УДК 004.89
DOI: 10.37661/1816-0301-2025-22-3-25-34

Оригинальная статья
Original Article

Система транскрибации речи и перевода с русского языка на китайский

Л. П. Кузьменков, В. А. Чуйко, Е. И. Козлова✉

Белорусский государственный университет,
пр. Независимости, 4, Минск, 220030, Беларусь
✉E-mail: kozlova@bsu.by

Аннотация

Цели. Целью проведенной работы является разработка архитектуры информационной системы для транскрибации и перевода речи, реализация ее блоков и тестирование их работы.

Методы. Рассмотрены существующие способы распознавания речи, проведен сравнительный анализ моделей распознавания речи и перевода текста. Процесс транскрибации речи включает в себя несколько последовательных этапов: сбор и предварительную обработку аудиосигнала, извлечение акустических признаков, непосредственное распознавание речи, постобработку и коррекцию текста, вывод результата. На этапе предобработки аудиосигнала используется комбинация специализированных библиотек, обеспечивающих подготовку данных для последующего анализа. Для нормализации параметров записи применяется библиотека librosa, позволяющая выполнять передискретизацию сигнала до стандартной частоты 16 кГц и преобразование его в монофонический формат. Для подавления фоновых шумов и выделения речевого компонента задействуется нейросетевая модель Demucs. Алгоритм спектральной субтракции дополнительно корректирует остаточные шумы. Сегментация речевой активности выполняется с использованием энергетического детектора из WebRTC, автоматически выделяя речевые фрагменты и удаляя паузы. Для реализации системы распознавания речи выбрана модель whisper-turbo (OpenAI) ввиду большей скорости обработки данных, позволяющей реализовывать потоковый режим работы системы, и меньших требований к вычислительной мощности машины. Модуль перевода разработанной интеллектуальной системы построен на модели T5-large-1024 (Text-to-Text Transfer Transformer), адаптированной для многоязычных задач.

Результаты. Предложен способ создания интеллектуальной системы распознавания речи – модульная архитектура системы распознавания и перевода речи, реализован прототип и замерены метрики. Система показала следующие результаты: для русско-английского перевода Cosine Similarity 0,6951, WER 0,529, BLEU Score 0,239; для каскадного русско-китайского перевода через английский язык Cosine Similarity 0,557, WER 0,748, BLEU Score 0,095. Исследования доказали, что применение каскадного перевода через английский язык повышает качество итогового текста на 32 % по метрике Cosine Similarity и на 25 % по BLEU Score по сравнению с прямым переводом. Результаты работы реализованного прототипа оказались удовлетворительными.

Заключение. Предложенная реализация системы распознавания речи может решать поставленную задачу с удовлетворительным для описанной проблемы качеством без рисков несанкционированного доступа к данным, поскольку работает без подключения к сети интернет. При использовании каскадного перевода через английский язык качество русско-китайского перевода улучшается на 32 % по метрике Cosine Similarity (с 0,423 до 0,557) и на 25 % по метрике BLEU Score (с 0,076 до 0,095). Предложенная информационная система может быть внедрена в образовательный процесс вне зависимости от учебной дисциплины, а также применена на выставках, конференциях, международных форумах. Возможен параллельный перевод на различные языки, что позволит всем участникам международных форумов активно участвовать в мероприятиях.

Ключевые слова: информационная система, агент, декодер, энкодер, трансформер, сверточные нейронные сети, транскрибация и перевод речи

Для цитирования. Кузьменков, Л. П. Система транскрибации речи и перевода с русского языка на китайский / Л. П. Кузьменков, В. А. Чуйко, Е. И. Козлова // Информатика. – 2025. − Т. 22, № 3. – С. 25–34. – DOI: 10.37661/1816-0301-2025-22-3-25-34.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 08.07.2025
Подписана в печать | Accepted 23.07.2025
Опубликована | Published 30.09.2025

Speech transcription and translation system from Russian to Chinese

Leonid P. Kuzmenkov, Vladislav A. Chuyko, A. I. Kazlova✉

Belarusian State University,
av. Nezavisimosti, 4, Minsk, 220030, Belarus
✉*E-mail: kozlova@bsu.by*

Abstract

Objectives. The aim of the work is to develop the architecture of an information system for transcription and translation of speech, implement its blocks and test their operation.

