

УДК 004.8:81'23:616.89

## НЕЙРОСЕТЕВОЙ ЛИНГВИСТИЧЕСКИЙ ТРЕНАЖЁР ДИАГНОСТИКИ И КОРРЕКЦИИ РЕЧЕВЫХ НАРУШЕНИЙ

<sup>+</sup> Абдуллаева М.И.<sup>1</sup>, Каримова М.Б.<sup>1</sup>

<sup>1</sup> Ташкентский университет информационных технологий имени  
Мухаммада ал-Хоразмий, Ташкент, Узбекистан

<sup>+</sup> m.abdullaeva@tuit.uz

**Аннотация.** На сегодняшний день разработка систем коррекции речи общего недоразвития речи у детей дошкольного возраста на основе внедрения речевых технологий является актуальной и приоритетной задачей. В данной статье описана структура лингвистического тренажера включающего в себя модули диагностирования на основе нечёткой логики, интерпретации и понимания запросов пользователей и генерации ответа на базе NLU-подходов, синтезирование речи на основе текста для построения удобной взаимосвязи между пользователем и системой, адаптивных и специализированных упражнений с использованием педагогических технологий, аналитического блока на базе специальных метрик, таких как метрика прогресса, темп прогресса, метрика освоения речевой единицы, метрика посещаемости, которые в совокупности представляют заключения ИИ логопеда. В работе каждая из модулей описаны как с точки зрения структуры, так и эффективных подходов. На детальные результаты моделей нейронных сетей лингвистического тренажера приведены литературные источники. На основе анализа выявлено, что метрика темп прогресса, вычисляемый лингвистическим тренажером, составляет 1,2%, тогда как традиционная реабилитация 0,6. При использовании тренажера темп формирования навыка в рамках рассматриваемой речевой единицы в среднем увеличилось ~1,7 раз, а показатель абсолютного прироста точности произношения (ΔР) освоения речевой единицы, которая указывает, что за одинаковый период реабилитации речи при использовании тренажера достигается ускорение процесса в 4,5–5 раз.

**Ключевые слова:** оценка произношения, модуль генерации речи, NLU-модуль, цифровая логопедия, лингвистический тренажёр.

### 1 ВВЕДЕНИЕ

Сегодня особое внимание уделяется разработке информационных систем, основанных на речевых технологиях. Разрабатываемые в настоящее время системы речевого интерфейса напрямую связаны с анализом [1] и синтезом речи [2]. Поскольку речевой диалог является основным способом человеческого общения, разработка интеллектуальной системы на основе речевых компонентов является одним из основных направлений искусственного интеллекта. Речевой интерфейс состоит из таких базовых речевых технологий, как анализ и синтез речи [3, 4]. В развитых странах, таких как США, Российская Федерация, Китай, Япония и др., ведутся научные исследования, направленные на развитие теоретических и практических работ, направленных на совершенствование существующих методов и алгоритмов анализа и синтеза речи, а также на разработку методов интеллектуальной обработки речевых сигналов. Сегодня разработка систем, способных обеспечить полноценный диалог человека и машины с использованием речи, является ключевой задачей в области цифровой обработки сигналов и искусственного интеллекта [5].

Во всем мире растёт число детей с различными видами языковых и речевых нарушений. Нарушения языка (НЯ) являются одним из наиболее распространённых типов особых образовательных потребностей (ООП), не только как основная потребность ребёнка, но и как вторичная область, связанная с другими типами ООП. НЯ является фактором риска для дальнейшего развития детей, будучи связанным с усилением поведенческих, эмоциональных и социальных трудностей, в частности, проблем со сверстниками и эмоциональных трудностей; трудностей с грамотностью, включая как чтение, так и письмо; и снижения уровня академической успеваемости. Риски, возникающие из НЯ в раннем детстве, могут также оказывать влияние в подростковом возрасте и во взрослой жизни.

Согласно статистике в [6] диагноз «НЯ» вырос с ~9,0 % в 2018 году до ~11,8 % в конце 2021 года и ~16,9 % в начале 2022 года. К сожалению, от года в год наблюдается заметное увеличение количества детей с НР проблемами. В связи с чем, разработка информационной системы (лингвистического тренажёра) для реабилитации слухоречевых навыков детей дошкольного возраста является актуальной задачей.

В данной статье описаны программные модули лингвистического тренажера для коррекции слуха и речи детей на основе речевых и интеллектуальных технологий с элементами последующего мониторинга результатов детей.

## 2 ОСНОВНАЯ ЧАСТЬ

Возраст 4-6 лет является наиболее благоприятным периодом для реабилитации речи, поскольку именно в это время происходит активное созревание ключевых зон мозга, отвечающих за понимание и продуцирование речи – областей Брока и Вернике. Логопеды отмечают, что в этом возрасте ребёнок особенно восприимчив к коррекции речевых трудностей, а сформированная игровая мотивация делает процесс обучения естественным, увлекательным и эффективным. Благодаря сочетанию нейрофизиологической готовности и высокой эмоциональной вовлечённости, использование тренажёров на базе речевых технологий и с широкой анимационной базой становится наиболее результативным способом поддержки речевого и когнитивного развития детей [7,8].

В работах [9,10] показано, что логопедическая диагностика речевых отклонений у детей основывается на комплексной оценке ключевых компонентов речевого развития, включая фонематический слух, звукопроизношение, лексико-грамматические навыки, связную речь и понимание устных инструкций. Эти диагностические направления особенно актуальны при выявлении общего недоразвития речи (ОНР), поскольку данное нарушение затрагивает все перечисленные уровни языковой системы. ОНР проявляется в стойких затруднениях формирования звуков, ограниченности словарного запаса, аграмматизмах и нарушениях связного высказывания, что существенно влияет на коммуникативное и когнитивное развитие ребёнка [11].

Таким образом, комплексная логопедическая диагностика, описанная в литературе, является необходимым условием своевременного выявления ОНР и определения индивидуальной коррекционной траектории, что позволяет минимизировать последствия нарушения и повысить эффективность последующей речевой реабилитации.

Таблица-1 отражает научно-установленное соответствие между уровнем ОНР и необходимыми коррекционными задачами.

**Таблица 1.** Таблица соответствия: уровень ОНР – задачи – тип упражнений

Уровень ОНР	Особенности	Что тренировать	Примеры заданий
<b>I (тяжёлый)</b>	Нет фразовой речи, отдельные слова, понимание нарушено	Понимание речи, базовые слова, звукоподражание	«Покажи предмет», «Повтори слово», звукоподражание, простые слоги
<b>II (средней тяжести)</b>	Есть фразы, но с ошибками, бедный словарь	Расширение словаря, грамматика, фраза	Составь фразу, выбор окончания, упражнения на согласование
<b>III (лёгкий)</b>	Фраза есть, но с грамматическими нарушениями, бедная связная речь	Связная речь, пересказ, сложные предложения	Пересказ по картинке, составь рассказ, найди ошибку

Для I уровня, характеризующегося отсутствием фразовой речи и выраженными трудностями понимания, доказанно эффективны упражнения на развитие базового понимания, расширение элементарного словаря и формирование первых слоговых структур. На II уровне, где речь представлена фразами с ошибками и ограниченным словарём, исследования показывают необходимость целенаправленного формирования лексико-грамматических категорий и структурирования фразы. На III уровне, при сохранении грамматических нарушений и недостаточной связности речи, приоритет отдается тренировке связных высказываний, пересказу и формированию сложных предложений. Таблица-1 фиксирует емкую и доказательную зависимость: чем выше уровень ОНР, тем более сложные языковые операции подлежат тренировке.

На рис.1 приведены главные составляющие модули тренажера по реабилитации речи и слуха детей с ОНР, которыми являются модуль синтеза речи, модуль оценки правильности речи детей, модуль компоненты.

