

## ДОСТУПНОСТЬ ОБРАЗОВАНИЯ ДЛЯ ЛЮДЕЙ С ОГРАНИЧЕНИЯМИ СЛУХА

Митюхин А.И.

*Институт информационных технологий БГУИР, г. Минск, Республика Беларусь*

*[mityuhin@bsuir.by](mailto:mityuhin@bsuir.by)*

Рассмотрены проблемы устранения барьеров к ступеням непрерывного образования для людей с глухотой или нарушением слуха за счет использования инновационных технических методов доступа. Представлена современная информационная программно-аппаратная система отображения устной речи в письменный текст с точки зрения эффективного ее применения людьми с ограничениями слуха.

Ключевые слова: инновации; непрерывное образование; доступ; образовательные ресурсы.

Согласно Конвенции ООН обеспечение равной доступности непрерывного образования инвалидам рассматривается как одно из основных прав, которое должно соблюдаться в стране [1]. Гарантируется право людям с физическими ограничениями иметь возможность получения непрерывного образования, профессиональных знаний и улучшения своих навыков. Практическая реализация этого права связана в первую очередь с получением достаточного доступа к образовательным ресурсам или в безбарьерную учебную среду. *По данным Программы развития ООН глобальный уровень грамотности для взрослых людей с инвалидностью составляет всего три процента, а для женщин с инвалидностью – всего один процент* [2]. В работе рассматриваются проблемы доступности для лиц с особыми потребностями по слуху, выявление технических и иных препятствий с точки зрения инклюзии.

Съемные слуховые системы или имплантанты, закрепляемые в среднем или внутреннем ухе, пока не решают проблемы инвалидов по слуху по разным причинам. К главным причинам можно отнести сложность и стоимость хирургических операций даже в западных странах [3], [4]. В этом случае основным барьером в образовании для людей нарушением слуха является отсутствие субтитров речевых и других аудио сигналов. Прогресс в информационной науке и

цифровой технологии распознавания речевых сигналов позволяют иметь новые инструменты и сервисы, обеспечивающие определенную степень доступности к непрерывному образованию, учебным материалам и иным образовательным ресурсам. Эффективное практическое использование инвалидами по слуху этих инструментов требует выявления основных их технических особенностей. В качестве такого инструмента выступает современная информационная программно-аппаратная система отображения устной речи в письменный текст. Эта же система является инструментом обеспечения коммуникации для лиц с особыми потребностями по слуху.

Для лучшего понимания проблем эффективного и правильного применения инструмента «речь–текст» следует на системном уровне рассмотреть технические особенности аппаратно-программной его реализации, рис. 1.

Основой технологии преобразования речи в текст является автоматическое распознавание речи. Как видно из рис. 1, после обнаружения и захвата речевого сигнала через микрофон, предварительной обработки (фильтрации), цифро-аналогового преобразования, цифровой спектральной обработки и формирования опорных векторов признаков речи (образов распознавания) решается задача распознавания речи. Надежная система декодирования текста должна содержать блок дополнительной точностной обработки (см. рис. 1) для улучшения читаемости текста за счет использования заглавных букв, расстановки знаков препинания и исправления грамматических ошибок, вносить контекстные изменения по мере необходимости и редактирование. Таким образом, система интегрирует речевые и звуковых сигналы, синтаксис, грамматику для надежного и точного формирования текста.