Methods. The existing methods of speech recognition are considered; a comparative analysis of speech recognition and text translation models is carried out. The speech transcription process includes several successive stages: collection and preliminary processing of the audio signal, extraction of acoustic features, direct speech recognition, post-processing and text correction, and output of the result. At the stage of audio signal pre-processing, a combination of specialized libraries is used to prepare data for subsequent analysis. To normalize the recording parameters, the Librosa library is used, which allows resampling the signal to a standard frequency of 16 kHz and converting it to a monophonic format. To suppress background noise and highlight the speech component, the Demucs neural network model is used. The spectral subtraction algorithm additionally corrects residual noise. Speech activity segmentation (VAD) is performed using an energy detector from WebRTC, automatically highlighting speech fragments and removing pauses. The whisper-turbo (OpenAI) model was chosen to implement the speech recognition system due to the higher data processing speed, which allows implementing the streaming mode of the system, and lower requirements for the computing power of the machine. The translation module of the developed intelligent system is built on the T5-large-1024 (Text-to-Text Transfer Transformer) model, adapted for multilingual tasks.

Results. A method for creating an intelligent speech recognition system is proposed - a modular architecture of the speech recognition and translation system, a prototype is implemented and metrics are measured. The system showed the following results: for Russian-English translation Cosine Similarity 0.6951, WER 0.529, BLEU Score 0.239; for cascade Russian-Chinese translation through English Cosine Similarity 0.557, WER 0.748, BLEU Score 0.095. Research has shown that the use of cascade translation through English improves the quality of the final text by 32% according to the Cosine Similarity metric and by 25% according to BLEU Score compared to direct translation. The results of the implemented prototype were satisfactory.

Conclusion. The proposed implementation of the speech recognition system can solve the task with quality satisfactory for the described problem without risks of unauthorized access to data, since it works without an Internet connection. When using cascade translation through English, the quality of Russian-Chinese translation improves by 32% according to the Cosine Similarity metric (from 0.423 to 0.557) and by 25% according to BLEU Score (from 0.076 to 0.095). The proposed information system can be implemented in the educational process regardless of the academic discipline, and also used at exhibitions, conferences, and international forums. Parallel translation into different languages is possible, which will allow all participants of international forums to actively participate in its events.

Keywords: information system, agent, decoder, encoder, transformer, convolutional neural networks, speech transcribing and translation

For citation. Kuzmenkov L. P., Chuyko V. A., Kazlova A. I. *Speech transcription and translation system from Russian to Chinese*. Informatika [*Informatics*], 2025, vol. 22, no. 3, pp. 25–34 (In Russ.). DOI: 10.37661/1816-0301-2025-22-3-25-34.

Conflict of interest. The authors declare of no conflict of interest.

Введение. В условиях активного развития международного сотрудничества в сфере образования ограничивающим фактором для увеличения числа зарубежных студентов является нехватка преподавателей, знающих язык, который доступен для понимания иностранцам. Одним из способов решения этой проблемы служит информационная система, состоящая из модуля, транскрибирующего речь, и модуля, осуществляющего перевод. Данные модули являются агентами, а система – многоагентной. Многоагентная система – это система, образованная несколькими взаимодействующими интеллектуальными агентами. Агент в общем смысле представляет собой любой объект, способный действовать и воспринимать информацию. Представленная в данной работе многоагентная система состоит из двух агентов: агента, преобразующего входящий аудиопоток или аудиофайл в текст на языке преподавателя, и агента, осуществляющего перевод текста, полученного от первого агента, с языка преподавателя на язык, доступный для иностранных студентов. На данный момент на рынке информационных услуг представлено множество агентов и систем.

Большинство современных сервисов для транскрибации и перевода речи, таких как Otter.AI, Beey, Google Cloud Speech API и др., требуют подключения к интернету для выполнения своих функций. Это связано с использованием облачных технологий и нейросетей, которые обрабатывают данные на удаленных серверах. Такие подходы обеспечивают высокую точность распознавания речи, поддержку множества языков и возможность обработки сложных аудиофайлов, но делают их зависимыми от интернет-соединения. При использовании готового решения возникают различные проблемы, одна из которых – безопасность.

Существующие решения. На данный момент есть множество приложений, сайтов и сервисов для перевода текстов с одного языка на другой, некоторые из систем перевода оснащены функцией распознавания и синтеза речи: Google Translate, Speech Logger, Яндекс Переводчик, Microsoft Translator, Talkao Translate.

Ключевым недостатком перечисленных сервисов является то, что их функциональность ограничена в офлайн-режиме только переводом. Большинство сервисов не предоставляют никаких услуг без подключения к сети интернет и имеют закрытый код. При использовании сервисов с закрытым исходным кодом возникает проблема кибербезопасности пользователей. Закрытый исходный код делает невозможным проверку объемов данных, собираемых сервисом. Пользователи вынуждены доверять заявлениям компаний о политиках обработки данных без возможности проверить декларированные утверждения. Закрытые системы не позволяют быть уверенным в отсутствии сбора персональных данных пользователя. Вышеперечисленные факторы создают риски утечек и неправомерного использования информации. В частности, для функции голосового ввода в условиях активного развития методик синтеза изображений или голоса, основанных на искусственном интеллекте (дипфейк), проблема утечек персональных и авторских данных крайне актуальна. Для демонстрации примера описанной выше проблемы рассматривается модель: преподаватель проводит два лекционных занятия в неделю, за месяц 9×2 ч ≈18 ч. Оцененного количества данных хватило бы для обучения диффузионной модели и фальсификации речи.

Методы разработки. На рис. 1 изображена архитектура прототипа информационной системы распознавания речи. На вход интеллектуальной системы подается аудиофайл, записанный с помощью микрофона. Затем с применением библиотеки librosa производится передискретизация полученного сигнала, так как модели распознавания речи корректно работают только с аудиофайлами с определенными частотой и глубиной дискретизации. Полученный аудиофайл подается на вход модели распознавания речи. Модель распознавания речи в общем случае со-

стоит из блока извлечения признаков, энкодера и декодера. Блок извлечения признаков представлен алгоритмом преобразования непрерывного сигнала в мел-спектрограмму (спектрограмма, показывающая частотное содержание аудиосигнала во времени, представленная шкалой Мел), и сверточной нейронной сетью, используемой для понижения размерности и уменьшения вычислительных затрат на следующих этапах. Шкала Мел – это перцептивная шкала, аппроксимирующая нелинейную частотную характеристику человеческого уха.

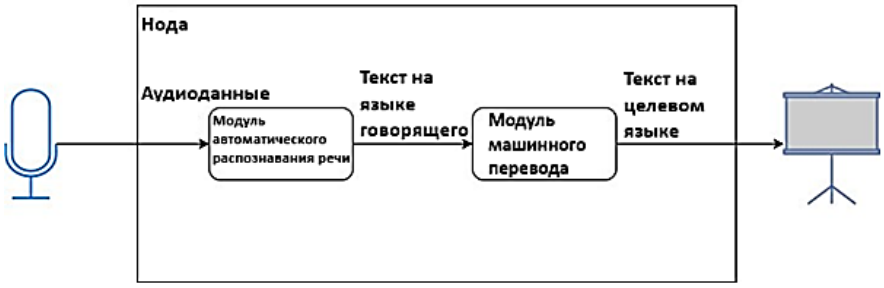

Рис. 1. Архитектура системы распознавания речи
Fig. 1. Architecture of the speech recognition system

В качестве энкодера используется трансформер или его вариации, что позволяет эффективно обрабатывать входные данные благодаря механизму многоголосового внимания и позиционному кодированию, которые учитывают контекст и последовательность токенов [1]. Декодер представлен трансформером – авторегрессионной моделью, основанной на архитектуре глубоких нейронных сетей. Сложная архитектура декодера, построенная на трансформерах, необходима для исправления ошибок транскрибации за счет предсказания вероятностей разных вариантов текста, который произнес человек на аудиозаписи, и выбора наиболее естественного варианта. Также существуют модели распознавания речи без энкодера, с соединенными блоками извлечения признаков, и с энкодером (например, конформер – архитектура, комбинирующая сверточные нейронные сети и механизм самовнимания).

Для оценки качества всей системы распознавания речи и отдельных агентов используются следующие метрики: косинусное подобие, алгоритм BLEU (bilingual evaluation understudy), расстояние Левенштейна (word error rate, WRE) [2]. Для получения значения косинусного подобия рассчитывается скалярное произведение векторных представлений двух текстов, полученных с помощью статической меры *tf–idf*:

$$tf - idf(t,d,D) = tf(t,d) \times idf(t,D). \quad (1)$$

Частота термина *tf* вычисляется по формуле

$$tf(t,d) = \frac{n_t}{\sum_k n_k}, \quad (2)$$

где n_t – число вхождений слова t в документ, а в знаменателе – общее число слов в данном документе.

Обратная частота документа *idf* вычисляется по формуле

$$idf(t,D) = \log_2 \frac{|D|}{\left|\{d_i \in D \mid t \in d_i\}\right|}, \quad (3)$$

где $|D|$ – число документов в коллекции; $\left|\{d_i \in D \mid t \in d_i\}\right|$ – число документов из коллекции D, в которых встречается t (когда $n_t \neq 0$).

Затем полученный результат делится на произведение норм векторных представлений.

BLEU – это автоматическая метрика для оценки качества текста, сгенерированного системой машинного перевода или распознавания речи, путем сравнения с эталонными (референсными)

переводами, выполненными человеком. BLEU анализирует совпадения *n*-грамм (последовательностей из *n* слов) между кандидатом (текст, который оценивается) и одним или несколькими референсами. Обычно используются *n*-граммы от одной до четырех (униграммы, биграммы, триграммы, четырехграммы).

BLEU вычисляется по формуле

$$\text{BLEU} := BP \cdot e^{(\sum_{n=1}^{\infty} \omega_n \ln p_n)}, \tag{4}$$

где p_n – модифицированная точность *n*-грамм, *BP* – штраф за краткость, ω_n – положительные веса, в сумме дающие единицу (в базовой версии метрики $\omega_n = 1/N$, $n = 1, 2, \ldots, N$). Модифицированная точность *n*-грамм рассчитывается по формуле

$$p_n := \frac{\sum_{C} \sum_{s \in C_n} \min(Count(s, C), \max_i (Count(s, R^{(i)})))}{\sum_{C} \sum_{s \in C_n} Count(s, C)}, \tag{5}$$

где Σ_c – суммирование по всем переводам-кандидатам *C* в текстовом корпусе; C_n – множество всех *n*-грамм, извлеченных из кандидата *C*; *s* – конкретная *n*-грамма из множества C_n; *Count*(*s*, *C*) – количество вхождений *s* в *C*; *Count*(*s*, $R^{(i)}$) – количество вхождений *s* в эталонный перевод $R^{(i)}$.

Штраф за краткость (brevity penalty) рассчитывается по формуле

$$BP = \begin{cases} 1, & c > r, \\ \exp(1 - \frac{r}{c}), & c \le r, \end{cases} \tag{6}$$

где *c* – длина корпуса кандидата, *r* – длина корпуса референса.

Длина корпуса кандидата рассчитывается по формуле

$$c := \sum_{C} |C|, \tag{7}$$

где $|C|$ – количество слов в кандидате *C*.

Длина корпуса референса *r*, представляемая как сумма длин наиболее близких по длине эталонных строк, вычисляется по формуле

$$r := \sum_{C} \arg\min_i \left| |C| - \left| R^{(i)} \right| \right|, \tag{8}$$

где $|C|$ – количество слов в кандидате, $|R^{(i)}|$ – количество слов в *i*-м эталонном переводе.