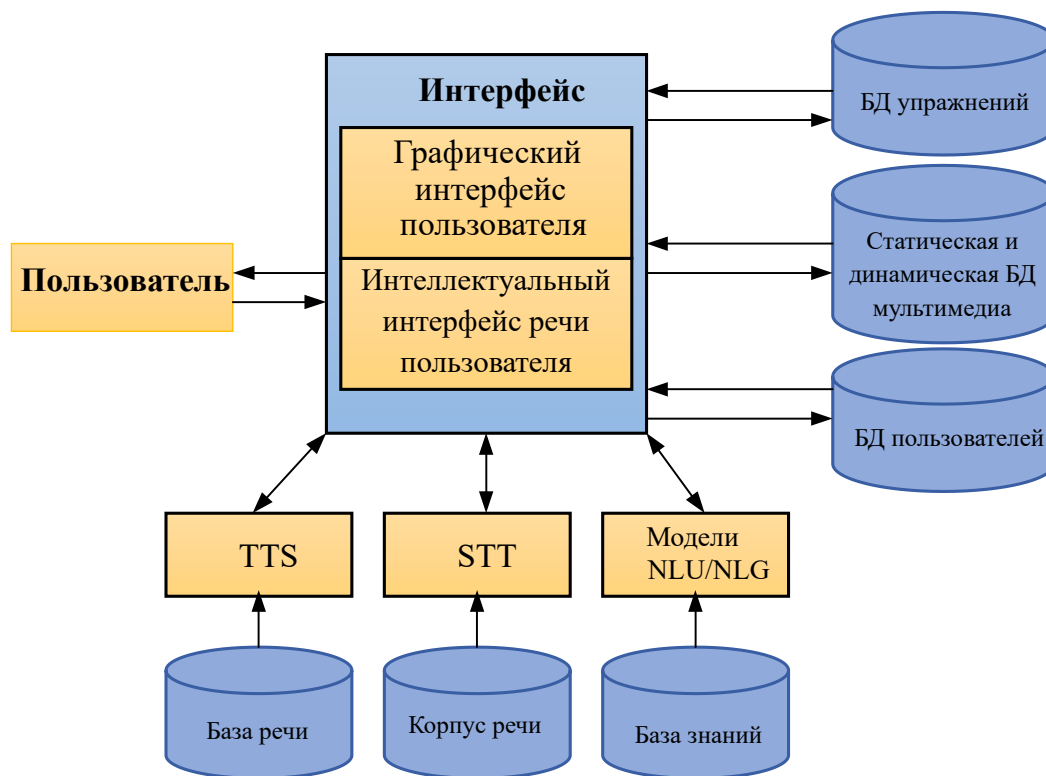


Рис. 1. Программные модули разрабатываемого лингвистического тренажёра

## 2.1 Модуль синтеза речи (TTS)

В системах с речевым интерфейсом для формирования голосового ответа необходима разработка модуля синтеза речи. Синтез речи – это компьютерное моделирование человеческой речи из текстового представления с использованием методов машинного обучения.

Первая система синтеза речи под названием «*têtes parlantes*» (говорящие головы), появилась в XVIII веке и была первопроходцем, однако представляла собой несовершенную имитацию человеческого голоса.

Естественная речь состоит из коартикуляции, т.е. эффекта соединения двух звуков вместе, и просодии на уровне слога, слова, предложения, которые не могут быть синтезированы простым соединением фонем.

Современные системы преобразования текста в речь (TTS) проходят через два главных этапа: front-end и back-end.

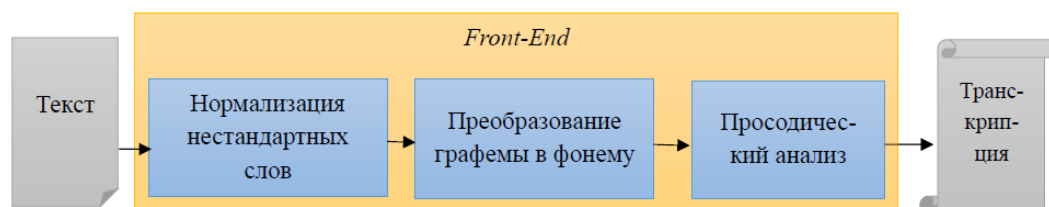


Рис. 2. Последовательность действий Front-End среды

Front-End этап, который преобразовывает текст в его транскрипционное представление. Back-end среда представляет из себя синтезатор, здесь производится генерация речи посредством преобразования каждой единицы транскрипции в звук с помощью выбранного метода и алгоритма.

Среди существующих методов синтеза речи по критериям естественности, качества и точности наиболее высокие показатели имеют современные TTS системы. Современные TTS системы опираются на глубокую нейронную сеть, состоящую из акустической модели и нейронного вокодера для

аппроксимации параметров и отношений между входным текстом и формой волны, которые составляют речь (рис. 4).

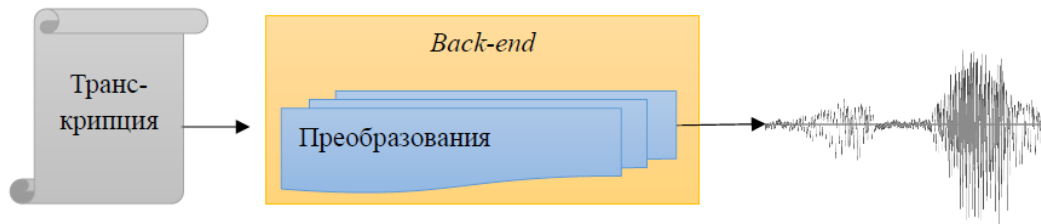


Рис. 3. Последовательность действий Back-end среды

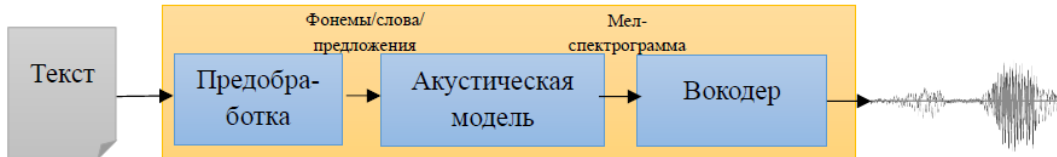


Рис. 4. Схема работы двухступенчатой системы TTS

## 2.2 Модуль оценки правильности речи (STT)

Для проверки правильности произношения был разработан модуль, который рассчитывает метрическую оценку основываясь на алгоритме Dynamic Waring (DW) с использованием акустической модели отклонения.

Пусть имеется эталонное произношение носителя языка, представленное последовательностью признаков временного ряда  $X=(x_1, x_2, \dots, x_n)$ ,  $x_i \in R^d$  и произнесение обучаемого  $Y=(y_1, y_2, \dots, y_m)$ ,  $y_j \in R^d$ , где обычно в качестве признаков используются 39-мерные векторы MFCC+ $\Delta$ + $\Delta\Delta$  ( $d = 39$ ). Требуется вычислить скалярную оценку правильности произношения  $\in [0, 100]$ , где 100 – идеальное совпадение с носителем.

Оптимальное DW-расстояние определяется как  $\min_{P \in \rho_n, m} \sum c(x_i, y_j)$ ,  $(i, j) \in P$  где  $c(x_i, y_j) = \|x_i - y_j\|^2$  – локальная евклидова стоимость,  $\rho_n, m$  – множество всех допустимых путей выравнивания с условиями монотонности и непрерывности. Данный функционал эффективно вычисляется динамическим программированием  $D(i, j) = c(x_i, y_j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\}$  с граничными условиями  $D(0, 0) = 0$ ,  $D(i, 0) = D(0, j) = +\infty$  (или суммирование по границам при разрешённых вставках удалений).

Прямое применение исходного DW-расстояние оказалось методологически ограниченным, поскольку его значения зависят от длительности речевого сигнала, не имеют верхней границы и демонстрируют низкую интерпретируемость в педагогической шкале. В связи с этим применена нормированная схема вычисления, обеспечивающая сопоставимость результатов и их согласованность с экспертными суждениями.

