

АВТОРАСПОЗНАВАНИЕ РЕЧИ И ОЦЕНКА ЕГО ПРИМЕНИМОСТИ В АВИАЦИИ НА СОВРЕМЕННОМ ЭТАПЕ

СҮЮНБАЕВА А.Ж.^{1*}, ШЕЛЕСТИЮК Е.В.²

*Суюнбаева Алтынгүл Жакиповна¹ - Кандидат филологических наук, доцент, Военный институт Сил воздушной обороны, г. Ақтөбе, Казахстан

E-mail: altin_suenbaeva@mail.ru, <https://orcid.org/0000-0002-0676-2028>

Шелестюк Елена Владимировна² - Доктор филологических наук, профессор, Челябинский государственный университет, г. Челябинск, Российская Федерация

E-mail: shelestiuk@yandex.ru, <https://orcid.org/0000-0003-4254-4439>

Аннотация. В статье представлены результаты эксперимента по сравнительной оценке эффективности двух современных программ распознавания устной речи – ChatGPT (на платформе <https://chatgptchatapp.com>) и Speech2Text (<https://speech2text.ru>) – при обработке аудиозаписей на трех языках: русском, английском и казахском. Выбор именно этих инструментов обусловлен их широкой доступностью, использованием передовых нейросетевых моделей и возможностью обработки разнообразных типов речи. Целью эксперимента было определить, насколько правильно и точно системы распознают речь и имеются ли отличия в распознавании речи на разных языках. Исследование охватывало такие формы устной коммуникации, как формализованная диктовка, юридический текст, рассказ на родном языке, спонтанный диалог. Для повышения валидности исследования в качестве исходных материалов использовались как запись относительно «чистой» речи, так и аудиофайлы с речью невнятной, сопровождаемой шумами, выраженными акцентами, нетипичной паузацией и интонированием – распространенные в реальных условиях общения в авиации и других профессиональных средах. Внимание уделялось точности транскрипции на трех языках, способности систем воспроизводить структуру текста, корректно передавать имена собственные, сохранять паузы, расставлять знаки препинания и разделять речь по спикерам – ключевые аспекты для анализа естественного диалога. Была разработана методика транскрибирования, учитывающая речевые и неречевые феномены. Обе системы преобразуют аудиофайлы в машиночитаемый текст, однако реализуют этот процесс по-разному. Speech2Text специализируется на высокоточной транскрипции с разделением на спикеров, автоматическим расстановкой тайм-кодов и знаков препинания, поддерживает широкий спектр форматов и не имеет ограничений по длительности файлов. Сервис использует нейросети для распознавания речи даже в условиях плохой слышимости, что делает его особенно полезным для обработки реальных, неидеальных записей. ChatGPTChatApp, представляющий собой многофункционального ИИ-ассистента на базе различных моделей (включая GPT-4o, DeepSeek, Qwen3 и другие), демонстрирует высокие способности к анализу речи, генерации текста, пониманию контекста и обработке документов. Хотя его основная направленность — не транскрибация, он достаточно эффективно обрабатывает аудиоданные, особенно в сочетании с другими задачами, такими как перевод, анализ содержания или генерация отчетов. Полученные в ходе эксперимента данные позволяют оценить сильные и слабые стороны каждой из систем и определить их применимость в зависимости от задачи – от точной транскрипции до комплексного анализа речевой коммуникации.

Ключевые слова: ASR-MT, формальная диктовка, спонтанная речь, многоязычная речь, автоматическое распознавание речи, когнитивная нагрузка, идентичность транскрипта оригиналу.

Введение. В настоящее время привлекают внимание различные программы с функцией распознавания речи (ASR), т.е. преобразования устного сообщения в текст, и их эффективное применение в разных сферах деятельности. Распознавание устной речи в данное время осуществляется в разных информационно-технологических устройствах. Эффективность применения распознавания устной речи (используя голос) заключается в скорости и доступности взаимодействия коммуникантов. На сегодняшний день данной функцией программы снабжается большинство устройств. Для автоматического распознавания устной речи используются различные программы и методы. Данный процесс осуществляется в любом гаджете (телефон, компьютер, планшет и т.д.), который имеет микрофон.

Развитие технологий автоматического распознавания речи (ASR) открывает новые возможности для их применения в различных профессиональных сферах, включая авиацию, судебную систему, образование и медицину. Особую актуальность представляет использование таких программных решений в многоязычной среде, где важным фактором становится способность программного обеспечения точно обрабатывать устную речь на разных языках с различными акцентами и в разных стилистических и функциональных контекстах – от формальной диктовки до свободного диалога.

Исследование направлено на выявление ключевых характеристик и ограничений современных программ распознавания речи в зависимости от языка, типа текста и его стилистической формы.

Результаты исследования могут быть применены в сфере авиации при осуществлении радиообмена (диалогическая речь пилота с диспетчером).

Авиация – одна из самых интернациональных отраслей, где пилоты, диспетчеры и технический персонал могут говорить на различных языках. Официальным языком международной авиации является английский, но в реальной практике встречаются ситуации, когда:

- контроллеры или пилоты используют неаутентичные варианты английского (например, с акцентами, характерными для азиатских, африканских или латиноамериканских стран);
- возникают случаи использования местных языков, особенно в региональной авиации или при чрезвычайных ситуациях;
- авиакомпании или государства стремятся к билингвизму (например, использование английского и французского в африканских странах).

Самыми сложными факторами в распознавании устной речи являются неразборчивое произнесение слов, акценты, ошибки и исправления, нелогичные паузация и интонация, посторонний шум, разделение на спикеров, а также различие акустических моделей для разных языков, затрудняющее определение смены языков в диалоге, адекватную фиксацию попеременной/одновременной/параллельной речи на разных языках.

Несмотря на трудности, внедрение и развитие ASR-систем способно кардинально улучшить коммуникацию в мультиязычной среде, повысить безопасность и оперативность принятия решений в сложных и связанных с риском областях, таких как авиация. Эти технологии при постоянном развитии, совершенствовании позволяют распознавать голосовые команды, контролировать информационные потоки в экстремальных условиях и точно воспроизводить естественные диалоги на разных языках. Отсюда следует актуальность представленного эмпирического исследования, в котором на основе сравнительного анализа двух современных ASR-систем — ChatGPT и Speech2Text — оценивается эффективность в распознавании устной речи на трех языках (русском, английском и казахском) в различных коммуникативных контекстах: от формализованной диктовки до спонтанного диалога.

Объектом статьи является сравнительное тестирование технологий автоматического распознавания речи (ASR), предметом – выявление эффективности распознавания устной речи на трех языках: казахском, русском и английском, идентичности или неидентичности транскриптов устному оригиналу. Целью исследования является определение точности, адекватности и нюансировки распознавания речи двумя инструментами (системами, программами) распознавания устной речи – ChatGPT (на платформе <https://chatgptchatapp.com>) и Speech2Text (<https://speech2text.ru>) – при обработке аудиозаписей на трех языках. Особое внимание обращается на казахский язык, который характеризуется меньшей доступностью автоматического распознавания речи вследствие несовершенства и дефицита мультиязычных систем ASR.

В качестве задач выступают:

1. Автоматическое распознавание голоса человека (устного сообщения) в виде текста (на экране телефона, компьютера и т.д.), т.е. физический электрический сигнал расшифруется при помощи аналого-цифрового преобразователя.

2. Анализ результатов эксперимента, определяющий необходимые рекомендации и перспективные исследовательские направления.

Материалы и методы исследования.

Методы и инструменты исследования включают аудиозапись, обработку материалов с помощью ChatGPT и Speech2Text, их транскрибацию и интерпретацию по разным параметрам: транскрипция с разделением на спикеров, автоматическая расстановка тайм-кодов и знаков препинания, понимание контекста, учет особенностей воспроизведения структуры текста, передача имен собственных, сохранение пауз и разделение речи по спикерам. В исследовании также применен статистический подход, методы когнитивной лингвистики, дискурсивного анализа и социолингвистики. Статистический анализ был использован в финальном этапе исследования для интеграции данных. Описательно-аналитические методы социолингвистики и дискурс-анализа применялись при анализе и интерпретации языкового материала из текстов.

Материал исследования включает аудиофайлы с фрагментами речи трехязычных носителей языка с разными компетенциями в трех языках (казахский как родной, русский как второй, английский как иностранный). Для объективной оценки использовались как заранее подготовленные оригинальные тексты, которые произносились вслух и затем подвергались автоматической обработке, так и естественные диалоги. Для повышения валидности исследования использовались как запись относительно «чистой» речи, так и аудиофайлы с речью невнятной, сопровождаемой акцентами, ошибками, нетипичной паузой и интонированием, распространенные в реальных условиях коммуникации.

В эксперименте участвовали 18 человек (курсанты Военного института Сил Воздушной обороны). Полученные результаты сравнивались с исходными текстами в виде печатного текста или аудиозаписей (в случае естественного диалога) с целью определить количество ошибок и примерное количество необходимых манипуляций для корректировки транскриптов. Далее представляем таблицу, которая показывает сравнение эффективности программ распознавания речи ChatGPT и Speech2Text на различных типах устной речи.

Таблица 1. Сравнение эффективности программ распознавания речи ChatGPT и Speech2Text на различных типах устной речи

Язык	Исходный текст (текст для диктовки)	Полученный текст (результат диктовки или свободной беседы в виде письменного текста)		Распознавания, примерное количество манипуляции для редактирования
		Сервис 1 Программа ChatGPT (https://chatgptchatapp.com)	Сервис 2 Программа Speech2Text (https://speech2text.ru)	
Аудиофайл 1 Язык: русский. Жанр: диктовка, чтение текста.	Статья 5. Употребление русского языка. В государственных организациях и органах местного самоуправления наравне с казахским официально употребляется русский язык. Согласно	Статья 5. Употребление русского языка. В государственных организациях, в органах местного самоуправления наравне с казахстаном официально употребляется русский язык. Согласно Постановлению	Статья 5. Употребление русского языка в государственных организациях, в органах местного самоуправления наравне с казахским официально употребляется русский язык. Согласно	Общее количество слов в оригинале – 56. ChatGPT: Количество неправильно распознанных слов – 5. 5 манипуляций. Speech2Text: Количество неправильно

	Постановлению Конституционного Совета РК от 8 мая 1997 года №10/2: «Данная конституционная норма понимается однозначно, что в государственных организациях и органах местного самоуправления казахский и русский языки употребляются в равной степени, одинаково, независимо от каких-либо обстоятельств».	Конституционного Совета РК от 8 мая 1997 года №10/2: «Данной конституционная – норма, понимание однозначно, что в государственных организациях и в органах самоуправления наравне с казахский и русский языки употребляются в равной степени. Независимо от каких-либо обстоятельств.	постановлению Конституционного совета республик Казахстана от 8 мая 1997 года номер 10 и 2: «Данная Конституция-Норма, понимание однозначно, что в государственных организациях и органах местного самоуправления казахский и русский языки употребляются в равной степени, одинаково, независимо от каких-либо обстоятельств.	распознанных слов – 7. 7 манипуляций.
Аудиофайл 2 Язык: английский Жанр: диктовка, чтение текста.	Alone at Sea Debra and Andrew Veal entered a 3,000-mile rowing race across the Atlantic from Tenerife to Barbados. There were thirty-four other rowing teams. They started the race together, but after two weeks Andrew left the race because he was frightened of the ocean. Debra didn't want to stop the race, and she decided to continue. She rowed the remaining 2,290 miles alone. The journey took 113 days. Debra arrived in Barbados seventy days after the winning team, but for most people she was the hero of the race. The editor of The Times wrote, 'The winner of the race is the girl who came last.'	Hello. Deborah and Andrew will enter a 3,000 mile race rowing across the Atlantic from Tenerife to Barbados. There were 34 other wrong teams. They started the race together, but after two weeks, Andrew left the race because he was frightened of the ocean. Deborah didn't want to stop the race, and she decided to continue. She wrote the remaining 2,290 miles alone. The journey took 113 days. Deborah arrived in Barbados 70 days after the winning team. But for most people, she was the hero of the race. The editor of the Times wrote, «The winner of the race is the girl who came last.»	Alone at sea Deborah and Andrew Veal entered a 3,000-mile-long rowing race across the Atlantic from Tenerife to Barbados. There were 34 other wrong teams. They started the race together, but after two weeks, Andrew left the race because he was frightened of the ocean. Deborah didn't want to stop the race, and she decided to continue. She rode the remaining 2,290 miles alone. The journey took 113 days. Deborah arrived in Barbados 70 days after the winning team. But for most people she was the hero of the race, the editor of the Times Road. The winner of the race is the girl who came last.	Общее количество слов в оригинале – 107. ChatGPT: Количество неправильно распознанных слов – 7. 7 манипуляций. Speech2Text: Количество неправильно распознанных слов – 7. 7 манипуляций.
Аудиофайл 3 Язык: казахский. Жанр: свободное говорение,	Спикер 1: — Ия, досым, сен қандай кітап оқып отырсың? Спикер 2: - Қандай кітап дегенде, Абай	Жадосым, на? Қандай кітап оқушысын? Қандай кітап дегенде, абай қонанбайдың Искендір атының поемасын оқып жатырдым. Өзім	Спикер 1: Достмын сенің қандай кітап қолдысыздар? Қандай кітап деңді? Абай-Хонамба етін... Искендер атты пайі масын қолды «Бзім	Общее количество слов в оригинале – 319. ChatGPT: Количество неправильно распознанных

