

УТВЕРЖДАЮ:

И.о. директора Государственного  
учреждения «Институт прикладной  
математики и механики»,  
доктор физ.-мат. наук,  
старший научный сотрудник.



С.Н.Судаков

2021 г.

## ОТЗЫВ

ведущей организации о диссертации **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)

### 1. Актуальность для науки и практики

Актуальность темы диссертации обусловлена необходимостью совершенствования систем автоматического распознавания речи с целью повышения точности распознавания речевого сигнала в условиях неподготовленного спонтанного разговора (диалога или полилога) и телефонного канала. Современное общество невозможно представить без систем взаимодействия человека и машины, и в этих системах автоматическое распознавание речи занимает одну из ведущих ролей.

Существующие на сегодняшний день системы распознавания слитной речи, рассчитанные на сотрудничающего диктора, на русском языке обеспечивают высокую точность распознавания (90% и более). Однако эффективность работы таких систем в условиях неподготовленной разговорной речи не достаточно высока. В связи с этим, задача разработки и совершенствования методов, алгоритмов и программных средств распознавания спонтанной русской речи, решению которой посвящена диссертация Пикалёва Я.С., является *актуальной*.

### 2. Новизна исследования и полученных результатов, выводов и рекомендаций, сформулированных в диссертации

В диссертации Пикалёва Я.С. получены следующие новые научные результаты.

1. Разработан метод автоматического определения позиции ударения в слове, использующий архитектуру Transformer в качестве базовой. Модель модернизирована за счет увеличения количества слоёв, использования методов градиентного отсечения и teacher forcing для оптимизации параметра скорости обучения. Доказано в результате экспериментов, проведенных на фонетических словарях общим объемом более 5 млн. записей, предложенная модернизация позволила снизить показатель пословной ошибки до 0,08, повысив точность модели по сравнению со стандартной на 10%.

2. Разработан метод генерации практических транскрипций англоязычных слов и слов-исключений на основе seq2seq модели, отличием которого является применение механизма обучения с подкреплением и метода beam-search для выбора наиболее вероятной последовательности символов. Доказано в результате эксперименте, что предложенная техника модернизации моделей типа sequence to sequence на основе внесения изменений в структуру алгоритма построения позволила повысить точность модели по критерию количества ошибочно сгенерированных символов на 0,8% и 3%, по критерию неправильно сгенерированных слов на 0,6% и 9% соответственно.

3. Разработан метод построения информативных признаков речевого сигнала, извлекаемых из малоразмерного линейного слоя глубокой нейронной сети. Метод отличается от существующих применением объединения ансамбля нейронных сетей с «узким горлом» и архитектуры ResNet-50. Доказано в результате экспериментов, проведенных на реальных записях русской слитной речи, что происходит снижение показателя пословной ошибки распознавания при обучении акустических моделей с использованием предложенных высокоуровневых информативных признаков на 2,7% по сравнению с моделью, извлекающей стандартные bottleneck-признаки.

4. Разработан метод нейросетевой классификации фонем, использующий нейросеть с временными задержками и двунаправленную нейросеть с долгой кратковременной памятью, а также механизм внимания в последнем скрытом слое сети. Доказано в результате экспериментов, проведенных на реальных записях русской слитной речи, что предложенная архитектура сети для классификации фонем позволяет сохранять высокую точность и быстродействие в задаче распознавания слитной русской речи, используя для своего обучения значительно меньше данных, чем известные системы от компании ЦРТ и Google.

Новизна полученных результатов не вызывает сомнения.

### **3. Значимость для науки и производства (практики) полученных автором диссертации результатов**

Научная значимость результатов диссертации определяется тем, что в ней проведена модернизация существующих алгоритмов обучения акустических моделей на основе глубоких нейронных сетей и построения акустических признаков речевого сигнала. В частности, предложенная технология извлечения робастных признаков из скрытых слоев иерархической мультимодульной нейросети вносит свой вклад в теорию понимания кодирования сигнала на основе глубоких нейросетей, которые на сегодняшний день рассматриваются как «черный ящик».

Практическая значимость диссертации состоит в создании системы распознавания русской слитной речи, которая демонстрирует высокую точность распознавания и быстродействие, но требует для своего обучения гораздо меньше ресурсов по сравнению с существующими в России и за рубежом альтернативными системами.

О большой практической значимости результатов работы свидетельствует их успешное использование при проведении нескольких научно-исследовательских работ, выполненных в Институте проблем

искусственного интеллекта. Результаты внедрения подтверждены соответствующей справкой, имеющейся в диссертации.

#### **4. Рекомендации по использованию результатов и выводов, приведенных в диссертации**

Результаты работы обладают большой научно-практической значимостью, что позволяет рекомендовать их к внедрению в организациях, активно занимающихся исследованиями и разработкой систем автоматического распознавания речи. Результаты и выводы, приведенные в диссертации, будут полезными как для усовершенствования систем автоматического распознавания разговорной русской речи, так и для создания таких систем для других языков.

#### **5. Обоснованность и достоверность научных положений, выводов и заключений**

Все представленные в диссертации научные положения, выводы и рекомендации являются полностью обоснованными и достоверными. Предложенные в диссертации идеи базируются на обобщении передового опыта ведущих мировых исследователей, работающих в области автоматического распознавания слитной речи. Теоретические выводы основываются на известных методах обработки речевого сигнала, согласуются с экспериментальными результатами по теме диссертации, полученными в рассматриваемой области, связанной с дикторонезависимым распознаванием слитной русской речи.