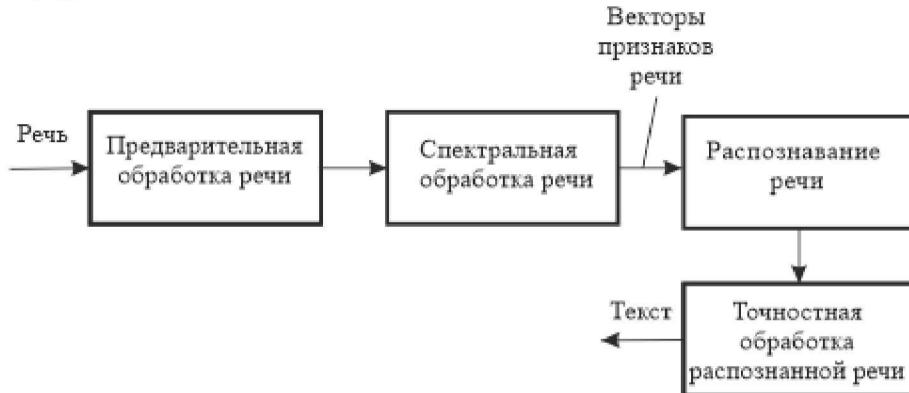


Рис.1. Обобщенная структурная схема системы речь-текст

Непосредственно задача распознавания речевого сигнала считается одной из самых сложных областей информатики, включающей в себя математику, цифровую обработку сигналов и изображений, кодирование, теорию вероятностей, лингвистику, статистику и пр. Решение этой задачи основывается на выполнении вычислительно затратных многоэтапных алгоритмов цифровой обработки сигналов, кодирования и методов машинного обучения с целью распознавания речевых шаблонов. На рис. 2 показана укрупненная схема реализации процесса распознавания.

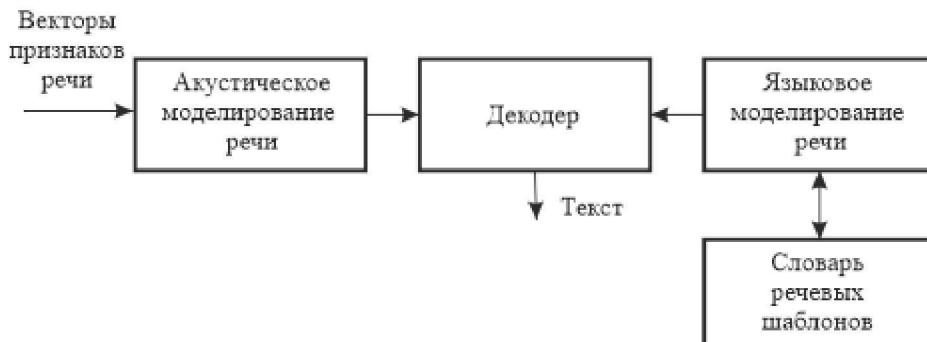


Рис.2. Обобщенная структурная схема блока распознавания речевого образа

Схема, рис. 2 преобразует речевой сигнал, представленный в виде векторов признаков последовательности фонем в слова, которые сравниваются с опорными образами словаря и декодируются в текстовом формате. Основными элементами схемы, рис.2, являются:

- акустическая модель, выполняющая распознавание звуков с помощью вектора признаков образа речи;
- языковая модель, выполняет предсказание вероятной последовательности слов фразы в контекстном значении, выбирая слова из словаря речевых шаблонов.
- в словаре реализуется процесс поиска соответствия звуковых сигналов речи в виде вектора признаков образа речи и слов словаря с возможностью определения отличий слов с одинаковым звучанием, используя контекст фразы. Тем самым, выбираются наиболее вероятные варианты слов и повышается точность распознавания речи.
- декодер на основе векторных оптимальных алгоритмов сравнений выходов моделей и словаря формирует текстовый образ входной речи, рис. 1.

В рамках НИР проводились экспериментальные исследования, связанные с разработкой рекомендаций по использованию систем речь–текст, структуры, которых показаны на рис. 1 и рис. 2 для людей с проблемами слуха. При выборе сравнения систем рассматривались системы распознавания речи, построенные на применении математических методов скрытых марковских моделей [5] и нейросетей [6]. Исследования касались режимов работы систем распознавания типа «Whisper Notes» [7]. Система может работать полностью *оффлайн под управлением операционных систем Windows, iPhone, Mac* без предварительного обучения языковой модели, рис 2. В этом случае не требовалось включения сравнительно затратного оборудования и времени на подготовку учебных текстов и технически не простое предварительное обучения языковой модели по заранее подготовленным аудио текстам лекций. Аудио файлы по математически насыщенным специальным дисциплинам («Теория информации» и «Цифровая обработка сигналов») должны были удовлетворять заданным техническим требованиям и характеристикам. Например, в форматах 16-битного PCM WAV или FLAC (без компрессии), частоте дискретизации 16 кГц или 44,1 кГц. Кроме того, исследования проводились для варианта работы систем распознавания речи в автономном режиме (не зависящем от сети). С целью ускорения обработки речевого сигнала для распознавания ограничивалось пространство поиска речевого шаблона, соответствующего текстовому слову, т. е. не использовался поиск в словаре сети.

Далее представляются анализ и основные практические результаты исследований по оценке точности и эффективности преобразования речи в текст на основе комплексного подхода решения проблемы доступности.

Для повышения точности распознавания необходимо учитывать следующие аспекты, непосредственно связанные с структурами систем, показанными на рис. 1 и рис. 2.

1. Необходимо упрощать синтаксис учебного материала.
2. Для варианта распознавания без предварительного обучения языковой модели желательно заранее сформировать достаточный объем словаря. Необходимо заранее подготовить список слов, научных текстов с их фонетическими транскрипциями и интегрировать его в систему распознавания речи. Объем словаря в значительной мере определяет точность распознавания речи, обеспечивая точность интерпретации звуковых сигналов со словами и словосочетаниями (лексическими единицами языка). Очевидно, **ограниченный словарь** системы не распознает не включённые слова. Кроме того, в этом случае системе распознавания сложно отличать похожие по звучанию слова и выбирать наиболее вероятные варианты из словаря. В общем случае, точность распознавания определяется не только объемом словаря, но и объемом обучающей выборки, которая включает в себя:

- количество часов аудио;
- разнообразие тематик речи;

– качество транскрипций.

Экспериментальные исследования показали, что точность распознавания постепенно увеличивалась по мере временного речевого продвижения по двум ранее названным лекционным курсам. Однако качество транскрипции оставляет желать лучшего. Покажем это на примере анализа ошибок.

**Пример. Эталонная фраза:** «Расширенный код Хэмминга исправляет ошибки кратностью один и обнаруживает две или более ошибок».

Экспериментально полученное **текстовое содержание**: «Расширенный кот Хеминга исправляет ошибки кратностью один и обнаруживает две ошибок».

*Орфографические ошибки:*

Расширенный → Расширеный;

Код → кот;

Хэмминга → Хеминга;

*Пропуски:*

или более;

*Грамматические ошибки:*

две и более ошибок → две ошибок

Оценим качество распознавания транскрипции, вычислив коэффициент ошибок *WER* (Word Error Rate) на уровне слов [5].

$$WER = \frac{S+D+I}{N} \cong 0,38, \text{ или } 38\%$$

где  $S = 3$  – число замен;

$D = 2$  – число пропусков

$I = 0$  – число вставок;

$N = 13$  – число слов в эталонной транскрипции.

Правильная работа системы распознавания на этапе языкового моделирования должна характеризоваться распознаванием не только отдельных слов, но и наиболее вероятной последовательностью слов с учетом контекстуального смысла фразы. Например, языковая модель должна распознать звук «Код Хэмминга», но не «Кот Хэмминга». Качество транскрипции в примере можно оценить, как **неудовлетворительное**. Ошибки затронули научные и технические термины, и смысл утверждения «две или более ошибок», что не допустимо для образовательного или научного текста. С увеличением объема речи текст постепенно улучшается.

3. Следует разделять длинные предложения на более короткие.

4. Делать паузы перед формулами. Объясняя или комментируя формулы следует каждую ее составляющую проговаривать отдельно, четко и сравнительно медленно, иначе система преобразует речь в бессмысленный текст. Большое количество формул, специальных математических символов, типа «символ Лежандра», определений, типа «матрица Джекобстола» резко уменьшает корректность и надежность распознавания. Научные термины часто искажаются, отображаются с ошибками. Пока еще автоматическое распознавание речи не справляется с формулами, символами и сложными терминами.