На первом этапе определяется оптимальный путь выравнивания  $P^*$  с минимальной суммарной стоимостью и его длина  $L^* = |P^*|$ . Далее выполняется нормировка расстояния по длине пути, что позволяет получать среднее акустическое отклонение на один сопоставленный кадр:

Вычисление длины оптимального пути выравнивания  $L^* = |P^*|$  (количество шагов в пути с минимальной стоимостью). Нормировка DW-расстояния по длине пути (path-normalized DW):

$$DTW_{norm}(X, Y) = \frac{1}{L^*} \sum_{(i, j) \in P^*} \|x_i - y_j\|^2.$$

Такой подход устраняет зависимость от абсолютной продолжительности сигнала и обеспечивает метрическую стабильность. Для преобразования нормированного расстояния в педагогически интерпретируемую шкалу применяется экспоненциальная функция вида:

$$S = 100 \cdot \exp(-\alpha \cdot DTW_{norm}(X, Y)),$$

где коэффициент  $\alpha > 0$  – калибруемый параметр чувствительности (в наших экспериментах  $\alpha = 0.045$  при использовании 39-мерных MFCC). Данная формула обладает следующими свойствами.  $S = 100$

при идеальном совпадении ( $DTW=0$ ). Монотонно убывает с ростом акустического отклонения. Соответствует субъективным оценкам экспертов в диапазоне 0–100 (проверено корреляцией Пирсона  $r = 0.91$ ). Для снижения вычислительной сложности и предотвращения патологического выравнивания (pathological warping) применена полоса Сакоэ–Тиббы радиусом  $r = \lfloor \max(n,m)/10 \rfloor$ , что уменьшает сложность с  $O(nm)$  до  $O(n \cdot r)$ . Также реализована симметричная версия DTW (Symmetric DTW) с шаговым шаблоном типа II по Itakura, обеспечивающая выполнение треугольного неравенства в приближении и повышающая стабильность оценки. Итоговая формула оценки правильности речи выглядит следующим образом:

$$S_{DTW}(X, Y) = 100 \cdot \exp \left( -\alpha \cdot \frac{1}{L^*} \sum_{(i,j) \in P^*} x_i - y_{j2} \right),$$

где  $P^* = \arg \min P \in \rho_{n,m}^r \sum_{(i,j) \in P} \|x_i - y_j\|_2$  - оптимальный путь в ограниченной полосе  $r$ .

### 2.3 Модуль понимания текста на основе NLU

С помощью модуля NLU все вошедшие в систему запросы обрабатываются, анализируются и понимаются «на уровне разума». Он интерпретирует высказывания ребенка, точно определяет, что именно пользователь хочет сказать или спросить, выделяет в высказывании ключевые для него элементы и правильно трактует запросы на подобии «повторить», «объяснить ошибку» или «дать упражнение». Вместе с модулем распознавания речи NLU-модуль определяет тип речевой ошибки, корректирует замены, пропуски, искажения и неправильную редукцию фонем, а также ошибки в просодии. Далее он соотносит проблему, выявленную в ходе диагностики, с базой методических рекомендаций и подбирает средства, соответствующие возрасту ребёнка и текущему уровню его речевого развития.

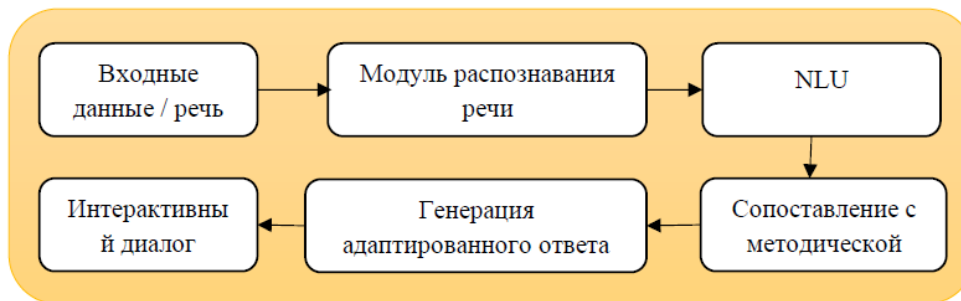


Рис. 5. Этапы обработки входного запроса в диалоговых системах

На выходе модуль NLU формирует комплексное интерактивное обратное связующее звено в адаптированной форме. К тому же, NLU-модуль полностью ответственный за проведение диалога: сохраняет контекст беседы, отвечает на уточняющие вопросы пользователей и в зависимости от реакции ребёнка динамически выбирает наиболее подходящую стратегию коррекции.

Для реализации модуля NLU и интерактивного блока «вопрос–ответ» в предлагаемой системе используется функциональная архитектура, опирающаяся на методы и результаты, полученные в исследовании Хужаяров И. и др. по разработке узбекских Q-A моделей на основе fine-tuning современных трансформерных архитектур (T5, Flan-T5, mT5 и Llama)[13].

Ключевой особенностью этой архитектуры является то, что NLU-модуль не просто классифицирует запрос, но и выполняет семантический анализ вопроса, определяет его контекстуальную релевантность и формирует входные представления (embeddings), необходимые для блока извлечения знаний (retrieval).

Благодаря использованию предобученных и дообученных моделей T5/mT5 и Llama, адаптированных авторами к узбекскому языку на корпусе более чем из 10 000 пар (контекст–вопрос–ответ), достигается корректная семантическая сегментация вопросов, понимание вариативных формулировок, включая разговорные и административно-правовые конструкции, более точное выделение ключевых сущностей (служба, документ, срок, реквизит), устойчивость к орфографическим и морфологическим вариациям узбекского языка.

В полученных исследованиях [14] были использованы модели T5, Flan-T5, mT5 и Llama, дообученные на локальном доменном датасете. Эти модели продемонстрировали высокие показатели на метриках BLEU, ROUGE и METEOR, что позволило включить их в интерактивный Q-A модуль нашей системы.

В рамках интеграции в разрабатываемую архитектуру NLU-модуль использует T5/mT5-модели для нормализации вопроса, переформулирования сложных запросов (query reformulation), выделения типа пользовательского запроса (информационный, справочный, транзакционный). Блок анализа (Analyze) использует fine-tuned модели Llama или Flan-T5 для оценки соответствия вопроса контексту базы знаний, ранжирование потенциальных ответов, выбор оптимального варианта ответа с учётом семантического сходства, вычисленного через cosine similarity или BM25. Интерактивная обратная связь реализуется посредством генерации уточняющих вопросов, если контекст неполный, объяснения ответа в естественном языке (explainability generation), адаптивной переформулировки ответа под уровень пользователя. Использование RAG-архитектуры (Retrieval-Augmented Generation), описанной в исследовании Kshirajagar и др., позволяет извлекать наиболее релевантные документы из ChromaDB, формировать расширенный контекст и генерировать ответы на узбекском языке с высокой точностью даже в условиях узкой предметной области (например, документы, регламенты, госуслуги, call-center).

Проведённые авторами эксперименты подтверждают, что адаптированные к узбекскому языку модели значительно повышают качество обработки запросов в закрытой доменной среде (в отличие от универсальных LLM, которые дают более общие или «галлюцинирующие» ответы).

Поэтому интеграция моделей, дообученных на локальном датасете, позволяет обеспечить высокую точность понимания вопросов в сфере госуслуг и колл-центров, улучшить интерактивность и точность диалога с пользователем, сократить количество ошибок благодаря тонкой настройке (fine-tuning) и улучшить работу всей цепочки:

*Понимание вопроса → Поиск самого близкого по значению контекста → Генерация ответа на основе контекста на входной вопрос → Обратная связь.*

Предлагаемая архитектура лингвистического тренажёра представляет собой многоуровневую цифровую систему, обеспечивающую полный цикл диагностики, обучения и мониторинга речевого развития ребёнка. На первом этапе осуществляется вход в систему, предусматривающий использование современных методов идентификации (логин и пароль, Face ID, голосовая аутентификация), что гарантирует безопасность данных и персонализацию обучения. После авторизации пользователь проходит этап диагностирования, включающий тестирование уровня речевого развития. На основе полученных результатов система автоматически формирует рекомендации и обеспечивает выбор индивидуального модуля, соответствующего речевому профилю ребёнка.

Каждый обучающий модуль включает комплекс упражнений, направленных на коррекцию отдельных компонентов речи, а выполнение заданий сопровождается интерактивной обратной связью: система анализирует ответы ученика, формирует корректирующие подсказки и ведёт автоматизированный диалог в процессе обучения. Для поддержания устойчивой учебной мотивации внедрён блок игровой геймификации, включающий награды, анимации и прогресс-бар, что повышает вовлечённость ребёнка и способствует регулярности занятий. Все результаты фиксируются и сохраняются в базе данных, обеспечивая возможность динамического мониторинга. На заключительном этапе встроенный ИИ-логопед анализирует накопленные данные, оценивает успеваемость, формирует отчёты и корректирует последующие обучающие траектории, создавая адаптивную, научно обоснованную и персонализированную систему речевой реабилитации.