пересказ/расказ.	Құнанбайдың «Ескендір» атты поэмасын оқып отырмын. Өзім жалпы Ескендір туралы білгім келіп жүрген. Жалпы әлемді жаулап алған дейді ғой. Спикер 1: . А, Александр Македонский ма? Спикер 2: Я, я, я. Естіп бе ең? Спикер 1: Я, я, естігенім бар. Мынандай бір қызықты, сол тұралы аңыз бар білесің бе? Спикер 2: Айтты берші, сол менде қызығып жүрмін соған. Спикер 1: Александр Македонскийдің басында бір мүйіз бар деген бір аңыз бар. Ол бір ауылға келіп тоқтайды. Бүкіл әлемді жаулап алып, жаулап алған жорықтар кезінде бір аулаға келіп тоқтайды. Со кезде, бүкіл ауылдың еркектері оның шашын алып беріп жатады. Со кезде, бірақ шашын алған кезде еркектердің со кезде басы кесіледі екен. Солай бүкіл ауылдың еркектері өліп, бір жас бала ғана қалады, оның кезегі келген кезде, оның шешесі жылап жібергісі келмейді. Бірақ жауынгерлер ол баланы алып кетеді. Ол бала Искендірдің, басындағы	жалықпайды. Искендір туралы білгім келіп жүрген. Жалпы әлемді жаулап алған дейді ғой. А, Александр Македонский ба? Я, я, я. Естесіз бе? Я, я, естеген бар. Мынандай бір қызықты мынандай тұралы аңыз бар білесің бе? Айтпайсың со менде қызығып жүрген болсаң. Бір Александр Македонскийдің басында бір мүйіз бар деген бір аңыз бар. Ол бір аулаға келіп тоқтайды. Бүкіл әлемді жаулап алып, жаулап алған жорықтар кезінде бір аулаға келіп тоқтайды. Со кезде, е-е, бүкіл ауылдың еркектері оның шашын алып береді. Со, бірақ шашын алған кезде еркектердің со кезде басы кесіледі екен. Солай бүкіл ауылдың еркектері өліп, бір жас баланы қалып, i-i, оның кезегі кейінге ізде оның шешесі жылап жібереді. Сенің бірақ жауынгерлер ол баланы алып кетеді. Ол бала Искендірдің, е-е, басындағы дулығасын шешкен кезде мүйізін көріп шошып кетеді. Бірақ Искендір оған қорқпа деп айтады да шашын алып беруге, i-i, бұйырады. Бала шашын алып береді. Бірақ Искендір ол жас баланы аяйды да, бірақ жібере салады, бірақ бір, i-i, талап қояды. Егер де осыдан біреуге мен мүйізім бар екенін айтсаң, мен сенің басыңды аламын дейді. Солай арада жылдар өтеді. Е-е, ана бала қатты, соны ішіне ұстай алмайды. Қатты	жалпы Искендер туралы білігім кең жүрген жалпы әлемі де жаулабоды» дейді. Александр Македонскийдің бәрі? Есіне сипірі! Есіне жалп, ноға... Нандай бір қызықты бұр... Сұрғал алғанғыз бір. Бір Александр Македонский басында бұр мүйіз бар, дейін бір аңыз бар. Ол бұр ауала еп тұқтайды. Бүкіл әлінің жолауалы жорұқтары кезінде бр ауала еп тұқтайды. Сон кезде бүкіл аудың еркектері оны шашын алып беруді. Бірақ шашыналған кезде, еркектерінің со кезде басы кезілді. Сонда бүкіл аудың еркектері өліп, бір жаспалағана қалады. Оның кезде кейінгізді оның шеше сұшылап жөрек сеніп, бірақ жауың келдер ол балаң алып күтіп. Ол бала Искендер басында болғасын шешкенгізді мүйізін көріп шоршып күтіп. Бірақ Искендер оған қорқпай айтабырды, шашын алып беруге бұйырады. Бала шашыналып береді, бірақ ескендер ол жаспалаға аяйттар, бірақ жөре салады, бірақ бір талап қояды. Егерде осынан бір өге мен нүйілім бар екен айтысар, незе ембасын ғалам дейді. Солай арады жылдар өтеді. Ана бала қатты, үшінде сұйылмайды, қатты айтық қысы кереді. Солай бір қарияға кейіп айтады. Осында	слов – 27. 8 манипуляций. Speech2Text: Количество неправильно распознанных слов – 71. 23 манипуляций.
------------------	--	---	---	---

