В.В. Евдокимова, **П.А. Скрелин, У.Е. Кочеткова**

г. Санкт-Петербург, Россия, Санкт-Петербургский государственный университет

ФОНЕТИЧЕСКИЕ ПРИЗНАКИ ЭМОЦИОНАЛЬНОГО, ПСИХОЛОГИЧЕСКОГО И ФИЗИЧЕСКОГО СОСТОЯНИЯ ЧЕЛОВЕКА В РЕЧЕВОМ СИГНАЛЕ

Последние десятилетия стали временем изменения отношения к человеческой речи только как к психофизиологическому процессу. Взрывное развитие информатики и вычислительной техники позволили концентрировать в стабильной форме громадные речевые массивы и анализировать их с помощью математических методов, разработанных для других научных и технических задач для смежных отраслей знания. Открылись возможности статистической обработки речи и ее моделирования. Эти возможности стали основой постановки широкого круга конкретных технических задач. Методы исследования речи нашли применение в самых различных отраслях жизни от контрразведки до медицины и туризма.

По мере усложнения этих решаемых задач возрастали требования к аппарату обработки речи. Оказалось, что чисто математический подход имеет свои ограничения и может оказаться недостаточен для учета тонких речевых эффектов. В полной мере это относится к многогранной задаче распознавания психофизиологических особенностей человека по голосу и речи. Такие аспекты ее решения, как дистанционное определение эмоционального состояния собеседника, его работоспособности, верификация речевых сообщений, качественный морфинг и еще ряд актуальных задач могут дать приемлемый результат только при сочетании математических методов обработки речевого процесса и его лингвистической интерпретации. Именно использование аппарата фонетики позволяет получить результаты качественно более высокого уровня.

Роль фонетиста, который ставит своей задачей исследование происходящих в языке фонетических процессов, остается неизменной и недооцененной, на фоне имитации речеобразования и речевосприятия прикладными системами.

Основная образовательная программа магистратуры «Искусственный интеллект в моделировании речевой деятельности» ставит своей целью подготовку специалистов, которые умеют работать с системами машинного обучения и искусственного интеллекта. При этом обучение включает освоение теоретических и прикладных дисциплин, связанных с изучением, описанием, обработкой и моделированием разных аспектов устной речи человека: от порождения сообщения до интерпретации его содержания.

В 2024 году программа была отмечена премией правительства Санкт-Петербурга в сфере образования. Заведующий кафедрой Павел Анатольевич

Скрелин и доценты Вера Вячеславовна Евдокимова и Ульяна Евгеньевна Кочеткова получили награду за разработку и реализацию магистерской программы «Искусственный интеллект в моделировании речевой деятельности».

Первым классом задач по использованию нейросетевых систем для решения фонетических задач можно выделить использование систем распознавания речи. В магистерской диссертации Поволоцкой А. была поставлена задача сбора корпуса эмоциональной речи, обработка и создание датасета, обучение нейронной сети распознавать негативные эмоции в речевом сигнале. Были Записаны 72 диктора мужского пола. В общий корпус записанных фраз в результате вошли 1442 аудио-фрагмента. Общий временной объем файлов составил: 1 час 17 минут. На основе результатов проведенного перцептивного эксперимента был сформирован второй набор данных, уже проверенный аудиторами и проведена задача машинного обучения. Проведенная работа показала важность использования этапа перцептивного эксперимента при распознавании эмоциональных и психофизиологических признаков в речевом сигнале.

Магистерская диссертация М.Д. Долгушина может рассматриваться как пример автоматической обработки речевого корпуса. Работа посвящена проблеме распознавания речи людей, переживших сильные эмоциональные потрясения. Исследование производится на материале корпуса интервью со свидетелями Холокоста. В исследовании были рассмотрены различные теоретические аспекты задачи, включая современные методы автоматического распознавания речи, психофизиологические особенности эмоциональной речи, речь пожилых людей, а также фонетическая интерференция русского языка с украинским, идиш и белорусским. В результате исследования создан корпус русской устной речи людей, переживших Холокост, названный RuOH. Был разработан алгоритм для распознавания речи с использованием нейросетевой модели Wav2Vec 2.0 и N- граммной языковой модели и проведен лингвистический анализ расхождений автоматических расшифровок с реализацией. Также создан прототип интерфейса для автоматической генерации субтитров. Результирующее качество распознавания достигло 35,88% WER. Исследование вносит вклад в области корпусной лингвистики, так как существующие речевые архивы требуют всегда расшифровки материала, особенно при работе с речью пожилых людей или людей в разных состояниях.