Использование цифрового лингвистического тренажёра позволяет значительно ускорить процесс речевой реабилитации: согласно исследованию, большинство детей, ожидающих терапию, за 12 месяцев самостоятельного ожидания показали лишь изменение на 0,58 балла по шкале TOM-I, и при этом 83% из них всё ещё нуждались в помощи специалиста [12]. Шкала нарушения (Impairment) в рамках Therapy Outcome Measures (ТОМ-I) представляет собой 6-балльную ординальную шкалу (0–5 баллов, с возможностью полубаллов), где 0 баллов соответствует отсутствию речевого нарушения, соответствующему возрастной норме, а 5 баллов – полному или крайне тяжёлому нарушению функциональной устной речи. Внедрение активных тренажёров с регулярными интерактивными заданиями, обратной связью и мониторингом результатов может сократить время до выраженного прогресса более чем на 50%, позволяя ребёнку перейти к успешной коммуникативной активности быстрее и с меньшими рисками задержек.

### 3 ДИАГНОСТИРОВАНИЕ

Диагностика общего недоразвития речи у детей требует оценки множественных параметров, которые не поддаются жёсткой бинарной классификации, таких как степень искажения фонемы, уровень просодических нарушений, стабильность артикуляции, динамика ошибок во времени и др.

Классические пороговые и вероятностные методы (например, threshold-based и softmax-классификация) плохо отражают промежуточные состояния, характерные для детской речи. Поэтому в архитектуре тренажёра реализована нечёткая диагностическая модель, которая позволяет формировать интерпретируемые логопедические заключения на основе акустических, семантических и поведенческих параметров.

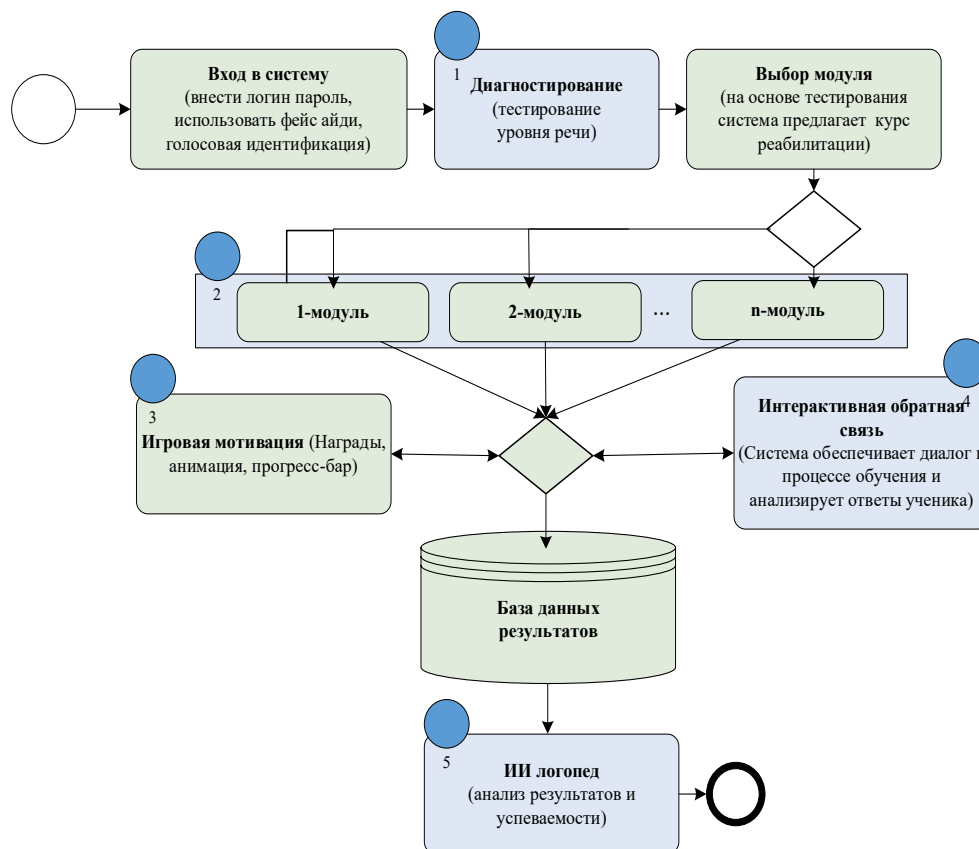


Рис. 6. Функциональная диаграмма процесса работы лингвистического тренажёра

Для формализации диагностических решений в системе коррекции речевых нарушений используется нечёткий модуль вывода, опирающийся на совокупность лингвистически значимых параметров. В качестве входных переменных нечеткие-системы выбраны показатели, отражающие ключевые аспекты акустико-артикуляционных и когнитивно-лингвистических особенностей произношения ребёнка.

Акустическое отклонение (А) характеризует уровень акустического расхождения меж эталонным и фактическим произнесением. Для его вычисления используется нормированная метрика DTW (Dynamic Time Warping), позволяющая учитывать вариативность временной структуры сигнала. Значения нормируются в диапазоне 0–1, где нулевое значение соответствует высочайшей близости к эталонной форме, а одиночное – максимальному отклонению.

Артикуляционная стабильность (S) отражает устойчивость артикуляционного паттерна ребёнка и вычисляется как дисперсия сегментированных фонем в серии повторных произнесений. Повышенная вариабельность артикуляции указывает на несформированность моторного контроля и типична для детей с ОНР.

Просодическая корректность (P) описывает ритмико-мелодические характеристики речи, включая продолжительность и структуру слогов, распределение ударности, соблюдение темпо-ритмических паттернов. Нарушения просодии считаются одним из ранних маркеров отклонений в речевом развитии и оказывают значительное воздействие на восприятие высказывания.

Лингвистическая точность (L) складывается NLU-модулем и отражает правильность выбора текстов и слогов, грамматико-семантическую адекватность ответа, а также дееспособность ребёнка следовать смысловой структуре задания.

Для каждой входной переменной ориентируется набор нечётких термов, таких как низкое, среднее и высокое значение показателя. Функции принадлежности задаются трапецеидальными и гауссовыми кривыми, собственно, что обеспечивает плавность переходов меж категориями и корректное моделирование реальной вариативности детской речи. Такое описание параметров разрешает

нечеткому-модулю возводить интерпретируемые диагностические заключения, учитывающие многомерную природу речевого нарушения.

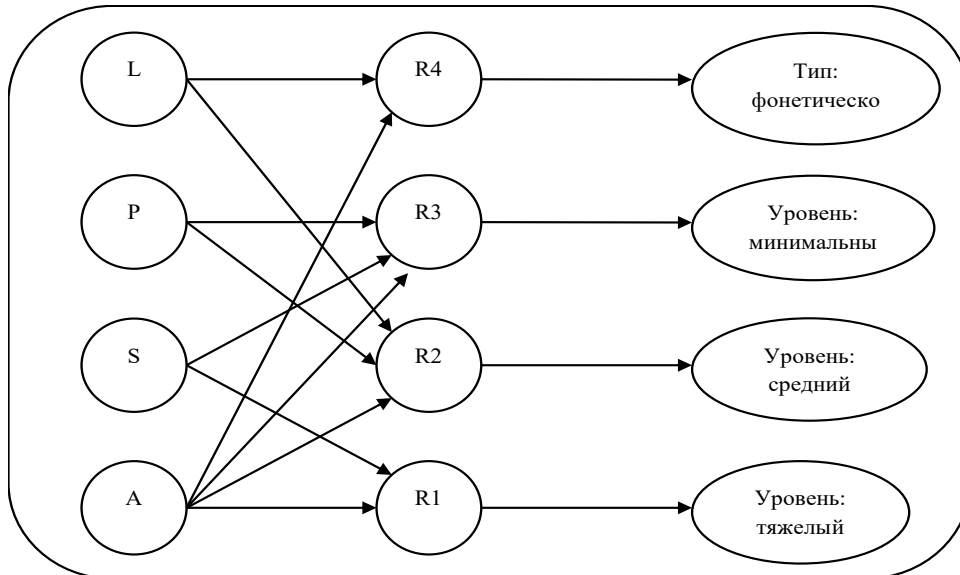


Рис. 7. Визуализация нечеткой базы правил, в рамку

Для моделирования нескончаемой природы речевых отклонений применяется параметризация функций принадлежности, описывающая уровень соотношения входных значений лингвистическим термам. Ниже приведён образчик формализации для переменной акустическое отклонение (A).

