

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

Использование речевых технологий при создании средств устного диалога человека с компьютером на сегодняшний день становится всё более распространённым. С учётом того, что в силу определенных трудностей, связанных с особенностями артикуляции, флексивностью языка, свободным порядком слов в предложении, отсутствием объёмного аннотированного речевого и нормализованного текстового корпуса, в сфере распознавания русскоязычной слитной речи успехи достигнуты только в пределах ограниченного словарного запаса. Поэтому не вызывает сомнений актуальность выбранной Пикалёвым Ярославом Сергеевичем темы диссертационной работы: «Совершенствование методов и программных средств распознавания слитной русской речи», направленной на повышение эффективности методов и программных средств дикторонезависимого распознавания слитной русской речи, позволяющих обеспечивать высокое качество распознавания, учитывать особенности орфоэпических норм русского языка.

Основными результатами диссертационной работы Пикалёва Ярослава Сергеевича, обладающими научной новизной, являются следующие:

1) предложена модификация модели глубокого обучения sequence-to-sequence на базе архитектуры Transformer для генерации практических транскрипций англоязычных слов и слов-исключений, обеспечивающая за счет применения механизма обучения с подкреплением повышение точности генерации транскрипции по критерию количества ошибочно сгенерированных символов и по критерию неправильно сгенерированных слов по сравнению со стандартной архитектурой;

2) выполнена модернизация архитектуры нейросети типа Transformer, стандартно используемой для задачи автоматического определения позиции ударения в слове, которая заключается в увеличении количества слоёв,

использовании методов градиентного отсечения и teacher forcing для оптимизации параметра скорости обучения, а также метода beam-search для выбора наиболее вероятной последовательности символов, которая позволила повысить точность определения позиции удараения на 10% по сравнению со стандартным подходом;

3) предложен подход к повышению робастности акустических моделей, использующий аугментацию и модификацию признаков для получения адаптивных и дискриминативных характеристик акустических признаков, а также нейросетевую параметризацию с помощью мультимодульной архитектуры нейросети в виде объединения ансамбля нейронных сетей с узким горлом и архитектуры ResNet-50, который позволяет извлечь из скрытых слоев информативные акустические признаки и обеспечивает их инвариантность к смене диктора и акустической обстановке и повышение точности распознавания;

4) для классификации фонем предложено объединение двух архитектур нейросетей: нейросети с временными задержками и двухнаправленной нейросети с долгой кратковременной памятью, – что позволило сохранять высокую точность на относительно небольшом обучающем наборе аудиоданных, свойственную end-to-end системам, для обучения которых требуется речевая база длительностью в десятки тысяч часов.

Достоинством работы представляются разработанные автором технологии повышения робастности акустических моделей, классификации фонем и нормализации текста, которые могут быть использованы при создании систем с речевым интерфейсом, обладающих высокой точностью распознавания и дикторонезависимостью.

В качестве замечания следует отметить, что в автореферате автор не раскрыл технологии обучения языковой модели, которая составляет важную часть системы распознавания слитной речи, ограничившись лишь упоминанием о триграммах и базовой архитектуре LSTM, хотя современные подходы к построению языковых моделей используют ВРЕ-словарь и сложные модификации архитектуры Transformer.

Указанное замечание не снижает уровень диссертационного исследования Пикалёва Ярослава Сергеевича.

Анализ автореферата показывает, что в целом диссертационное исследование Пикалёва Ярослава Сергеевича является законченной научно-

квалификационной работой, вносящей существенный вклад в развитие методов и средств человеко-машинного взаимодействия, соответствует научной специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки), соответствует всем критериям действующего «Положения о присуждении ученых степеней» (п. 2.2), предъявляемым к кандидатским диссертациям, а ее автор – Пикалёв Ярослав Сергеевич заслуживает ему присуждения ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (по отраслям) (технические науки).

Кандидат технических наук по специальности

05.13.11 – Математическое и

программное обеспечение

вычислительных машин,

комплексов и компьютерных сетей

W. B.

Ирина Сергеевна Кипяткова

Адрес: 199178, г. Санкт-Петербург, 14-я линия, д. 39

тел.: 8 (812) 328-04-21

E-mail: kipyatkova@iias.spb.su

Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН), старший научный сотрудник

Согласна на автоматизированную обработку моих персональных данных

Wm. W.

Ирина Сергеевна Кипяткова

Подпись Кипятковой И.С. подтверждена:

