

КОММУНИКАЦИЯ СЛАБОСЛЫШАЮЩИХ ЛЮДЕЙ ПОСРЕДСТВОМ СИСТЕМЫ ПРЕОБРАЗОВАНИЯ РЕЧЬ-ТЕКСТ

Митюхин А.И.

Институт информационных технологий БГУИР, г. Минск, Республика Беларусь
mitvuhin@bsuir.by

В статье рассматриваются подход, основой которого являются радиоэлектронные коммуникационные средства последнего поколения 4G, 5G. Технические возможности коммуникации спецификации 4G, 5G позволяют повысить эффективность доступа людей с проблемами слуха к непрерывному образованию, профессиональной подготовке, упростить социальную адаптацию в современное общество.

Ключевые слова: образование; коммуникация; достоверность; ограниченные возможности по слуху; распознавание речи.

В мировом сообществе постепенно осознается, что интеграция людей с особыми потребностями в жизнь общества возможна только путем непрерывного образования и профессиональной подготовки на базе концепции безбарьерного доступа к образованию и профессии [1]. Это понимание согласуется со статьей 24 Конвенции ООН о правах инвалидов (2008) относительно общего образования: «Никто не должен быть исключен из системы общего образования» [2]. В странах с 4.0 индустрией [3] и современными базовыми принципами социальной и образовательной политики происходит ускорение тенденции интеграции людей с нарушениями слуха в школы и другие образовательные институты, включая колледжи и университеты [4]. Исследования в рамках ООН на примере общеобразовательных школ показывают, что в интегрированной среде результаты обучения лучше, чем у сопоставимых учеников в специальных школах [2]. В связи с наличием в Беларуси кодекса социального обеспечения инвалидов и принятием Конвенции ООН о правах лиц с инвалидностью потребность в усовершенствовании процесса обучения возрастает с каждым годом. Естественным продолжением процесса инклюзии людей с проблемами слуха в общество является включение их в систему высшего образования. На сегодняшний день существует большая потребность в усовершенствовании системы образования для людей с ограниченными возможностями, особенно в высших учебных заведениях, в том числе технического профиля. В реальных условиях ограничения коммутативных возможностей у студентов с ослабленным слухом особенно трудно идет изучение специальных технических дисциплин. Возникают большие сложности при передаче, записи, понимании технической, научной терминологии и пр. Основная проблема

здесь – правильный и однозначный перевод и понимание смысла специальной терминологии изучаемой дисциплины.

В многочисленной научно-технической, медицинской и педагогической литературе описаны четыре способа коммуникации с людьми с нарушениями слуха: вербальный, невербальный, посредством жестового языка, письменный. Невербальная коммуникация охватывает все видимые физические сигналы. Однако чтение по губам не является эффективным методом. Исследования показывают, что чтение по губам имеет ограниченную эффективность. Даже самый подготовленный читатель по губам может понимать только 20 процентов слов.

Использование цифровых слуховых аппаратов, медицинской технологии кохлеарного имплантирования улучшает процесс коммуникации людей с частичной или полной потерей слуха. Однако имплантирование слухового аппарата возможно только для тех пациентов, у которых слуховые сенсорные клетки во внутреннем ухе больше не функционируют, а слуховой нерв не поврежден. К сожалению, стоимость проведения операции внедрения кохлеарного импланта даже в индустриально развитых странах высока. Например, в Германии стоимость лечения, включая реабилитационные мероприятия, составляет 30000–40000 евро (покрывается государственными медицинскими страховыми компаниями).

Многие люди с нарушениями слуха предпочитают письменную форму коммуникации в сочетании с другими формами общения. Письменная форма коммуникации основана на преобразовании звука речи в непрерывный текст. Причем текст может сопровождаться дополнениями в виде изображений, диаграмм, графиков и др. В статье рассматриваются современный метод получения слушателем с недостаточным уровнем слуха передаваемой речевой информации в текстовом виде с использованием средств связи последнего поколения G. Представлены некоторые эмпирические идеи на названную тему с учетом более чем десятилетнего опыта коммуникации со студентами с частичной потерей слуха, преподавания специальных дисциплин в интегрированной группе студентов ИИТ БГУИР.