Функции принадлежности для акустического отклонения A:

**Низкое отклонение:**

$$\mu_{low}(A) = \exp\left(-\left(\frac{A}{0.15}\right)^2\right).$$

Такая гауссова форма разрешает тактично отражать случаи, когда ребёнок практически буквально воспроизводит эталонный образец, а промахи имеют случайный характер.

**Среднее отклонение:** Трапецеидальная функция принадлежности с вершинами (0.15, 0.25, 0.35, 0.45) собственно, что соответствует плавному переходу от слабых к выраженным акустическим нарушениям.

**Высокое отклонение:**

$$\mu_{high}(A) = 1 - \exp\left(-\left(\frac{A - 0.4}{0.2}\right)^2\right).$$

Функция обеспечивает однообразный подъем принадлежности по мере наращивания отличия и действительно характеризует тяжкие нарушения.

Такой выбор параметров отражает плавный, градуальный характер речевых промахов у ребят, где границы меж нормой и патологией часто размыты.

Диагностика основывается на экспертной базе правил, разработанной вместе с логопедами-дефектологами Узбекистана подключить сколько правил подключили. Примеры правил:

**Если A = высокое и S = нестабильная**

**то** уровень нарушения = тяжёлый.

**Если A = среднее и P = частично нарушена и L = корректная**

**то** уровень нарушения = средний.

**Если A = низкое и S = стабильная и P = норма**

**то** уровень нарушения = минимальный.

**Если A = высокое, но L = корректная**

**то** нарушение вероятнее фонетическое, а не семантическое.

...

Экспертная база правил имеет порядка 60 продукционных правил, отражающих взаимосвязи меж акустическими, артикуляционными и лингвистическими параметрами речи. База правил разрешает классифицировать образ обнаруженного отличия фонетическое, фонематическое, просодическое или же комбинированное, обеспечивая интерпретируемость результатов и соотношение методическим принципам логопедической практики.

В рамках нечёткой системы логопедической диагностики положение речи ребёнка описывается обилием входных диагностических признаков  $X = \{X_1, X_2, \dots, X_n\}$ , каждый из коих отражает отдельный степень речевой организации и представляется в облике лингвистической переменной с надлежащими терминами нечётких множеств.

$X_1$  - акустический показатель, характеризующий уровень отличия произношения от эталонного образца (нормализованное DTW-расстояние, вариативность формантных частот и временных параметров);

$X_2$  - просодический показатель, отражающий стабильность темпа, ритмической структуры и интонационного контура речи;

$X_3$  - артикуляционный показатель, описывающий точность реализации звуков, наличие пропусков, искажений и замен;

$X_4$  - лингвистический показатель, характеризующий лексико-грамматическую правильность и семантическую адекватность речевого ответа.

Выходной переменной системы считается диагностический показатель  $Y$ , который определяет образ и уровень выраженности речевого нарушения.

Ассоциация меж входными признаками и диагностическим заключением формализуется с использованием набора продукционных правил вида:

$$\text{Если } X_1 = A_i \wedge X_2 = B_j \wedge X_3 = C_k \wedge X_4 = D_l \Rightarrow Y = E_m,$$

где  $A_i, B_j, C_k, D_l$  – нечёткие термины (например, низкий, средний, высокий, стабильный, нестабильный), а  $E_m$  – соответствующий класс речевого нарушения или же степень его выраженности.

Например, правило «Если  $X_1 = \text{высокий} \wedge X_2 = \text{нестабильный} \Rightarrow Y = \text{Тяжелый}$ », интерпретируется как высокий уровень акустического отличия в сочетании с нестабильной просодикой, что соответствует тяжкой форме речевого нарушения.

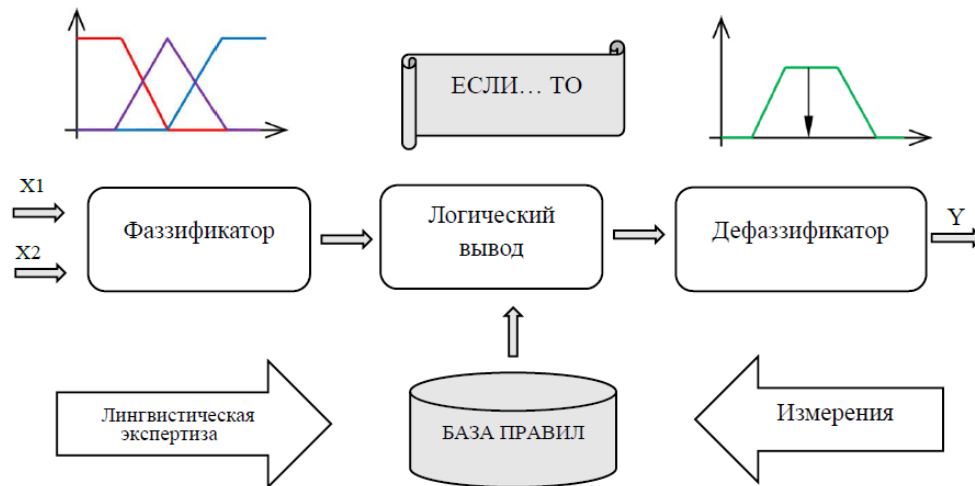


Рис. 8. Архитектура нечеткой логики

Для реализации диагностического ядра используется классическая схема нечёткого вывода Мамдани, которая наиболее адекватно отражает экспертный характер логопедической оценки. Агрегирование лингвистически релевантных показателей осуществляется с использованием стандартных операций. Логическая связка AND реализована оператором минимума  $\mu_{AND}(x) = \min(\mu_i(x), \mu_j(x))$ , а логическая связка OR – оператором максимума  $\mu_{OR}(x) = \max(\mu_i(x), \mu_j(x))$ .

После применения базы правил формируется результирующая функция принадлежности  $\mu_{out}(z)$ . Для дефаззификации используется метод центра тяжести (Center of Area):

$$D = \frac{\int z \cdot \mu_{out}(z) dz}{\int \mu_{out}(z) dz}.$$

Полученное значение  $D \in [0,1]$  преобразуется в клиническую шкалу выраженности нарушения, от 0,0 до 0,3 означает лёгкое нарушение, от 0,3 до 0,6 – среднее, и от 0,6 до 1,0 – выраженное. Такой расклад разрешает формировать интерпретацию, количественно воспроизводимую и сопоставимую оценку речевого дефекта.

Главным составляющей разработанной архитектуры считается интеграция двух гетерогенных источников информации: акустических параметров произнесения и семантико-лингвистических признаков, извлекаемых средствами NLU. Такое сочетание разрешает формировать больше истинный и клинически обоснованный диагностический вывод.

Акустический модуль (DTW-сравнение, нормализованные расстояния, акустические эмбеддинги модели речи) обеспечивает беспристрастную оценку отклонений звучания, включая нарушения артикуляционной базы, искажения частей, редукцию звуков, некорректности проигрывания продолжительности и временной структуры слога. Данные параметры отражают большей частью фонетический составляющую речевого нарушения.

NLU-модуль воплотит в жизнь когнитивно-лингвистическую интерпретацию высказывания, определяя семантическую правильность ответа, соотношение памяти, уместность выбранного слова или же слога, а также речевое намерение ребёнка. Эти показатели отражают степень сформированности фонематического восприятия, понимание задания и умение оперировать языковыми единицами.

Интеграция результатов акустического и NLU-анализа предоставляет ряд значимых преимуществ. Во первых, обеспечивает различие искажённого звука от неправильного выбора языковой единицы, собственно что считается принципиальным для разграничения фонетических и фонематических нарушений. Во вторых, учитывает контекст предъявленного задания и специфику речевой реакции ребёнка. В 3, разрешает обнаружить смешанные формы нарушений (фонетико-фонематические, просодические). В четвертых, снижает возможность ложноположительных срабатываний, вызванных наружными шумами, эмоциональным напряжением или же вариативностью произнесения, а также повышает стабильность и точность итогового диагностического вывода.