Метрика BLEU позволяет объективно сравнивать качество автоматически сгенерированных текстов с эталонными, учитывая совпадения *n*-грамм и корректируя результат с помощью штрафа за краткость, чтобы избежать завышения оценки за слишком короткие ответы.

WRE – расстояние Левенштейна, нормированное по количеству слов. Расстояние Левенштейна – это метрика, измеряющая по модулю разность между двумя последовательностями символов [3]. Она определяется как минимальное количество односимвольных операций (например, вставки, удаления, замены), необходимых для превращения одной последовательности символов в другую.

На рис. 2 изображена архитектура state-of-the-art модели Whisper компании OpenAI, используемой как «начало» в предлагаемой информационной системе. Модель работает локально, без подключения к сети интернет, что решает проблему безопасности персональных данных пользователя [4].

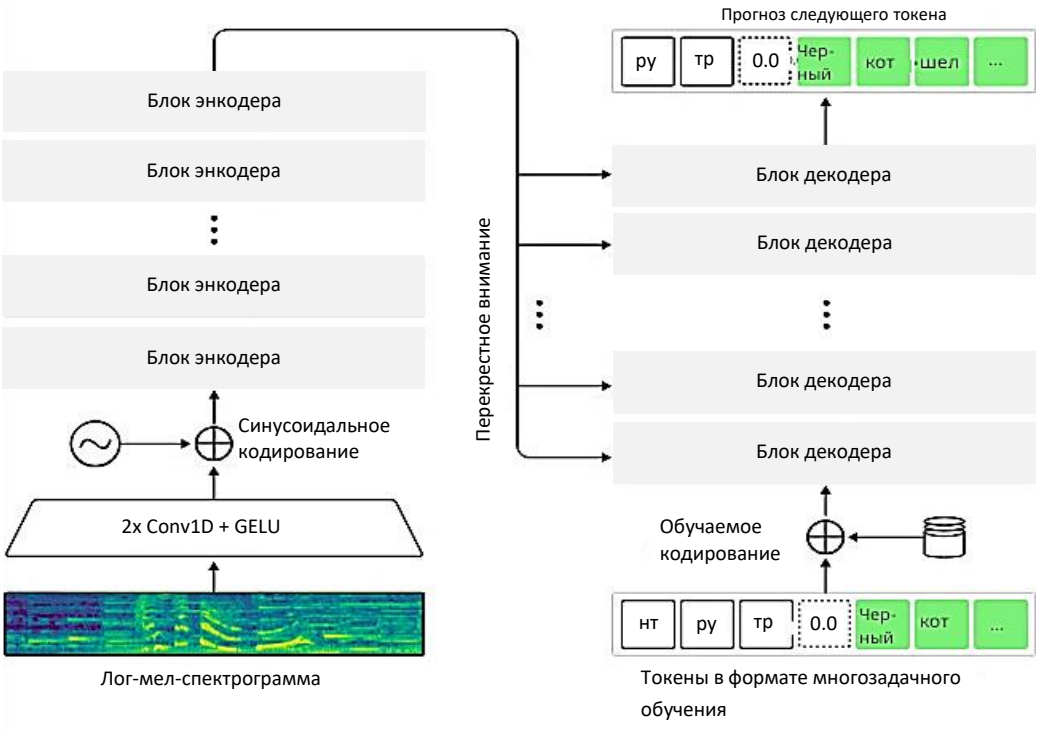

Рис. 2. Архитектура модели Whisper

Fig. 2. Whisper model architecture

Для оценки качества модели вышеперечисленные метрики рассчитываются на основании сравнения зачитанного в микрофон текста и текста, полученного после распознавания. Сравнительная характеристика разных моделей распознавания речи приведена в табл. 1.