	қорықпа деп айтады да, шашын алып беруге бұйырады. Бала шашын алып береді. Бірақ Искендір ол жас баланы аяйды да, бірақ жібере салады, бірақ бір, талап қояды. Егер де осыдан біреуге мен мүйізім бар екенін айтсаң мен сенің басыңды аламын дейді. Солай арада жылдар өтеді. Аа, ана бала қатты, соны ішіне ұстай алмайды. Қатты айтқысы келеді. Солай бір қарияға келіп айтады. Осындай іштей бір сырым бар еді. Соны қайда шығару керек деп, қайда айтсам болады деп, шал болса ауылдағы бір іші кеуіп қалған бір құдықты көрсетеді. Сол құдыққа барып жан, бар даусыңмен айқайла дейді. Сені ешкім естімейді, бірақ ішіндегіні шығарасың дейді. Солай барып, бала құдыққа барып айқайлайды. «Ескендірдің басында мүйізі бар» деп үш рет айқайлайды. Барып, ішіндегісін шешеді. Барып шығарғанша айқайлайды. И, арада бір-екі апта өткеннен кейін сол құдықтың жанынан керуен шығады. Керуенде бір әші болады екен. Құдықтың ішін ашса, құдықтың түбіне қараса ішінде қамыстар жатыр. Сол қамыстың, біреуін үзіп алып содан сыбызғы жасайды. Сыбызғыны, ән	айтқысы келеді. Солай бір қарияға келіп айтады. Осындай іштей бір сырым бар еді. Соны қайда шығару керек деп, қайда айтсам болады деп, i-i, шал болса ауылдағы бір іші, қуық қалған бір құдық көрсетеді. Сол қуыққа барып жаның, бар даусыңмен айқайлайды. Сені ешкім естімейді, бірақ ішіндегіні шығарасың дейді. Солай барып, е-е, бала қуыққа барып айқайлайды. Искендірдің басында мүйізі бар деп үш рет айқайлайды. Барып, i-i, ішіндегісін шешеді. Барып шығарып айқайлайды. И, i-i, арада бір-екі апта өткеннен кейін сол қуықтан жаңа керюінен шығады. Керюенде бір әнші болатын екен. Қуықтың ішіне ашса, құдықтың түбіне қараса ішінде қамыс өсіп жатыр еді. Сол қамысты i-i, біреуін үзіп алып соның сыбызғысын істейді. Сыбызғыны, i-i, ән ойнайын деп шерткен кезде, i-i, сыбызғыдан бір i-i, дауыстар шығады. “Искендір	үште бір сырын бар еді. Соң қайда шығар өріп болып тұрп, қайда етсен болып тұрп. Шал болса, ауылдағы бір үші кеуіп қалан бір құлық тұрп ұрыстыр. Сол құлыққа барып, жан бар да ұстымен айғалайды. Сені ешкімесімейт, бірақ үшінде егені шығарасын дейді. Солай барып, барлық құлыққа барып айғаледі. Искендердің бастыңдан бұйызды бар, өшірет айғалет, бар үшінде үйсін үшін, барлық шарғаша үйін айғалет. Е, арада бір екі апты өткенінде, сол құлықтан жаңынан керуенді жатыр. Керуенде бір әші болады кең, құлықтың түміле қараса, үшінде қамыстар жатыр. Сол қамысты бір ең үзі болып, солын сұбызғы жатыр.	
--	--	--	---	--