Таким образом, разработанный нечёткий-модуль делает роль интеграционного слоя, объединяющего беспристрастно измеряемые акустические параметры с семантической и контекстной информацией NLU-блока. Такое заключение обеспечивает полный, интерпретируемый и клинически валидный анализ устной речи в рамках автоматизированной логопедической диагностики.

Результатом работы нечёткого модуля считается структурированная диагностическая оценка, включающая степень выраженности нарушения, рекомендованный образ упражнения.

Использование нечётко-логического подхода предоставляет немаловажные преимущества по сравнению с традиционными пороговыми методами, потому что разрешает тактично обрабатывать промежуточные, размытые состояния, характерные для речи дошкольников, где границы меж нормой и ошибкой нестабильны и континуальны. Нечёткие множества обеспечивают плавные переходы меж уровнями отклонений, а интерпретируемость правил делает выводы системы понятными для логопедов, педагогов и опекунов, собственно что тем более важно в медицинской и образовательной практике. Благодаря стойкости к акустическим и поведенческим артефактам – таким как вариативность произнесения, беспокойство ребёнка или же домашние шумы – система показывает надёжность в реальных критериях использования. Гибкая настройка функций принадлежности и лингвистических термов разрешает адаптировать систему под особенности узбекского языка, учитывая различия в фонетике, ударении и слоговой структуре. Вместе сформированная с логопедами-дефектологами Узбекистана база правил обеспечивает соотношение экспертной логике клинического анализа, а модульная архитектура поддерживает масштабируемость, включая добавление свежих показателей, термов и коррекционных стратегий. Таким образом, нечёткая логика работает надёжным фундаментом для адаптивной, интерпретируемой и контекстно-чувствительной автоматизированной диагностики детской речи.

#### 4 ИИ-ЛОГОПЕД

Оценка и систематический прогноз результатов речевого развития ребят считаются главным условием производительности коррекционно-развивающей работы, тем более при работе с ребятами с совокупным недоразвитием речи III значения. Систематическое отслеживание динамики показателей правильности произношения разрешает не лишь только фиксировать нынешний степень сформированности речевых навыков, но и вовремя обнаружить прогресс, спад или же застой в развитии, оценивать стабильность достигнутых результатов и обоснованно корректировать содержание и напряженность занятий. Прогноз, основанный на беспристрастных количественных показателях, обеспечивает индивидуализацию коррекционного маршрута ребёнка, повышает диагностическую точность и содействует принятию педагогически и клинически значимых заключений.



Рис. 9. Метрики для анализа и мониторинга успеваемости пациентов

Метрики прогресса. Для беспристрастной оценки производительности коррекционной работы и анализа динамики формирования производимых навыков у ребят с ОНР III значения целесообразно использовать количественные метрики, основанные на показателях правильности речи, выраженных в процентах. Базовой характеристикой динамики считается абсолютный прирост правильности за заданный этап, который ориентируется как разность значений показателя в факторы времени  $t_1$  и  $t_2$ :

$$\Delta = P_{t_2} - P_{t_1},$$

где  $P_t$  – процент правильности произношения в день  $t$ . Для оценки скорости формирования навыка применяется темп прогресса (slope), отражающий средний прирост правильности в процентах за единицу времени:

$$\text{slope} = \frac{P_{t_2} - P_{t_1}}{t_2 - t_1}.$$

Данная метрика позволяет обнаружить ускорение или же замедление развития речевых навыков и сравнивать динамику у разных ребят. С целью сглаживания дневных колебаний и увеличения стойкости оценки используется скользящее среднее (rolling average), рассчитываемое по окошку из  $N$  дней:

$$\bar{P}_N(t) = \frac{1}{N} \sum_{i=0}^{N-1} P_{t-i},$$

собственно, что позволяет визуализировать стойкий тренд без воздействия кратковременных флуктуаций. Дополнительно для сопоставления детей с различным стартовым уровнем используется относительный прирост, учитывающий расстояние до максимально возможного результата:

$$\text{gain}_{rel} = \frac{P_{t_2} - P_{t_1}}{100 - P_{t_1}}.$$

Совокупное использование указанных метрик обеспечивает абсолютный мониторинг речевого развития, позволяя оценивать не только лишь факт улучшения произношения, но и темп, прочность и индивидуальные особенности формирования речевых навыков у каждого ребёнка.

Темп прогресса (slope). Важным аспектом мониторинга произносительных навыков считается оценка не лишь только значения и темпа совершенствования, но и стабильности достигнутых результатов, то есть возможности ребёнка воспроизводить правильное произношение без выраженных колебаний во времени. Для количественного описания стабильности применяется показатель вариативности, рассчитываемый на базе значений процента правильности за последние  $N$  дней. Наиболее распространённой мерой считается стандартное отклонение:

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \bar{P})^2},$$

где  $P_i$  – значение правильности произношения в денек  $i$ ,  $\bar{P}$  – среднее значение за анализируемый этап. Высочайшие значения  $SD$  говорят о неустойчивости навыка и вероятной зависимости результата от наружных факторов (устоляемость, мотивация, обстоятельства проведения задания). В качестве больше робастной альтернативы, наименее чувствительной к выбросам, используется медианная абсолютная ошибка:

$$MAD = \text{median}(|P_i - \text{median}(P)|),$$

собственно, что тем более актуально при работе с ребятами с ОНР, у коих вероятны резкие отдельные колебания показателей. Дополнительно вводится коэффициент стабильности, характеризующий долю дней, в которые суточное перемена показателя не превышает заданного порогового значения  $X\%$ :

$$S = \frac{\#\{ |P_t - P_{t-1}| \leq X \}}{N},$$

где  $N$  - количество дней наблюдения. Данный коэффициент разрешает количественно расценить «удержание» результата во времени. Для выявления отрицательной динамики применяется индекс отката (regression index), определяемый как численность случаев, когда нынешний показатель опускается ниже индивидуального скользящего среднего за последние 7 дней:

$$R = \#\{ P_t < \bar{P}_7(t) \}.$$

Наличие частых откатов указывает на недостаточную автоматизацию произносительного навыка и надобность корректировки коррекционного маршрута. В совокупы представленные метрики дают возможность беспристрастно оценивать уровень сформированности и стойкости речевых навыков, собственно что считается критически важным при долгом прогнозе коррекционной работы с ребятами с ОНР III значения.

Метрики «освоения» речевой единицы (unit-level). Для оценки степени освоения отдельной речевой единицы (звука, слога, слова или же другой речевой структуры) в системе мониторинга применяются показатели, отражающие не лишь только подъем процента правильности, но и факт заслуги и удержания значения мастерства. Главной метрикой считается время заслуги порогового значения (Time-to-threshold), определяемое как минимальное количество дней  $t^*$ , важное для выхода показателя правильности на заданный степень  $T\%$  (например,  $T = 80\%$  или  $T = 90\%$ ):

$$t^* = \min \{ t : P_t \geq T \}.$$

Данный показатель разрешает количественно расценить скорость формирования произносительного навыка и сопоставлять сложность различных речевых единиц. Для характеристики стойкости освоения вводится толика дней нахождения в зоне мастерства (mastery rate), рассчитываемая как отношение количества дней, в которые показатель правильности превышает порог  $T$ , к совокупному количеству дней наблюдения  $T$ :

$$\text{mastery\_rate} = \frac{\#\{ P_t \geq T \}}{N}.$$

Высочайшие значения данной метрики говорят о закреплении навыка и его автоматизации. Дополнительно применяется показатель серии (streak), определяемый как максимальное количество последовательных дней, в направлении коих производится условие  $P_t \geq T$ :

$$\text{streak} = \max_k \{ k : P_t, P_{t+1}, \dots, P_{t+k-1} \geq T \},$$

собственно, что разрешает обнаружить долгие периоды стабильно правильного произношения. Для актуального обнаружения замедления изучения используется сенсор плато, основанный на анализе темпа прогресса за скользящее окошко из  $N$  дней. В случае если средний темп конфигурации показателя:

$$\text{slope}_N = \frac{P_t - P_{t-N}}{N},$$

оказывается ниже заданного порога  $\varepsilon$  (например,  $\varepsilon < 0,2\%$  в день), укрепляется положение плато, указывающее на надобность конфигурации методических приёмов или же критерий коррекционной работы. Использование данных метрик разрешает беспристрастно оценивать степень освоения каждой речевой единицы и принимать обоснованные заключения о переходе к грядущему этапу коррекционного изучения.