Основной целью исследования Р.Д. Герман является автоматическая интерпретация особенностей иронического высказывания и разработка соответствующих моделей для его распознавания. В данной работе исследуются два подхода: обучение полносвязной нейронной сети на основе акустического датасета и дообучение модели Wav2Vec 2.0. Каждая из этих моделей была оценена с точки зрения их производительности и эффективности в задаче распознавания иронии. Разработанный алгоритм

позволил собрать акустические данные из корпуса иронической речи и обучить полносвязную нейросетевую модель. Исследование имеет важное значение для современных задач по обработке информации и автоматизированному анализу аудиоданных, а также способствует исследованию механизмов иронической коммуникации в русском языке. В данной работе проверка возможностей автоматической системы распознавания иронии в речи проводилась на основе тщательно разработанного корпуса иронической речи, созданного на Кафедре фонетики и методики преподавания иностранных языков. Проведенный ранее подробный акустический анализ показал, какие фонетические характеристики играют роль в восприятии и порождении иронии. Работа с нейросетевой моделью дала возможность оценить релевантность большого общего списка фонетических признаков и оценить возможности Wav2Vec 2.0 для решения подобных задач.

Нейросетевые технологии могут быть использованы и в довольно сложных задачах работы с эмоциональной окраской речевого сигнала. Опыт фонетических работ в области распознавания эмоций не показывает высокую эффективность на протяжении десятилетий. Однако необходимость работы с данным аспектом остается важной в автоматических системах общения человека с компьютерными системами. В работе Гуркова И.Е. исследуется автоматическое обнаружение и нейтрализация эмоциональной речи с помощью современных методов обработки естественного языка, таких как сентимент-анализ и перенос стиля. Набор данных состоит из расшифрованных записей экстренных вызовов, размеченных по эмоциональным категориям. Модели машинного обучения, обученные на лингвистических и акустических признаках, достигли около 62% точности в классификации эмоций. Для нейтрализации эмоциональных высказываний с сохранением семантики было проведено дообучение seq2seq модели ruT5 на данных, полученных путём промптинга большой языковой модели (GPT-4) к нейтральному перефразированию эмоциональных текстов. При дообучении модели были достигнуты значения метрики ParaScore 0.83 и высокие оценки в ходе эксперимента по человеческому оцениванию результатов, продемонстрировав способность нейтрализовать эмоции, сохраняя при этом смысл. Полученные результаты демонстрируют потенциал использования методов сентиментного анализа и переноса стиля для улучшения качества цифровой коммуникации.

При определении качества речевого сигнала и характеристик голоса не менее важно представлять, принадлежит и ли голос именно этому человеку. Применение технологий воспроизведения голоса человека, идентично существующему, стало возможным благодаря стремительному развитию нейросетей и искусственного интеллекта, которые научились анализировать и воспроизводить тембр, интонацию, паузы. Этот процесс с одной стороны помогает индустрии озвучки, где клонированные голоса способны заменить или быть помощником в работе дикторов и актёров озвучки. Польза

технологии видна в медиа и развлекательной индустрии, в коммерческих и социальных проектах, таких как виртуальные помощники, голосовые ассистенты и телемедицинские платформы. Голосовые интерфейсы, работающие на основе клонированного голоса, могут обеспечить персонализированное общение с клиентами и пациентами, усиливая доверие и улучшая пользовательский опыт.

В задачи данной работы входило провести синтеза голоса профессора Л.Р. Зиндера. Для этого был осуществлен ряд подготовительных процедур:

- сбор аудиоматериала;
- отбор наиболее репрезентативных отрезков речи, отражающих не только тембральную окраску, но и индивидуальные особенности интонирования, паузации, артикуляции;
- сегментация на аудиофайлы оптимальной длительности (15-20 секунд);
- обработка аудиоматериала: чистка фоновых шумов, удаление экстралингвистических единиц (например, вдохов) и сверхдлинных пауз.

Далее на основе различных комбинаций аудиофайлов было создано несколько моделей голоса. Был проведен ряд перцептивных тестов, чтобы определить модель, результаты генерации которой наиболее близко соответствовали характеристикам голоса, представленного в исходном материале. Экспериментальным образом были подобраны значения настраиваемых параметров: интонационной вариативности (variability), приближенности к исходному материалу (similarity), усиления индивидуальных речевых особенностей (style exaggeration). После завершения работы над моделью голоса дополнительные меры по управлению генерируемым материалом осуществлялись методами промпт-инжиниринга. Внесение дополнительных инструкций в промпт (синтезируемый текст) позволяет в определенной мере управлять эмоциональной окраской синтезируемой речи, произношением конкретных слов, паузацией, интонационным оформлением, расстановкой синтагматических границ и тональных акцентов.

Среди фонетических работ, которые проводятся на Кафедре фонетики в рамках направления распознавания психофизиологического состояния человека важно отметить направление анализа современных методов диагностики афазий при болезни Альцгеймера (БА) с использованием автоматических систем на основе технологий искусственного интеллекта (AI) и обработки естественного языка (NLP), а также разработка предложений по созданию системы диагностики БА по речи, работу над определением логопедических и речевых нарушений совместно с логопедами ЯГПУ им. Ушинского. Важным направлением работ также является работа с Мариинским театром Санкт-Петербурга по направлениям определения фонетических характеристик певческой речи в норме и патологии.

Представленные работы показывают возможности использования систем автоматической обработки речевого с учетом фонетических знаний об исследуемых явлениях. Синтез использования фонетических признаков и

возможности существующих методов искусственного интеллекта показывают высокие результаты и потенциал таких исследований.