работу отличает глубокое теоретическое обоснование рассматриваемых вопросов с практическим применением основных положений.

В период обучения в аспирантуре Пикалёв Я.С. продолжал работать по совместительству младшим сотрудником в ГУ ИПИИ, успешно совмещая учёбу и научную деятельность в аспирантуре с работой, являясь исполнителем основных этапов фундаментальной НИР «Исследование и разработка методов семантического анализа и интерпретации потоков данных интеллектуальными системами».

Слитная речь характеризуется высокой вариативностью фраз, зависимостью их произнесения от темпа речи и эмоционального состояния диктора. Дополнительно усложняет задачу распознавания речи наличие фонового шума, наложение нескольких голосов в одном аудиоканале. При распознавании синтетических языков, к которым относится русский, возникает ряд трудностей, связанных с недостаточным количеством аннотированных речевых баз и нормализованных текстов для обучения языковых моделей, создание этих ресурсов усложняется флективностью русского языка в силу и свободным порядком слов в предложении. Кроме того, фонетические особенности русского языка (редукция гласных, ослабление и выпадение согласных) **увеличивают акустическую** вариативность для русскоязычной слитной речи.

В связи с этими проблемами в сфере распознавания русскоязычной слитной речи на сегодняшний день успехи достигнуты только в пределах ограниченного

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Вх. № 16/181
25.05.2021

словарного запаса. Тем самым актуальной остаётся проблема разработки методов, алгоритмов и программных средств, обеспечивающих повышение точности распознавания слитной русскоязычной речи. Подобная проблема рассматривается в диссертационной работе Пикалёва Ярослава Сергеевича, что свидетельствует об **актуальности** работы.

Для изучения технологий языкового и акустического моделирования, а также распознавания речевых сигналов автор проделал большую работу по анализу литературных источников. В ходе диссертационного исследования соискателем получены и проанализированы научные результаты, отличающиеся новизной и практической значимостью, сделаны соответствующие выводы. Научная новизна полученных результатов заключается в совершенствовании нейросетевых методов анализа речевых и текстовых данных, используемых при разработке систем распознавания речи, а именно:

1) получили дальнейшее развитие нейросетевые методы автоматического построения транскрипции за счет:

a. модернизации архитектуры нейросети типа Transformer, стандартно используемой для определения позиции ударения, которая заключается в увеличении количества слоёв, использовании методов градиентного отсечения и teacher forcing для оптимизации параметра скорости обучения, что позволило повысить точность определения позиции ударения по сравнению со стандартной моделью Transformer;

b. модернизации sequence to sequence модели для генерации практических транскрипций англоязычных слов и слов-исключений, заключающейся в использовании механизма обучения с подкреплением RL-block и метода beam-search для выбора наиболее вероятной последовательности символов, что позволило повысить точность модели по критериям количества ошибочно сгенерированных символов и неправильно сгенерированных слов;

2) впервые предложена модель нейросетевой параметризации для извлечения информативных акустических признаков, основанная на иерархической мультимодульной архитектуре, представляющей собой объединения ансамбля нейронных сетей с «узким горлом» и архитектуры ResNet-50. Использование данной модели позволяет повысить точность распознавания по сравнению с моделью, извлекающей стандартные bottleneck-признаки;

3) получили дальнейшее развитие методы нейросетевой классификации фонем за счет использования механизма внимания в последнем скрытом слое сети, включающей в себя сеть с временными задержками и двухнаправленную нейросеть с долгой кратковременной памятью, что позволило сохранять высокую точность на относительно

небольшом обучающем наборе аудиоданных, в отличие от end-to-end систем, для обучения которых требуется речевая база длительностью в десятки тысяч часов.

Обоснованность и достоверность научных результатов исследования подтверждается значительным количеством численных исследований, для проведения которых автором проделана огромная работа по сбору и обработке речевых и текстовых данных, находящиеся в открытом доступе, с целью создания собственного аннотированного речевого корпуса.

Практическую значимость результатов диссертационной работы составляют вычислительные алгоритмы автоматической разметки аудиоданных, получения транскрипций и робастных дикторонезависимых акустических признаков, что может использоваться для создания аннотированных речевых баз данных, в системах синтеза и распознавания речи.

Диссертация Пикалёва Ярослава Сергеевича на тему «Совершенствование методов и программных средств распознавания слитной русской речи» является законченным самостоятельным логично структурированным и достоверным научным исследованием, посвященным развитию теоретических аспектов нейросетевых методов обработки и анализа данных и практических решений автоматизации процессов анализа и классификации массивов речевых и текстовых данных.

Диссертационная работа соответствует требованиям Положения о присуждении ученых степеней ВАК при МОН ДНР, а ее автор, Пикалёв Ярослав Сергеевич, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

Научный руководитель:

канд. техн. наук, доцент, доцент
кафедры компьютерных технологий
ГОУ ВПО «ДОНЕЦКИЙ
НАЦИОНАЛЬНЫЙ
УНИВЕРСИТЕТ»



Т.В. Ермоленко

Личную подпись Ермоленко Т.В.

ЗАВЕРЯЮ

Ученый секретарь
ГОУ ВПО «ДОНЕЦКИЙ
НАЦИОНАЛЬНЫЙ
УНИВЕРСИТЕТ»
канд. филол. наук



М.Н. Михальченко