УДК 621.3.049.77–048.24:537.2

ПРОГРАММНОЕ СРЕДСТВО ПЕРЕВОДА ВИДЕО И АУДИО ЗВОНКОВ НА ЛЮБОЙ ЯЗЫК В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ

Симерова Е.И.

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Республика Беларусь

Научный руководитель: Станкевич А.Д. – ассистент кафедры ПИКС

Аннотация. В статье рассматривается программное средство для перевода аудио и видео звонков в режиме реального времени с использованием различных алгоритмов. Рассматриваются ключевые этапы распознавания речи, анализа и перевода текста, синтеза речи, основные аспекты синхронизации перевода и видеодорожки в режиме реального времени.

Ключевые слова: аудио и видео звонок, перевод, режим реального времени

Введение. С развитием технологий и глобализации важность перевода аудио и видео звонков на различные языки в реальном времени значительно возросла. Этот процесс включает в себя сложные алгоритмы и программные средства (ПС), обеспечивающие не только точный перевод, но и синхронизацию аудио и видео потоков. В данной статье рассматриваются основные этапы процесса синтеза речи, анализа и перевода текста, а также методы синхронизации перевода и видео дорожки.

Основная часть. Перевод аудио и видео звонков на любой язык в режиме реального времени представляет собой сложный и многогранный процесс, обусловленный необходимостью обеспечения минимальной задержки в обработке данных для обеспечения непрерывного и высококачественного воспроизведения звуков и изображений. Задержки могут привести к десинхронизации, эхо и другим проблемам, негативно сказывающимся на качестве звонков [1].

Для сокращения времени задержек могут применяться различные методы:

– использование высокопроизводительных компьютеров, способных эффективно обрабатывать аудио и видео данные в реальном времени;

– обеспечение скоростного интернет-соединения с высокой пропускной способностью для быстрой передачи данных между участниками звонка;

– применение специализированного оборудования, такого как наушники и микрофоны, способных обеспечить высокую четкость звука и изображения;

– использование технологий подавления шума для оптимизации ПС путем уменьшения внешних помех и шумов, что способствует повышению качества звонков.

Подавление шума – это процесс сокращения или удаления нежелательных акустических сигналов, таких как фоновый шум, эхо и другие, из аудио- или видеосигнала. Этот процес способствует повышению качества и ясности речи, а также снижает нагрузку на процессор и интернет-соединение [2]. Существует несколько методов подавления шума:

1 Спектральное затворение основано на анализе спектра сигнала и выявляет частоты, на которых присутствует шум, подавляя их и сохраняя при этом частоты, содержащие речь.

2 Глубокое обучение использует нейронные сети для обучения модели подавления шума на большом объеме данных. Благодаря этому модель способна адаптироваться к различным типам шума и обеспечивать высокое качество подавления шума в режиме реального времени.

3 Улучшение микрофона включает в себя использование специализированных устройств для улучшения работы микрофона. Примеры таких улучшений включают в себя подавление эхо, автоматическую регулировку громкости, фильтрацию низких частот.

Процесс перевода звонков в режиме реального времени опирается на использование алгоритмов автоматического распознавания речи [3]. Акустические модели применяются для преобразования аудио сигнала в текст, который затем может быть переведен на другой язык. Синтез речи, в свою очередь, позволяет конвертировать текст на другом языке обратно в аудио сигнал, что позволяет собеседнику услышать его.

Распознавание речи представляет собой процесс преобразования речевого сигнала в цифровую информацию. Для автоматического распознавания речи используются акустические модели, принимающие на вход признаки небольшого участка аудио сигнала, называемого фреймом, и выдают распределение вероятностей различных фонем – элементарных звуков, позволяющий отличить одно слово от другого – на этом фрейме [4].

Процесс преобразования аудио в текст можно разделить на несколько этапов:

1 Анализ сигнала. Система получает голосовой сигнал, записывает его и отправляет на сервер. Затем сервер производит очистку сигнала от шумов и помех, а также разделяет запись на фонемы – звуковые единицы в языке. Каждый фонемный фрагмент проходит через акустическую модель, которая определяет, какие звуки были произнесены.

2 Расшифровка аудио. Речевые фрагменты записи сравниваются с эталонными произношениями слогов из акустической модели, используя методы машинного обучения для подбора фонетических вариантов произнесенных слов и определения их контекста.

3 Преобразование речи в текст. С помощью языковых моделей алгоритм определяет порядок слов и подбирает нераспознанные слова на основе контекста для преобразования речи в текст.

Процесс перевода текста можно разделить на несколько ключевых этапов:

1 Анализ и разбор текста. Алгоритм анализирует исходный текст, разбирая его на отдельные слова, фразы и предложения.

2 Перевод слов и фраз. Алгоритм находит соответствующие переводы для каждого слова или фразы в исходном тексте.

3 Составление переведенного текста. Алгоритм составляет переведенный текст, используя найденные переводы слов и фраз.