Таблица 1

Сравнительная характеристика моделей распознавания речи

Table 1

Comparative characteristics of speech recognition models

Модель *Model*	Vram, Gb	Время распознавания четырехминутного текста *Recognition time for a four-minute text*	Cosine similarity	Word error rate	BLEU Score
vosk-model-small-ru-0.22	0,5	19 с 355 мс	0,6399	0,6050	0,1957
vosk-model-ru-0.42	2	2 мин 20 с	0,7560	0,4798	0,3699
whisper-tiny	1	10 с 938 мс	0,4813	0,8382	0,0404
whisper-base	1	11 с 481 мс	0,5564	0,7746	0,074
whisper-small	2	32 с 428 мс	0,6412	0,6570	0,144
whisper-medium	5	4 мин	0,7156	0,5145	0,364
whisper-large	10	4 мин	0,8303	0,499	0,3185
whisper-turbo	**6**	**1 мин 58 с**	**0,7293**	**0,4566**	**0,4117**

Наилучшим качеством транскрибации речи обладают модели whisper-large и whisper-turbo. Для реализации системы распознавания речи в данной работе выбрана модель whisper-turbo ввиду большей скорости обработки данных, позволяющей реализовывать потоковый режим работы системы, и меньших требований к вычислительной мощности машины. Затем полученный текстовый файл подается на вход модели переводчика, зачастую также основанной на вариациях трансформеров.

Результаты. На рис. 3 изображена архитектура модели T5 (Text-to-Text Transfer Transformer), выбранная как модель для обучения на большом корпусе текстов на английском, русском и китайском языках.

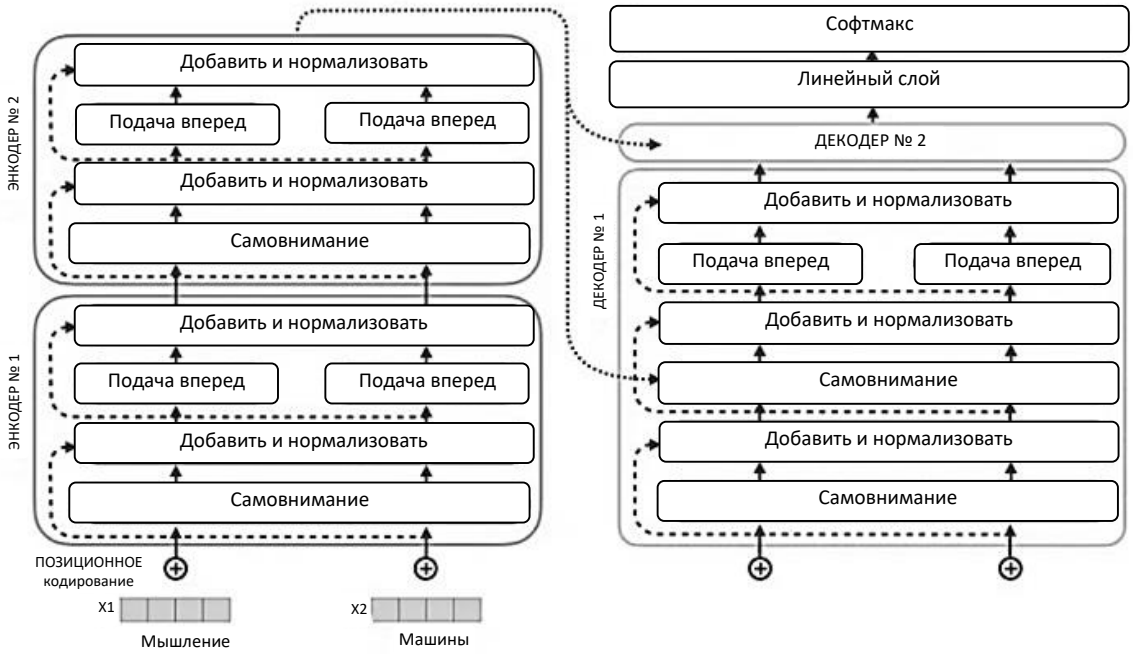

Рис. 3. Архитектура модели T 5
Fig. 3. Architecture of the T 5 model

Сравнительная характеристика разных моделей перевода приведена в табл. 2.