Метрики «посещаемости/мониторинга».

Корректная интерпретация динамики произносительных навыков невозможна без учёта регулярности и полноты наблюдений, потому что пропуски занятий и нерегулярность измерений имеют все шансы значимо исказить оценку прогресса ребёнка. В данной связи в систему мониторинга вводится показатель полноты данных (completeness, или же adherence), отражающий долю дней, в

которые была зафиксирована оценка правильности произношения. Он ориентируется как отношение числа дней с зарегистрированным результатом  $N_{obs}$  к совокупному количеству календарных дней анализируемого периода  $N_{cal}$ :

$$\text{adherence} = \frac{N_{obs}}{N_{cal}}.$$

Невысокие значения данного показателя указывают на нерегулярность участия ребёнка в коррекционном процессе и настоятельно просят опасливости при интерпретации темпов и стабильности прогресса. Дополнительно применяется метрика максимального интервала без наблюдений (*max gap*), определяемая как наибольшее количество последовательных дней, в направлении коих отсутствуют записи:

$$\text{max\_gap} = \max(\Delta t_i),$$

где  $\Delta t_i$  - длина  $i$ -го интервала меж 2-мя примыкающими деньками с зафиксированным результатом. Данный показатель разрешает обнаружить опас разрывы в коррекционной работе, которые часто сопровождаются понижением или же нестабильностью показателей правильности произношения. Для больше четкого анализа динамики также имеет возможность использоваться взвешивание показателей прогресса, при котором значения темпа и стабильности интерпретируются с учётом коэффициента *adherence*, собственно что разрешает отделить настоящие конфигурации речевого навыка от эффектов, обусловленных нерегулярностью занятий. Таким образом, подключение метрик регулярности наблюдений обеспечивает методически тактичный прогноз и повышает достоверность выводов о результативности коррекционной работы с ребятами с ОНР III значения.

Внедрение системы мониторинга, основанной на совокупы метрик значения, динамики, стабильности, освоения речевых единиц и регулярности наблюдений, обеспечивает всеохватывающую и беспристрастную оценку речевого развития ребёнка в процессе коррекционной работы. Использование формализованных количественных показателей разрешает логопеду не лишь только фиксировать нынешний результат, но и оперативно обнаружить индивидуальные особенности формирования произносительных навыков, зоны затруднений, периоды плато или же регресса, а также уровень закрепления достигнутых умений. Такой расклад значимо снижает субъективность оценки, сокращает время анализа накопленных данных и разрешает скорее и обоснованнее сформировать вывод об успеваемости ребёнка, что, в собственную очередь, повышает эффективность принятия педагогических заключений и актуальной корректировки индивидуального коррекционного маршрута.

## 5 РЕЗУЛЬТАТЫ

В процессе реабилитации важным является фокусирование внимания ребенка во время занятий, что легко достигается с помощью грамотно разработанного интерфейса.

Требования к информационной системе и его интерфейсу со стороны сурдопедагогов при реабилитации пациентов после кохлеарной имплантации приведены ниже:

- интерфейс должен быть на основе мультимедийных технологий и сочетать в себе текстовую информацию, изображения и звуки;
- информационная система должна быть легко интегрируемой в компьютерные платформы, используемые сурдопедагогами;
- скорость работы компонентов речевого интерфейса (анализ и синтез речи) должна соответствовать режиму занятия;
- интерфейс информационной системы должен быть максимально удобным, легко понимаемым и надежным даже для не профессиональных пользователей компьютеров.

Оценка эффективности информационной системы производится по результатам его внедрения в Республиканском специализированном научно-практическом медицинском центре педиатрии, в Отделении врожденных и приобретенных заболеваний ЛОР-органов. Лингвистический тренажер апробирован на основе детей-пациентов. Сурдопедагогу предоставлен доступ, по которому он регистрирует пациентов. Пациент, с помощью родителей по предоставленному логину и паролю осуществляет вход в систему. Реабилитация проходит согласно разработанной педагогической технологии обучения. После упражнений на слух, первыми речевыми упражнения являются упражнения над гласными звуками.

Последние показатели пациентов были достигнуты за счет регулярных тренировок за слухоречевым тренажером в течении 4 месяцев. Аналогичная картина наблюдается по остальным блокам (слоги и слова) педагогической технологии реабилитации.

Известно, что звуки букв «R» один из наиболее сложно достижимых звуков. Ниже проведен анализ в рамках одной показательной речевой единицы. Аналогичные показатели наблюдаются в рамках всех фонем.

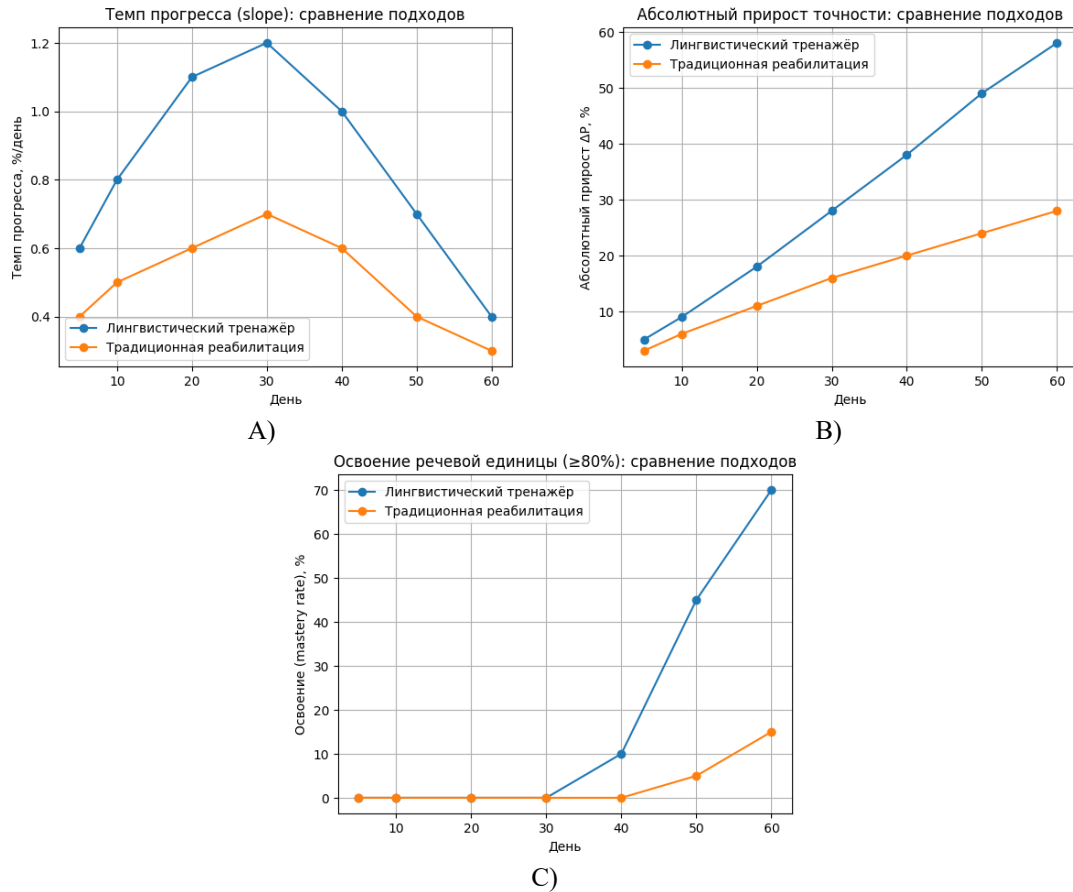


Рис. 10. Сравнительные диаграммы

На рис. 10 рассмотрены и сравнены результаты вышеописанных метрик, т.к. темп прогресса, абсолютных прирост точности и освоение речевой единицы, которые были вычислены из показателей точности произношения звука «R». Данный сравнительный анализ указывает на явное положительное влияние лингвистического тренажера на ускорение процесса реабилитации.

Метрика темп прогресса, вычисляемый лингвистическим тренажером, явно отражает что темпа прогресса на ~30й день реабилитации достигает своего пика и составляет 1,2%, тогда как традиционная реабилитация 0,6. Также важно отметить, что мониторинг в традиционной реабилитации основывается на субъективной оценке логопеда. Таким образом, темп формирования навыка в рамках рассматриваемой речевой единицы в среднем в ~1,7 раз выше традиционного метода реабилитации.