	ойнайын деп шерткен кезде, сыбызғыдан бір дауыстар шығады. «Ескендір			
Аудиофайл 4 Язык: русский. Жанр: свободный диалог.	Спикер 1: — Вот, Макс, я сижу и думаю, что сегодня по распорядку дня после четвертой пары? Спикер 2: — Сегодня какой день недели? Спикер 1: Сегодня четверг. Спикер 2: В четверг у нас, по-моему, спортивно-массовая работа. Спикер 1: — Давай тогда скажем ребятам, чтобы сегодня сыграли в футбол. Спикер 2: — Давай, если командир разрешит, можно сыграть. Спикер 1: — Кого бы ты хотел взять в команду? Спикер 2: — Я хотел взять в команду Байдербека и Аравова, а на воротах Биг Пау (Big Power). Спикер 1: А почему именно Биг Пау? Спикер 2: Биг Пау, потому что он занимался и хорошо стоит на воротах, и у него высокий рост. Спикер 1: — Ну, можно сказать. Но ты же знаешь, что я тоже хорошо играю. И у вас, по-моему, шансов мало будет нас победить. Спикер 2: — Ну, это время покажет. Спикер 1: — Все тогда, увидим.	Без разделения на спикеров Вот, Макс, я сижу и думаю, что сегодня по распорядку дня после четвертой пары. Сегодня какой день недели? Сегодня четверг. Четверг у нас, по-моему, спортивно-массовая работа. Давай тогда скажем ребятам, чтобы сегодня сыграли в футбол. Давай, если командир разрешит. Можно сыграть. Э-э, кого бы ты хотел взять в команду? Я хотел взять в команду Бадербекова и Яроал - оно ворота - Бикпаов. А почему именно Бикпаова? Бикпаов, потому что он занимался и хорошо стоит на воротах, и у него высокий рост. Ну, можно сказать. Но ты же знаешь, то что я тоже хорошо играю. И у вас, по-моему, шансов мало будет нас победить. Ну, это время покажет. Всё тогда, увидим.	Спикер 1: — Вот, Макс, я сижу и думаю, что сегодня по распорядку дня после четвертой пары? Спикер 2: — Сегодня какой день недели? Сегодня четверг. В четверг у нас, по-моему, спортивная масса, работа. Спикер 1: — Давай тогда скажем ребятам, чтобы сегодня сыграли в футбол. Спикер 2: — Давай, если команде разрешить, можно сыграть. Спикер 1: — Кого бы ты хотел взять в команду? Спикер 2: — Я хотел взять в команду Байдербека и Аравова, а на воротах Биг Пау. А почему именно Биг Пау? Биг Пау, потому что он занимался и хорошо стоит на воротах, и у него высокий рост. Спикер 1: — Ну, можно сказать. Но ты же знаешь, что я тоже хорошо играю. И у вас, по-моему, шансов мало будет нас победить. Спикер 2: — Ну, это время пока. Спикер 1: — Все, тогда увидим.	Общее количество слов в оригинале – 112. ChatGPT Количество неправильно распознанных слов – 4 Количество нераспознанных смен спикеров – 7. Количество неверно распознанных пауз – 0. 11 манипуляций. Speech2Text Количество неправильно распознанных слов – 4 Количество нераспознанных смен ролей спикеров – 4. Количество неверно распознанных пауз – 2. 10 манипуляций.

Результаты и их обсуждение.

Рассмотрим результаты аудиофайла 1 (русский язык, диктовка) из таблицы 1. Оригинальный текст состоит из 56 слов. На сервисе ChatGPT были обнаружены 5 ошибочно распознанных слов, например, «казахстаном» → «казахским», «Постановлению» → «Поставновлению», пропущено словосочетание «местного самоуправления». Для исправления текста требуется 5 манипуляций (корректировок ошибок). Результаты анализа на сервисе Speech2Text показывают 7 ошибок. Например, «Постановлению» → «постановлению»; «Конституционного Совета РК» → «Конституционного совета республики Казахстан». Манипуляции: 7 (коррекция ошибок и пунктуации).

Опираясь на вышеприведенные материалы, можно утверждать, что оба сервиса показывают высокую точность (91–93% для ChatGPT, 87–90% для Speech2Text) для русского официального текста. Следует отметить, что ChatGPT в целом лучше справляется с юридической терминологией, сохраняет структуру текста, но допускает ошибки в названиях («казахстаном»). Speech2Text точнее воспроизводит формулировки, но требует больше правок из-за пропусков, неверных падежей, ошибок согласования, нерасслышанных окончаний слов.

Обратим внимание на аудиофайл 2 (английский язык). Оригинальный текст состоит из 107 слов. Оба сервиса (ChatGPT и Speech2Text) допустили по 7 ошибок (93% точности). Примеры ошибок: «Debra» → «Deborah»; «rowed» → «wrote» (в ChatGPT), «rode» (в Speech2Text); «The editor of The Times wrote» → «The editor of the Times Road» (Speech2Text). Манипуляции: 7 для обоих сервисов.

Таким образом, сервисы демонстрируют достаточную точность, но допускают лексические ошибки и теряют смысловые детали, не понимая контекст, что критично для понимания текста, в частности:

- ChatGPT: «wrote» вместо «rowed»;
- Speech2Text: «rode» вместо «rowed», путает имена нарицательные и собственные («Road» вместо «wrote»).

Кроме того, в обоих случаях зафиксированы грамматические ошибки.

В итоге определяется, что оба сервиса не идеальны для английской речи в отношении распознавания имен собственных и сложных глагольных форм.

