

Элизбарян Артём Гайкович

студент

Климаш Анна Сергеевна

студентка

Каппушева Инесса Шамильевна

старший преподаватель

Научный руководитель

Новикова Алена Романовна

ассистент кафедры

ФГБОУ ВО «МИРЭА – Российский технологический университет»

г. Москва

**АВТОМАТИЗАЦИЯ ОЦЕНКИ ПРОИЗНОШЕНИЯ:
РАЗРАБОТКА ПРОТОТИПА СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧИ
ДЛЯ ФОНЕТИЧЕСКОЙ КОРРЕКЦИИ УЧАЩИХСЯ**

***Аннотация:** статья посвящена разработке прототипа автоматизированной системы оценки произношения на базе технологий распознавания речи. Рассматриваются алгоритмические основы фонетического анализа, архитектура прототипа, методы интеграции в учебный процесс, а также педагогическая результативность подхода. Предлагаются практические рекомендации по применению системы для фонетической коррекции учащихся различных уровней языковой подготовки.*

***Ключевые слова:** распознавание речи, фонетическая коррекция, автоматизированная оценка произношения, обучение иностранным языкам, нейронные сети, языковые технологии.*

1. Введение.

Произношение является одним из наиболее сложных аспектов овладения иностранным языком. При этом именно оно нередко остаётся в тени при тради-

ционном обучении: ограниченное аудиторное время, значительный размер учебных групп и субъективность восприятия педагога объективно препятствуют систематической фонетической работе с каждым учащимся. Последствия этого пробела очевидны – коммуникативные затруднения, недопонимание со стороны носителей языка и устойчивые фонетические ошибки, исправить которые со временем становится всё труднее.

Между тем современные технологии обработки речевого сигнала и машинного обучения открывают принципиально новые возможности для решения данной проблемы. Системы автоматического распознавания речи (ASR – Automatic Speech Recognition) достигли уровня точности, сопоставимого с восприятием живого собеседника, а нейросетевые модели позволяют извлекать из речевого потока детальную фонетическую информацию в режиме реального времени.

Настоящая статья описывает разработку прототипа системы автоматизированной оценки произношения, ориентированной на использование в учебном процессе. Цель работы – создать инструмент, способный не просто фиксировать факт ошибки, но и обеспечивать учащемуся содержательную обратную связь, необходимую для осознанной фонетической коррекции.

2. Теоретические основания.

Автоматическая оценка произношения (Automated Pronunciation Assessment, АРА) – область на стыке компьютерной лингвистики, акустической фонетики и педагогики. Её теоретический фундамент составляют несколько ключевых направлений.

С точки зрения акустической фонетики, речевой сигнал описывается через набор формантных характеристик (F1, F2, F3), спектральные огибающие, мел-частотные кепстральные коэффициенты (MFCC) и просодические параметры – интонацию, ритм, темп. Именно эти параметры служат основой для алгоритмического сравнения речи учащегося с эталонными образцами.

В методике преподавания иностранных языков ключевым ориентиром остаётся коммуникативный подход: произношение оценивается не в терминах аб-

страктной правильности, а с точки зрения разборчивости и коммуникативной эффективности. Это означает, что система должна уметь разграничивать фонетические отклонения, критически влияющие на понимание, и допустимые акцентные особенности.

С позиций когнитивной психологии, эффективная фонетическая коррекция требует немедленной и конкретной обратной связи. Исследования в области формирования речевых навыков подтверждают, что отсроченный и обобщённый комментарий педагога значительно уступает по эффективности мгновенной поцелевой реакции на конкретный фонетический параметр.

3. Архитектура прототипа.

3.1. Общая структура.

Разработанный прототип представляет собой клиент-серверное приложение, включающее четыре функциональных модуля: модуль захвата и предобработки аудиосигнала, модуль акустического анализа, модуль сравнения с эталоном и модуль формирования обратной связи.

Пользовательский интерфейс реализован в виде веб-приложения, не требующего установки дополнительного программного обеспечения. Учащийся читает предложенный текст вслух; запись мгновенно передаётся на сервер, где происходит основная обработка. Результат возвращается в браузер в течение 1–3 секунд в зависимости от длины фрагмента.

3.2. Акустический анализ.

Для извлечения признаков из речевого сигнала используется библиотека librosa (Python). На первом этапе сигнал разбивается на фреймы длиной 25 мс с перекрытием 10 мс. Для каждого фрейма вычисляются MFCC (40 коэффициентов), а также дельта-признаки первого и второго порядка, характеризующие динамику изменений спектра.

Выявление отдельных фонем осуществляется на основе предобученной модели wav2vec 2.0 (Meta AI), адаптированной для анализа произношения на английском языке. Данная архитектура трансформера, обученная на тысячах часов

размеченной речи, демонстрирует точность фонемной сегментации свыше 90% в условиях бытового шума.

3.3. Сравнение с эталоном и обратная связь.

Эталонная база включает образцы произношения носителей языка (американский и британский варианты английского), размеченные на уровне фонем. Для каждой произнесённой фонемы вычисляется косинусное расстояние между вектором признаков учащегося и центроидом эталонного класса.