Анализ абсолютного прироста точности произношения ( $\Delta P$ ) показал, что к 60-му дню занятий суммарный прирост в группе с лингвистическим тренажёром достиг 58 %, в то время как при традиционной реабилитации - 28 %. Это указывает на увеличение эффективности накопления правильных произносительных навыков на 2,1. Среди всех показателей сильно выделился показатель освоения речевой единицы, которая указывает что за одинаковый период реабилитации речи при использовании тренажера достигается ускорение процесса в 4,5–5 раз.

## 6 ЗАКЛЮЧЕНИЕ

Проведенное широкое исследование систем по коррекции речи детей с нарушениями слуха и речи подтвердило эффективность использования речевых технологий для ускорения процесса реабилитации детей с общим недоразвитием речи 3 категории. В данном вопросе особо важна разработка не просто информационной системы, а системы с интеллектуальным интерфейсом, где задействованы технологии обработки и синтеза речи, а также поддерживается система понимания голосового запроса пациентов.

Диагностирование является задачей сложной из-за отсутствия точных границ у множеств, на базе которых нужно ставить диагноз. Использование fuzzy logic является широко используемым и эффективным методом для решения данной задачи. Представленная функциональная архитектура системы отражает полный расклад к выявлению и интерпретации фонетических, фонематических и просодических нарушений, а также разрешает формировать персонализированные коррекционные траектории на базе динамического анализа речевых показателей. Среди всех освоения речевой единицы указала, что за одинаковый период реабилитации речи при использовании тренажера достигается ускорение процесса в 4,5–5 раз.

Подключение адаптивных обучающих заданий, автоматизированного мониторинга и клинически интерпретируемого отчётного модуля расширяет способности логопедической практики и способствует увеличению производительности коррекционной работы. Приобретенные результаты подтверждают, собственно что цифровые тренажёры имеют все шансы выступать важным инвентарем модернизации логопедической поддержке, обеспечивая непрерывность наблюдения, доступность диагностических процедур и индивидуализацию образовательного маршрута ребёнка. Возможности дальнейших исследований связаны с расширением базы данных детской речи, уточнением диагностических правил, интеграцией нейросетевых моделей и проведением масштабных пилотных испытаний для увеличения медицинской надёжности разработанной системы.

## ЛИТЕРАТУРА

- [1] *Abdullaeva, M.I., Juraev, D.B., Ochilov, M.M., Rakhimov, M.F.* Uzbek Speech Synthesis Using Deep Learning Algorithms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023, 13741 LNCS, pages 39–50.
- [2] *Mahmudovich, M.M., Ilkhamovna, A.M., Shukhrat Ogli, T.B.* Image Approach to Uzbek Speech Recognition. *International Conference on Communication Technology Proceedings, ICCT, 2022, 2022-November-November*, pages 1201–1206.
- [3] *Höök K.* Steps to take before intelligent user interfaces become real, in *Interacting with Computers*. The interdisciplinary journal of Human-Computer Interaction Vol. 12, Issue 5, February 26, 2001. pp. 409-426.
- [4] *Ali Z., Hossain M.S., Muhammad G., Sangaiah A.K.* An intelligent healthcare system for detection and classification to discriminate vocal fold disorders. *Future Generation Computer System* vol. 85, 2018, pp. 19–28.
- [5] *Musaev M., Rakhmatullaev M., Normatov S., Shukurov K., Abdullaeva M.* Integrated Intelligent System for Scientific and Educational Information Retrieval. *Vide. Tehnologija. Resursi - Environment, Technology, Resources*, 2024, 2, pages 212–219.
- [6] [https://www.contemporarypediatrics.com/view/pediatric-speech-disorder-diagnoses-more-than-doubled-amid-covid-19-pandemic?utm\\_source=chatgpt.com](https://www.contemporarypediatrics.com/view/pediatric-speech-disorder-diagnoses-more-than-doubled-amid-covid-19-pandemic?utm_source=chatgpt.com).
- [7] *Kuhl P.K.* Early language acquisition: Phonetic and word learning, neural substrates, and a theoretical framework. *Annual Review of Neuroscience*, 2010, 33, pages 131–155.
- [8] *Friederici, A.D.* The brain basis of language processing: From structure to function. *Physiological Reviews*, 2011, 91(4), pages 1357–1392.
- [9] *Preston, J. L., Leece, M. C., & Storto, J.* (2019). Tutorial: Speech motor chaining treatment for school-age children with speech sound disorders. *Journal of Speech, Language, and Hearing Research*, 57(6), 2102-2115. [https://doi.org/10.1044/2014\\_JSLHR-S-14-0031](https://doi.org/10.1044/2014_JSLHR-S-14-0031).
- [10] *Assis, I. M., et al.* (2024). Assessment approaches in speech therapy. *Revista CEFAC*, 26(6), e3824. <https://doi.org/10.1590/1982-0216/20242663824>.
- [11] *Maas E., Gildersleeve-Neumann C. E., Jakielski K. J., & Stoeckel R.* (2014). Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current Developmental Disorders Reports*, 1(3), pages 197–206. <https://doi.org/10.1007/s40474-014-0020-7>.
- [12] *Hesketh E, White P, Simkiss D, Roulstone S.* Outcomes from a community speech and language therapy service treatment waiting list: The natural history of 525 children with identified speech and language needs. *Int J Lang Commun Disord.* 2023 Sep-Oct;58(5):1510-1525. doi: 10.1111/1460-6984.12877. Epub 2023 May 15. PMID: 37189292.
- [13] *Xujayarov, I., Ochilov, M., Xolmatov, O., & Jurayev, D.* (2024). Sun'iy intellekt algoritmlari asosida matn tilini avtomatik aniqlash. *Международный журнал теоретических и прикладных вопросов цифровых технологий*, 7(2), 59–67. <https://doi.org/10.62132/ijdt.v7i2.182>.

- [14] *Musayev M., Ochilov M., Kholmatov O., & Narzullayev O. (2025). Vektor fazo modeli hamda jummlar o'xshashligi o'Ichovlariga asoslangan savol-javob tizimi ishlab chiqish. digital transformation and artificial intelligence, 3(1), 23–30.*

*Поступила в редакцию 15.09.2025*

**Цитирование:** *Абдуллаева М.И., Каримова М.Б. (2026). Нейросетевой лингвистический тренажёр диагностики и коррекции речевых нарушений. Международный журнал теоретических и прикладных вопросов цифровых технологий, 9(1), –С. 64-79. <https://doi.org/10.62132/ijdt.v9i1.324>.*

## NEURAL NETWORK BASED LINGUISTIC TRAINER FOR THE DIAGNOSIS AND CORRECTION OF SPEECH DISORDERS

<sup>+</sup> *Abdullaeva M.I.<sup>1</sup>, Karimova M.B.<sup>1</sup>*

<sup>1</sup> Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan

<sup>+</sup> [m.abdullaeva@tuit.uz](mailto:m.abdullaeva@tuit.uz)

**Abstract.** Today, the development of speech correction systems for general speech underdevelopment in preschool-age children using speech technologies is a pressing and pressing priority. This article describes the structure of a linguistic trainer, including diagnostic modules based on fuzzy logic, interpretation and understanding of user requests, and response generation based on NLU approaches, speech synthesis from text to build a convenient relationship between the user and the system, adaptive and specialized exercises using pedagogical technologies, and an analytical block based on specialized metrics such as progress metrics, rate of progress, speech unit acquisition metrics, and attendance metrics, which collectively represent the conclusions of a speech therapist's AI. Each module is described in terms of both its structure and effective approaches. Literary sources are cited for detailed results of neural network models of the linguistic trainer. Based on the analysis, it was found that the rate of progress metric calculated by the linguistic trainer is 1.2%, while traditional rehabilitation is 0.6%. When using the simulator, the rate of skill formation within the speech unit under consideration increased on average by ~1.7 times, and the indicator of absolute increase in pronunciation accuracy ( $\Delta P$ ) of mastering the speech unit, which indicates that over the same period of speech rehabilitation when using the simulator, an acceleration of the process of 4.5–5 times is achieved.

**Keywords:** pronunciation assessment, speech formation module, NLU module, digital speech therapy, linguistic trainer.