Таблица 2
Сравнительная характеристика моделей перевода
Table 2
Comparative characteristics of translation models

Модель *Model*	Нужен интернет *Need internet*	Vram, Gb	Время перевода *Translation time*	Cosine similarity	Word error rate	BLEU Score
utrobinmv/utrobinmv/t5_translate_en_ru_zh_small_1024	Нет	<3	24 с 583 мс	0,7185	0,5892	0,215
utrobinmv/t5_translate_en_ru_zh_large_1024	Нет	3	49 с 629 мс	0,6951	0,529	0,2387
utrobinmv/t5_translate_en_ru_zh_large_1024_v2	Нет	>4	1 мин	0,6312	0,6618	0,1723
google translate	Да	<1	0 с	0,7798	0,3714	0,4056
Яндекс переводчик	Да	<1	0 с	0,7295	0,5539	0,2925

Необходимость подключения к сети интернет для осуществления перевода создает угрозу безопасности. В табл. 2 добавлены для сравнения метрики качества перевода моделей с открытым кодом и коммерческих сервисов. Для описываемой интеллектуальной системы была выбрана модель utrobinmv/t5_translate_en_ru_zh_large_1024, имеющая следующие значения качества:

Cosine Similarity: 0,6951,

word error rate: 0,529,

BLEU Score: 0,23 879 138 591 567 485.

Метрики качества перевода распознанного текста с русского языка на китайский и обратно:

Cosine Similarity: 0,4232,

word error rate: 0,8233,

BLEU Score: 0,07 639 410 899 117 652.

Видно, что данные результаты далеки от желаемых. Для улучшения качества перевода после проведения ряда исследований и тестов было решено использовать английский язык как промежуточный для получения более качественного перевода, поскольку результаты оценки качества прямого перевода с английского на русский и с китайского на английский выше, чем прямого перевода с русского на китайский.

Метрики качества перевода распознанного текста с русского языка на английский, затем на китайский, затем обратно на английский и потом на русский представлены ниже:

Cosine Similarity: 0,5571,

word error rate: 0,7484,

BLEU Score: 0,0 954 180 649 678 097.

Анализируя полученные результаты по метрикам, можно отметить, что при переводе на китайский язык улучшение результата происходит при использовании промежуточного перевода на английский язык и далее – на русский как при прямом, так и обратном проходе.

Метрики качества интеллектуальной системы в приложении к английскому языку:

Cosine Similarity: 0,4107,

word error rate: 0,8786,

BLEU Score: 0,07 190 972 701 489 706.

Метрики качества интеллектуальной системы в приложении к китайскому языку:

Cosine Similarity: 0,5470,

word error rate: 0,7977,

BLEU Score: 0,1 076 237 727 294 554.

Достоинством описанной архитектуры системы является простота замены отдельных модулей, что позволяет проверять множество вариантов агентов без изменения общей архитектуры для проведения сравнительного анализа при поиске оптимального решения.

В противовес описанной каскадной архитектуре, состоящей из блоков транскрибации аудио- и машинного перевода текста, можно поставить прямые модели, где одна нейросетевая модель сразу преобразует аудиосигнал в перевод без промежуточного текстового представления на исходном языке. Сравнение данных методов представлено в работе Apple, где авторы показали, что даже при близких BLEU-оценках согласованность между транскриптом и переводом у каскадных систем выше, чем у монолитных моделей. Данный подход может улучшать качество перевода, но зачастую снижает точность транскрипта [5]. Аналогичную картину описали исследователи из Vicomtech: в их эксперименте пользователи отдали предпочтение каскадной схеме в 42 % случаев против 17 % за прямую схему, несмотря на сопоставимый уровень автоматических метрик [6]. В описываемом случае перевода с русского языка на китайский проблемой является малое количество параллельных корпусов, содержащих русскоязычные аудиоданные и соответствующие китайские текстовые транскрипции. Она усугубляется значительными типологическими различиями между русским и китайским языками, относящимися к разным лингвистическим семьям.

Тестирование показало, что для русско-английского перевода система достигает значений Cosine Similarity 0,695, WER 0,529 и BLEU Score 0,239.