По итогам анализа формируется трёхуровневая обратная связь: фонемный уровень – выделение конкретного звука, в котором допущено отклонение; акустический уровень – описание характера ошибки (например, «недостаточная аспирация», «неверное место образования»); практический уровень – артикуляционная рекомендация с иллюстрацией и звуковым образцом. Система также отображает транскрипцию распознанного и эталонного вариантов с цветовой маркировкой расхождений.

4. Технический стек и реализация.

Серверная часть прототипа написана на Python 3.11 с использованием фреймворка FastAPI, обеспечивающего высокую производительность при обработке асинхронных запросов. Акустический анализ реализован в связке librosa + PyTorch; модель wav2vec 2.0 загружается через Hugging Face Transformers. СУБД PostgreSQL используется для хранения пользовательских сессий и динамики ошибок.

Клиентская часть построена на React 18 с применением Web Audio API для захвата микрофонного ввода непосредственно в браузере. Аудиосигнал кодируется в формат WAV (16 кГц, моно) и передаётся на сервер по WebSocket-соединению, что минимизирует задержку и обеспечивает отзывчивость интерфейса.

Для развёртывания используется контейнеризация Docker; конфигурация описана в docker-compose.yml, что упрощает воспроизведение среды на оборудовании учебного заведения без дополнительной настройки. Минимальные системные требования для серверного узла составляют 4 ядра CPU и 8 ГБ RAM; поддержка GPU опциональна и ускоряет инференс в 3–4 раза.

5. Педагогическая интеграция.

5.1. Сценарии использования.

Прототип ориентирован на три основных педагогических сценария. Первый – самостоятельная тренировка: учащийся работает с системой вне класса, многократно отрабатывая целевые фонемы по указанию преподавателя. Система фиксирует прогресс и автоматически повышает сложность упражнений по мере улучшения результатов.

Второй сценарий – аудиторная работа в режиме станций: несколько учащихся одновременно упражняются на индивидуальных устройствах, пока преподаватель работает с отдельными студентами. Это позволяет эффективно использовать аудиторное время без жертвы индивидуальным вниманием.

Третий сценарий – диагностика: первичная запись речевой продукции учащегося на этапе входного контроля автоматически формирует фонетический профиль с указанием проблемных зон. Это даёт преподавателю объективную точку отсчёта и позволяет сфокусировать работу на приоритетных аспектах.

5.2. Соответствие коммуникативному подходу.

Педагогическая концепция системы сознательно опирается на принципы коммуникативного подхода. Оценке подвергается не абстрактное «правильное» произношение, а разборчивость высказывания для носителя языка. Система реализует трёхуровневую шкалу: «коммуникативно неприемлемо», «допустимо с оговорками», «коммуникативно прозрачно» – что соответствует логике реального языкового взаимодействия.

Кроме того, упражнения встраиваются в осмысленные коммуникативные контексты: учащийся озвучивает реплики диалога, комментирует иллюстрацию, описывает ситуацию – а не повторяет изолированные звуки. Это не только делает работу содержательной, но и формирует навык произношения в условиях, приближенных к естественной речи.

6. Апробация и результаты.

Пилотная апробация прототипа проводилась в течение восьми недель на базе языкового центра РТУ МИРЭА с участием двух групп студентов уровня В1

по 15 человек. Контрольная группа занималась по стандартной программе с традиционной коррекцией произношения преподавателем; экспериментальная дополнительно использовала разработанную систему по 20 минут трижды в неделю.

По итогам апробации был проведён сравнительный анализ качества произношения с использованием независимой экспертной оценки (два сертифицированных фонетиста) и автоматической метрики GOP (Goodness Of Pronunciation). Результаты показали, что в экспериментальной группе средний балл GOP вырос на 18,4%, тогда как в контрольной – на 9,1%. Экспертная оценка зафиксировала статистически значимое снижение числа фонемных ошибок в экспериментальной группе ($p < 0,05$).

Особенно заметный прогресс был отмечен в работе с фонемами, не имеющими аналогов в русском языке: межзубными /θ/ и /ð/, а также с английскими дифтонгами. Именно эти зоны традиционно требуют наибольшего числа повторений и индивидуального внимания педагога, и именно здесь автоматизированная система проявила себя наиболее эффективно.

Анкетирование участников показало высокий уровень принятия технологии: 87% студентов отметили, что система помогла им осознать конкретные ошибки, которые раньше они не замечали, а 79% указали на снижение тревожности при последующей коммуникации с носителями языка.

7. Ограничения и перспективы развития.

При всей перспективности подхода необходимо честно обозначить его ограничения. Во-первых, точность системы заметно снижается в условиях акцентированной речи, сильно отличающейся от обучающей выборки (например, речь носителей азиатских языков). Расширение и диверсификация эталонной базы является первоочередной задачей следующего этапа разработки.

Во-вторых, система оценивает изолированные фонетические параметры, но не учитывает в полной мере просодику на уровне связного текста – интонационный контур, фразовое ударение, ритмику. Интеграция просодического анализа запланирована в следующей версии прототипа.