Для экспериментальных работ результаты получены на естественном речевом материале, при этом тестовые выборки ни по произнесениям, ни по составу дикторов не пересекались с обучающими данными.

При этом использованы данные сравнения представленных в диссертации методов и алгоритмов с существующими методами и алгоритмами, применяемыми для решения аналогичных задач.

Установлено качественное соответствие между полученными автором результатами и результатами, представленными в независимых источниках, при решении задачи распознавания слитной русской речи.

Обоснованность и достоверность результатов диссертации подтверждается тем, что в ней использованы современные методики проведения исследования, широко распространенные в прикладных науках, а также речевые корпуса, которые были записаны в естественных условиях и являются представительными с точки зрения таких параметров, как характеристики канала, наличие фоновых шумов различной природы, длительность речевого сигнала, количество дикторов, их пол, возраст, социальный статус и место проживания.

#### **6. Оценка содержания диссертации, ее завершенность в целом, замечания по оформлению**

Диссертация оформлена грамотно, с логическими переходами между главами. В *первой* главе сделан анализ современных подходов к распознаванию речи. Приведен обзор предшествующих работ по проблеме дикторонезависимости и робастности акустических моделей. Во *второй* главе описан подход к созданию собственного аннотированного речевого корпуса для обучения акустической и языковой моделей. *Третья* глава посвящена

описанию системы автоматического получения транскрипции, приведены особенности фонетики русского языка, архитектуры нейросетей для определения позиции ударения, получения транскрипции слов-исключений и практической транскрипции для вставок на латинице. В **четвертой** главе дано описание технологии повышения робастности акустических моделей, основанной на глубоком и машинном обучении. Предлагаемая технология использует информативные акустические признаки, извлечённые из иерархических нейросетевых моделей, а также на гибридные акустические модели, обученные на основе машинного и глубокого обучения с применением дискриминативного подхода. Полученные модели обеспечивают инвариантность к голосам дикторов и различным акустическим обстановкам. Предлагается модель нейросетевой классификации фонем, а также приведены численные исследования, подтверждающие качество предложенных моделей. В **пятой** главе представлена структура разработанной системы автоматического распознавания слитной русской речи, приведены результаты экспериментальных исследований, позволяющие сравнить качество распознавания авторской системы с существующими решениями.

**Заключения**, содержащиеся в диссертации, подтверждают все научные положения и выводы, сделанные автором, и завершенность диссертационной работы в целом.

#### **7. Общие замечания**

1) в работе использовалась аугментация речевых данных, позволяющая повысить робастность акустических моделей. Наложение аудиозаписей шумов, музыки и голоса другого диктора является одним из этапов аугментации. Стоило провести численные исследования и сравнить качество распознавания авторской системы с существующими решениями в условиях шумов подобного рода, что могло подтвердить преимущество авторской системы, выражающееся в устойчивости работы с зашумленным сигналом;

2) в пятой главе при описании реализации модуля денормализации текста указывался размер ВРЕ-словаря, равный 32 тыс. токенов. Известно, что размер ВРЕ-словаря может оказать влияние на качество языковой модели, а следовательно, и на предсказание последовательности символов. В связи с этим стоило аргументировать выбор объема ВРЕ-словаря;

3) при оформлении списка литературы не везде соблюдены требования ГОСТа.

В целом указанные замечания не снижают ценности и практической значимости диссертационной работы.

#### **8. Соответствие автореферата основным положениям диссертации**

Автореферат представляет собой хорошо продуманное краткое изложение материала диссертации. В нем компактно изложены как основные положения диссертации, так и их доказательная база, которые позволяют оценить научную новизну и практическую значимость проведенной работы.

#### **9. Подтверждения опубликованных основных результатов диссертации в научной печати**

Основные результаты диссертации представлены в 17 опубликованных научных работах, в том числе 5 публикаций в изданиях, входящих в Перечень рецензируемых научных изданий ВАК РФ и ДНР, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук, и 2 публикации в изданиях, включенных в систему Российского индекса научного цитирования.

В автореферате указан личный вклад автора. Полученные автором результаты прошли апробацию на одной всероссийской и 9 международных научно-технических конференциях.

**10. Заключение о соответствии диссертации критериям, установленным Положением «О присуждении ученых степеней»**

Таким образом, диссертация Пикалёва Ярослава Сергеевича является научно-квалификационной работой, в которой содержится решение задачи повышения эффективности дикторонезависимой системы автоматического распознавания слитной русской речи, учитывающей её особенности и адаптирующейся под любую предметную область, имеющей значение для развития методов и средств человеко-машинного взаимодействия, что соответствует требованиям п. 2.2 Положение «О присуждении ученых степеней», утвержденного постановлением Совета Министров ДНР от 12.02.2015 г. № 2-13, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

Отзыв обсужден и одобрен на заседании научного семинара отдела теории управляющих систем «Современные проблемы в машинном обучении и анализе данных» Государственного учреждения «Институт прикладной математики и механики» «16» сентября 2021 г., протокол № 5.

Учёный секретарь ГУ «ИПММ»,  
кандидат технических наук

А.Ю.Максимова



*Максимова А.Ю.*  
*Ильченко М.Н.*