Как известно, в последние 10–13 лет в области цифровой обработки и распознавания речевых сигналов [5] были получены сравнительно хорошие результаты по точностным характеристикам как распознавания, так и разборчивости речи. На основе использования новых математических алгоритмов цифровой обработки сигналов и высокопроизводительных цифровых процессоров, радиоэлектронных аппаратно-программных средств в период введения известных ограничительных мер (COVID-19), удалось без значительного ущерба для качества обучения осуществлять образовательный процесс дистанционно [6]. Заметим, на решение этой сложной научно-технической задачи реально ушло более 60-и лет. Появилась реальная техническая возможность получить функциональную компоненту вида «речь-текст» и использовать ее при работе с людьми с особой потребностью по слуху. Можно утверждать, что преобразование речи в письменный образ – это сильный стимул для разработки новых, более эффективных подходов для преподавания и обучения инвалидов по слуху.

На занятиях легко обнаруживалось, что студенты с нарушениями слуха испытывают трудности в освоении специальных дисциплин, курсов, где присутствуют новые научные термины, определения, понятия и пр. Коммуникации с преподавателями во многих случаях становились малоэффективными или недостаточными для усвоения таких дисциплин, несмотря на использование современных слуховых аппаратов. Очевидным становилось, что на общение с такими студентами следует запланировать достаточное количество времени.

На кафедре ФМД ИИТ БГУИР были выполнены научные и экспериментальные исследования [7], связанные с получением количественных оценок достоверности (точности) распознавания речи. Для преобразования речи в текст и повышения точности транскрипции используются различные алгоритмы обработки сигналов и вычислительные методы. Экспериментальные исследования выполнялись для систем обработки речевых сигналов, которые обеспечивают минимально возможную ошибку распознавания. В этом случае используется математический алгоритм, основой которого является операция свертки. На базе этой операции строятся оптимальные сверточные нейросети, обеспечивающие

минимально возможную ошибку преобразования «речь-текст» с наилучшей транскрипцией текста. Рассматривались решения обработки цифровых речевых сигналов, полученные в научных и инженерных подразделениях компаний Microsoft Research и Google с использованием сверточных рекурсивных нейросетей. Была разработана методика оценки точности распознавания, выполнены сравнительные эксперименты с использованием аппаратного оборудования, показанного на рис. 1. Оценка точности характеризовалась частотой ошибок в словах для заданной скорости обработки. В качестве сохраненных эталонных шаблонных элементов (базового словаря) использовались простые слова, слова научно-технического характера и текст из учебных материалов для специальных дисциплин (теория информации, теория помехоустойчивого кодирования, теория цифровой обработки сигналов и изображений, отрывки из записанных лекции и др.). Речь преподавателя произносилась сравнительно медленно с четким обозначением пауз между словами, предложениями и т. д.



Рисунок 1 – Структурная схема экспериментальной установки

В качестве основных аппаратных компонентов установки применялись:

1. USB-Микрофон – Beyerdynamic Fox (DE).
2. Проводной канал – кабель SCP USBC–5M (U.S.).
3. Приемное устройство – Razer Blade Stealth 13 GTX 1650 RZ 09-0310 (U.S.).
4. Видеоадаптер – Satechi USB-C Hybrid Multi Port Adapter ST-UCHSEM (U.S.).
5. Проводной видеоканал – Wyrestorm EXP-HDMI-H2-5M (U.S.).
6. Дисплей – Sony KD-50X81J.

Эксперименты показали, что основное влияние на частоту ошибок в словах и, соответственно, на автоматическое эффективное распознавание речи и правильное преобразование вербального формата в письменный (текстовый) формат оказывают такие факторы:

1. Выбор сертифицированного оборудования (рис. 1) с анализом технических спецификаций и производства. В частности, следует использовать профессиональные однонаправленные usb-микрофоны с кардиоидной диаграммой направленности.
2. Акустические и шумовые характеристики учебной аудитории.
3. Речь говорящего, в которой произношение носит специфически заметные особенности: громкость, тембр, значительный вес высокочастотных спектральных составляющих.
4. Нестационарный характер речевого сигнала на временном интервале произношения слов, предложений, когда происходит резкое изменение дисперсии спектральных компонент речевого фрагмента.
5. Специфическая лексика специальных технических дисциплин.
6. С увеличением времени «речевой коммуникации» по сравнительно одной теме, например, тема лекции «Энтропия блокового источника», в которой имеются повторения одних и тех же слов, фраз, точность преобразования «речь-текст» становится выше в тех экспериментах, где использовались рекурсивные нейросети. Происходит постепенный этап

оптимального настраивания процесса распознавания речи с оптимизацией транскрипции получаемого текста за счет использования в таких системах математических алгоритмов предсказания какое слово следует ожидать следующим.

7. Физическое состояние человека.

Результаты исследований приводят к следующим выводам:

– наибольшее влияние на точностные характеристики системы преобразования «речь-текст» оказывают факторы 1–7, приведенные выше;

– точность распознавания речевых образов и преобразования их в текстовый формат для обычных слов достаточно велика и составила 91–93%. Следует отметить, что за последние пять лет верность классификации речевых образов возросла с 75% до подтвержденных значений, полученных экспериментально на установке, рис. 1. Сравнительно высокие значения точностных параметров позволяют рассчитывать на использование техники преобразования «речь-текст» как при общем образовании людей с ограничениями по слуху, так и для модернизации образовательного процесса и повышения качества учебного процесса для различных типов учреждений образования;

– точность распознавания для более сложных слов особенно терминов технического характера в экспериментах находилась в пределах 86–91%.

В процессе исследовательской работы возникали и другие вопросы, которые требуют дальнейшего теоретического и экспериментального изучения с целью решения проблемы прямой коммуникационной потребности доступа к образованию людей с ограничениями по слуху.

Литература

1. Митюхин, А.И. Высшее образование для людей с ослабленным слухом в ИИТ БГУИР / А.И. Митюхин. // Непрерывное профессиональное образование лиц с особыми потребностями: сб. ст. II Междунар. науч.-практ. конф. (Респ. Беларусь, Минск, 14–15 дек. 2017 года. – Минск : БГУИР, 2017. – С. 88–89.

2. Disability [Электронный ресурс]. – Режим доступа: <https://www.un.org/development/desa/disabilities>. – Дата доступа: 11.11.2023.

3. Dokument nicht gefunden [Электронный ресурс]. – Режим доступа: https://www.bmbf.de/bundesministerium_für_bildung_und_forschung. – Дата доступа: 11.11.2023.

4. Митюхин, А.И. Коммуникационные составляющие инклюзивного обучения в университете / А.И. Митюхин. // Непрерывное профессиональное образование лиц с особыми потребностями: сб. ст. III Междунар. науч.-практ. конф. (Респ. Беларусь, Минск, 5 дек. 2019 года). – Минск: БГУИР, 2019. – С. 87–89.

5. Митюхин, А.И. Цифровая обработка речи и анализ изображений : учеб.-метод. пособие / А.И. Митюхин. – Минск : БГУИР, 2016. – 72 с

6. 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. (Респ. Беларусь, Минск, 9–10 дек. 2021 года) / редкол. : А. А. Охрименко [и др.]. – Минск : БГУИР, 2021. – С. 202–205.

7. Цык, А.В. Анализ характеристик распознавания речи компьютерных приложений : автореф. дисс. магистра технических наук : 1-45 80 01 / А.В. Цык ; науч. рук. А.И. Митюхин. – Минск : БГУИР, 2022. – 7 с.

COMMUNICATION OF HEARING-IMPAIRED PEOPLE THROUGH A SPEECH-TO-TEXT CONVERSION SYSTEM

Mitsiukhin A.

Institute of information technologies BSUIR, Minsk, Republic of Belarus

The article discusses the approach, which is based on the latest generation of radio-electronic communication means 4G, 5G. The technical capabilities of communication of the 4G, 5 G specification make it possible to increase the efficiency of access of people with hearing impairments to continuing education, professional training, and simplify social adaptation to modern society.

Keywords: Education; Communication; Credibility; Hearing Disability; Speech Recognition.