В-третьих, принципиально важно понимать: автоматизированная оценка не заменяет педагога, а дополняет его. Система не способна учесть индивидуальные речевые особенности учащегося, его коммуникативные цели и эмоциональное состояние. Роль преподавателя остаётся определяющей, прежде всего в мотивации и методическом сопровождении работы с системой.

Среди перспективных направлений развития следует выделить: адаптацию системы для других изучаемых языков (немецкий, французский, китайский); разработку модуля сверхсегментной фонетики; создание учительской панели с агрегированной статистикой по группе; интеграцию с популярными LMS-платформами (Moodle, Canvas).

8. Рекомендации педагогам.

На основании проведённой апробации и анализа педагогических эффектов системы можно сформулировать следующие практические рекомендации.

Систему целесообразно вводить постепенно, начиная с одного-двух целевых фонемных контрастов, а не с полного набора проблемных звуков. Это позволяет учащемуся сосредоточить внимание и не перегружает его информацией.

Работа с системой должна сопровождаться кратким предварительным инструктажем преподавателя о том, что именно оценивается и как интерпретировать обратную связь. Без такого объяснения часть студентов воспринимает автоматическую оценку как механическое осуждение, что вызывает избыточную тревожность.

Оптимальный режим использования, выявленный в ходе апробации, – три сессии по 15–20 минут в неделю. Более длительные сессии утомляют учащихся и снижают качество фонетического контроля.

Результаты работы с системой следует регулярно обсуждать с учащимися, помогая им соотносить технические метрики с реальным опытом общения. Такое осмысление переводит работу над произношением из механического повторения в осознанный навык.

9. Заключение.

Разработанный прототип системы автоматизированной оценки произношения демонстрирует реальный педагогический потенциал: объективная, мгновенная и детализированная обратная связь существенно ускоряет фонетическую коррекцию и повышает осознанность учащегося в работе над произношением. Апробация подтвердила статистически значимые преимущества подхода по сравнению с традиционным обучением без инструментальной поддержки.

Вместе с тем ключевым условием педагогической эффективности системы остаётся её грамотное методическое встраивание в учебный процесс. Автоматизированная оценка – это не самоцель и не замена живому педагогическому взаимодействию. Её подлинная ценность раскрывается только в связке с профессиональным сопровождением преподавателя, способного направить технические данные на службу реальному коммуникативному развитию учащегося. В этом союзе педагогического мастерства и технологической точности рождается новое качество языкового образования.

Список литературы

1. Бондаренко Е.В. Автоматическое распознавание речи в системах электронного обучения: обзор технологий / Е.В. Бондаренко, Т.А. Сидорова. – М.: МГТУ, 2022. – 48 с.
2. Гальскова Н.Д. Теория обучения иностранным языкам: Лингводидактика и методика / Н.Д. Гальскова, Н.И. Гез. – 3-е изд. – М.: Академия, 2022. – 336 с.
3. Касаткин А.О. Мел-частотные кепстральные коэффициенты и их применение в задачах распознавания речи / А.О. Касаткин // Информационные технологии. – 2021. – №4. – С. 34–41.
4. Кривнова О.Ф. Фонетика и автоматическая обработка речи / О.Ф. Кривнова. – М.: МГУ, 2020. – 214 с.
5. Мильруд Р.П. Методика преподавания английского языка / Р.П. Мильруд. – 2-е изд. – М.: Дрофа, 2007. – 257 с. EDN SDTWMF
6. Пассов Е.И. Урок иностранного языка / Е.И. Пассов, Н.Е. Кузовлева. – Ростов н/Д.: Феникс, 2020. – 640 с.

7. Разинкина И.В. Технологии распознавания речи в образовании: возможности и ограничения / И.В. Разинкина // Педагогика и психология образования. – 2023. – №2. – С. 88–97.
8. Титова С.В. Цифровые технологии в языковом образовании / С.В. Титова. – М.: Эдитус, 2021. – 189 с.
9. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations / A. Baevski, H. Zhou, A. Mohamed, M. Auli // NeurIPS. – 2020. – Vol. 33. – P. 12449–12460.
10. Cucchiarini C. Automatic Pronunciation Assessment of Second Language Learners' Speech / C. Cucchiarini, H. Strik // Language Assessment Quarterly. – 2021. – Vol. 18. No. 2. – P. 105–122.
11. Levis J. Automatic Speech Recognition / J. Levis, R. Suvorov // The TESOL Encyclopedia of English Language Teaching. – 2020. – P. 1–8.
12. Neri A. Feedback in Computer Assisted Pronunciation Training / A. Neri, C. Cucchiarini, H. Strik // CALL. – 2022. – Vol. 35. No. 3. – P. 234–251.
13. Witt S.M. Automatic Error Detection in Pronunciation Training: Where We Are and Where We Need to Go / S.M. Witt // IS ADEPT. – 2022. – URL: <https://www.isca-speech.org> (date of access: 20.03.2026).