Заключение. Система распознавания речи в своем настоящем виде может решать поставленную задачу с удовлетворительным для описанной проблемы качеством без рисков несанкционированного доступа к данным, поскольку работает без подключения к сети интернет. При использовании каскадного перевода через английский язык качество русско-китайского перевода улучшается на 32 % по метрике Cosine Similarity (с 0,423 до 0,557) и на 25 % по метрике BLEU Score (с 0,076 до 0,095). Предложенная информационная система может быть внедрена в образовательный процесс вне зависимости от учебной дисциплины. При некоторых доработках агенты распознавания речи и перевода текста смогут оперировать любыми доменными терминами и аббревиатурами, что позволит устранить необходимость изучения преподавателем языка иностранных граждан и (или) изучения иностранными гражданами языка, доступного для преподавателя. Кроме сферы образования предложенная интеллектуальная система может быть применена на выставках, конференциях, международных форумах. Возможен параллельный перевод на различные языки, что позволяет участникам международных форумов активно участвовать во всех мероприятиях.

Модель для перевода работает локально, переобучать ее нецелесообразно, но в будущем существует возможность дообучения на лекционных материалах для повышения точности перевода доменных терминов и аббревиатур. Возможна замена модели автоматического распознавания речи на модель, дообученную на доменной лексике, для увеличения точности транслитерации и перевода терминов, а также аббревиатур, которые часто встречаются в лекционных материалах и на которых модель общего назначения компании OpenAI не обучалась [7]. В дальнейшем планируется реализовать потоковую обработку речи, упростить развертывание системы, дообучить транскрибирующего агента и агента для перевода на доменной лексике.

Вклад авторов. *В. А. Чуйко* разработал концепцию работы, провел ряд исследований, критический анализ модели и текста статьи. *Л. П. Кузьменков* осуществил разработку системы, провел исследования и написал текст статьи. *Е. И. Козлова* провела критический анализ содержания статьи и подготовила окончательный вариант работы для публикации.

References

1. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., …, Polosukhin I. *Attention Is All You Need*, 2017. Available at: https://arxiv.org/abs/1706.03762 (accessed 12.05.2025).

2. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation. *40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, Jule 2002*, pp. 311–318.

3. Tzoukermann E., Miller C. Evaluating automatic speech recognition in translation. *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas, Boston, MA, March 2018*, vol. 2: MT Users' Track, pp. 294–302.

4. Sperber M., Setiawan H., Gollan C., Nallasamy U., Paulik M. Consistent transcription and translation of speech. *Transactions of the Association for Computational Linguistics (TACL)*, 2020, vol. 8, pp. 695–709.

5. Etchegoyhen T., Arzelus H., Gete H., Alvarez A., Torre I. G., …, Fernandez E. B. Cascade or direct speech translation? A case study. *Applied Sciences*, 2022, vol. 12, iss. 3, pp. 1097.

6. Radford A., Kim J. W., Xu T., Brockman G., McLeavey C., Sutskever I. *Robust Speech Recognition via Large-Scale Weak Supervision*, 2022. Available at: https://arxiv.org/abs/2212.04356 (accessed 12.05.2025).

7. Kumar L. A., Renuka D. K., Chakravarthi B. R., Mandl T. *Automatic Speech Recognition and Translation for Low Resource Languages*. Wiley-Scrivener, 2024, 496 p.

Информация об авторах

Кузьменков Леонид Павлович, студент, Белорусский государственный университет.
E-mail: salamandradrags@yandex.ru

Чуйко Владислав Александрович, магистр физико-математических наук, старший преподаватель, Белорусский государственный университет.
E-mail: Vchuyko@bsu.by

Козлова Елена Ивановна, кандидат физико-математических наук, доцент, Белорусский государственный университет.
E-mail: kozlova@bsu.by

Information about the authors

Leonid P. Kuzmenkov, Student, Belarusian State University.
E-mail: salamandradrags@yandex.ru

Vladislav A. Chuyko, M. Sc. (Phys.-Math.), Senior Lecturer, Belarusian State University.
E-mail: Vchuyko@bsu.by

Alena I. Kazlova, Ph. D. (Phys.-Math.), Assoc. Prof., Belarusian State University.
E-mail: kozlova@bsu.by