Существует несколько видов алгоритмов перевода текста:

1 Статический машинный перевод основан на анализе больших объемов параллельных текстов на разных языках. Подход хорошо работает для простых предложений, но имеет ограничения при переводе идиоматических выражений.

2 Нейронные сети используются для создания моделей, способных улавливать сложные зависимости между словами и фразами. Этот подход может обеспечить более точные и естественные переводы, особенно для контекстно зависимых выражений.

3 Гибридные модели комбинируют различные подходы для достижения более точных результатов. Они могут использовать как статический машинный перевод, так и нейронные сети, чтобы сбалансировать преимущества обоих подходов.

Важно отметить, что алгоритмы перевода текста не всегда обеспечивают высокую точность и естественность перевода. Они могут сталкиваться с проблемами, такими как полисемия (многозначность слов), идиоматические выражения и другие сложности языка.

Алгоритмы синтеза речи используются для генерации компьютерной речи на основе текста [5]. Процесс синтеза речи может быть разделен на следующие этапы:

1 Анализ текста. Программа анализирует входной текст, разбивая его на отдельные фонемы, слова и фразы.

2 Генерация звуков. На основе фонем и других звуковых элементов программа генерирует голосовые сигналы.

3 Модуляция речи. Программа настраивает скорость, высоту тона, громкость и другие параметры речи для создания естественного звучания.

Процесс перевода видео звонка включает следующие этапы, которые немного отличаются от перевода аудио:

– получение и обработка аудио потока;

– создание субтитров и/или новой аудиодорожки на основе перевода;

– синхронизация перевода и видео ряда;

– вывод видео для получателя звонка в виде субтитров и/или новой аудиодорожки.

Для синхронизации перевода и видеодорожки в режиме реального времени необходимо точно определить моменты начала и окончания речи в аудио потоке и соответствующие моменты в видео потоке. Это означает, что ПС, используемое для перевода видео звонка, должно быть способно определять эти моменты в режиме реального времени. При этом следует учесть возможные задержки как при самом переводе текста, так и при передаче данных между участниками видео звонка.

После точной синхронизации перевода и видеоряда, вывод видео для получателя звонка может представляться в виде субтитров, которые будут соответствовать произносимому голосу, или новой аудиодорожки, где голос будет приведен в желаемый переведенный язык. В обоих случаях, задача ПС для перевода видео звонка состоит в том, чтобы обеспечить точное соответствие перевода и видео, чтобы получатель звонка мог легко понимать и следить за происходящим в видео звонке.

Заключение. Разработанное ПС представляет собой инновационное решение, способствующее преодолению языковых барьеров и обеспечивающее эффективную коммуникацию на международном уровне. Указанные аспекты включают в себя сокращение времени задержек, подавление шума, алгоритмы автоматического распознавания речи, процессы преобразования аудио и текста в речь, а также алгоритмы перевода текста и синтеза речи. Эти компоненты являются фундаментальными для обеспечения высокого качества и точности перевода в режиме реального времени. В дальнейшем исследовании необходимо уделить внимание улучшению алгоритмов перевода и синтеза речи, а также оптимизации синхронизации перевода с видео потоком. Кроме того, учёт многообразия языков и культурных особенностей пользователей также представляет существенное значение для дальнейшего совершенствования ПС в этой области.

Список литературы

1. Кулаков, В. В. Принципы работы программного средства для перевода аудио- и видеозаписей. Журнал инженерных наук. – 2018. – № 6. – С. 87-95

2. Методы улучшения речи и шумоподавления [Электронный ресурс]. – Режим доступа: https://habr.com/ru/companies/ru_mts/articles/584308/. – Дата доступа: 13.10.2023.

3. Цветков, А. В. Применение методов машинного обучения для повышения точности программного средства перевода аудио и видео звонков. Цифровые технологии и нелинейная динамика. – 2019. – Т. 11. – № 2. – С. 77-86.

4. Чернявский, С. С. Сравнительный анализ систем распознавания речи для перевода аудио в режиме реального времени // Международный научно-исследовательский журнал. – 2018. – № 6. – С. 12-21.

5. Климов, А. В. Методы и алгоритмы распознавания и синтеза речи в программах перевода. Компьютерные исследования и моделирование. – 2020. – № 5. – С. 78-88.

UDC 621.3.049.77–048.24:537.2

SOFTWARE TOOL TO TRANSLATE VIDEO AND AUDIO CALLS INTO ANY LANGUAGE IN REAL TIME

Simerova E.I.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Stankevich A.D. – assistant of the department of ICSD

Annotation. The article discusses a software for real-time translation of audio and video calls using various algorithms. The main stages of speech recognition, text analysis and translation, speech synthesis, as well as the main aspects of synchronisation of video tracks in real time are considered.

Keywords: audio and video call, translation, real-time modeУДК 621.376:621.396