5. Необходимо заранее подготовить визуальные средства (**глоссарий научных терминов** с пояснениями и визуальными примерами, схемы, формулы и пр.). В начале обучения языковой модели система распознавания не понимает «новые» термины, специальные выражения. Процесс понимания с учетом контекстуального смысла фразы улучшается с увеличением времени обучения языковой модели.

6. Важно учитывать технические характеристики микрофона. Точность распознавания улучшилось при переходе на однородный микрофон с суперкардиоидной диаграммой направленности «beyerdynamic M 201» (DE). Отмечалось уменьшение влияние фоновых акустических шумов.

7. Для повышения точности распознавания следует учитывать особенностях голоса конкретного человека, тембр, темп речи и пр.

8. Для повышения точности распознавания желательно использовать модели Whisper

Notes Large-v2 или Whisper Notes Large-v3. Но в этом случае требуется компьютер с высокими техническими характеристиками и мощным GPU (Graphics Processing Unit), например, NVIDIA. Для Whisper Notes Large-v2 получены значения усредненной точности в метрике WER. Точность распознавания речи на математическом контенте находилась в диапазоне

$$1 - \text{WER} = 65 \div 80\%.$$

Значительное число ошибок возникает из-за неправильного распознавания специальных или аббревиатурных терминов, например, «Собственная функция», «БПФ». Исследования показывают, что существуют явные проблемы при передаче математической научной информации с использованием систем автоматического распознавания речи. Определенной проблемой является то, что системы распознавания речи с повышенными характеристиками точности относятся к платным или подписным вариантам. Оплата сервиса IBM Watson Speech-to-Text составляет от 0.02 \$ за минуту.

Благодаря последним достижениям в области распознавания речи следует ожидать появление более надежных инструментов, обеспечивающих более широкий доступ к непрерывному образованию людей с ограничениями слуха.

#### Литература

1. <https://www.un.org/en/academic-impact> [Электронный ресурс].
2. <https://www.un.org/en/disability-higher-education-workforce-preparedness-students-disabilities> | [Электронный ресурс].
3. Митюхин, А.И. Улучшение коммуникационных возможностей доступа к образованию для слабослышащих людей / А.И. Митюхин / Непрерывное профессиональное образование лиц с особыми потребностями: сб. ст. IV Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 9–10 декабря 2021 года) / редкол.: А. А. Охрименко [и др.]. – Минск: БГУИР, 2021. С. 202–205.
4. Митюхин, А.И. / Коммуникация слабослышащих людей посредством системы преобразования речь-текст / А.И. Митюхин / Непрерывное профессиональное образование лиц с особыми потребностями: сб. ст. V Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 14 декабря 2023 года) /– Минск: БГУИР, 2023. – С. 200 – 203.
5. Huang, X. An Overview of Modern Speech Recognition / X. Huang, L. Deng / Handbook of Natural Language Processing. – Chapman & Hall/CRC, 2010.
6. Dong, Yu. Automatic Speech Recognition. A Deep Learning Approach / Yu Dong, Li Deng. – Springer-Verlag, London, Microsoft Research Redmond, WA, USA, 2015.
7. <https://www.whispertranscribe.com> [Электронный ресурс]

## ACCESSIBILITY OF EDUCATION FOR PEOPLE WITH HEARING IMPAIRMENTS

Mitsiukhin A.I.

*Institute of Information Technologies, Minsk, Republic of Belarus*

The article examines the problems of eliminating barriers to continuing education for individuals with hearing impairments and deafness through the use of innovative technical access methods. A modern information software and hardware system for displaying oral speech into written text in real time is presented from the point of view of its effective use by people with hearing impairments.

Keywords: innovation; continuous education; access; educational resources.