По поводу транскрипций речи в аудиофайле 3 (на казахском языке) отметим следующее. Оригинальный текст состоит из 319 слов. На сервисе ChatGPT обнаружены ошибки, всего 27 слов. Примеры: «досым» → «жадосым»; «жалпы» → «жалықпайды»; «Ескендір» → «Искендір». Всего потребовалось 27 манипуляций (искажение слов и пунктуации).

На сервисе Speech2Text обнаружены ошибки: 58 слов. Например, «оқып отырсың» → «қолданыпсың»; «Абай Құнанбайдың» → «Абай Хонамба етін»; «поэмасын» → «пайі масын»; почти не совпадает с оригиналом: «Естіп пе ең» → «Есіне сипір». Манипуляции: 58 (коррекция ошибок, искажение, пропуски слов и пунктуации). Как мы видим, оба сервиса при распознавании на казахском языке допустили значительные ошибки. На наш взгляд, это объясняется с ограниченным использованием казахского языка, т.е. казахский язык является малораспространенным языком.

В результате проведенного анализа установлено, что в сервисе ChatGPT почти 89% слов распознается точно, часто встречаются орфографические, грамматические, пунктуационные ошибки. В целом, содержание и логическая структура текста совпадает. Распознанный текст не искажает фундаментальный смысл оригинала. Он является семантически адекватным, несмотря на технические недостатки распознавания. Содержание и логическая структура сохранены, что критически важно для применения ASR в таких сферах, как авиация, где нужно понимать суть сообщения, даже если не все слова распознаны идеально.

По анализу сервиса Speech2Text: 70,5% слов распознается правильно, но имеется много

Әлеуметтік-гуманитарлық ғылымдар-Социально-гуманитарные науки-Social and humanities sciences
ошибок, среди которых есть орфографические, грамматические, синтаксические, пунктуационные. Структура многих предложений искажена. Пропущены слова и фрагменты текста, многие слова неправильно распознаны. Из-за многих ошибок содержание и логическая структура текста не совпадает, транскрипт с трудом понимается и не передает финальный смысловой посыл. Speech2Text в данном случае показал не очень удовлетворительные результаты для казахского языка, что связано с ограниченной поддержкой малораспространённых языков. 70,5% точности — недостаточно для критически важных задач, таких как авиация, где даже потеря одной фразы может быть фатальной.

Рассмотрим аудиофайл 4 (русский язык, диалог). Оригинальный текст состоит из 112 слов. На сервисе ChatGPT были обнаружены ошибки: 4 слова (96% точности). На наш взгляд, проблемы возникли из-за не распознавания 7 смен спикеров, что усложняет восприятие диалога. Бросается в глаза нераспознавание имен собственных: «Биг Пау» → «Бикпаов», «Байдербека» → «Бадербекова», «Аравова» → «Яроал»; манипуляции: 12-13 (коррекция структуры и слов). На сервисе Speech2Text были обнаружены ошибки: 4 слова (96% точности). Трудности возникли из-за нераспознавания 4 смен спикеров и 2 пауз. Сохранил структуру диалога лучше, чем ChatGPT, но не полностью.

Итак, по лексической точности (слова) обе системы показали одинаковый результат: процент искажений: ~3.6%, точность: ~96.4%. ChatGPT передал смысл и паузы, но не разделил спикеров, допустил искажение имен, что является не просто «ошибкой в слове», а потенциально критической ошибкой в авиационной коммуникации. Speech2Text лучше сохранил структуру диалога, но допустил семантическую ошибку («время пока»), что критично в авиации.

На основании изложенного материала можно утверждать, что оба сервиса показывают высокую точность в распознавании слов, но плохо справляются с диалогами: 1) ChatGPT теряет структуру (нет разделения на спикеров) и допускает ошибки в именах собственных, особенно на казахском языке; 2) Speech2Text частично сохраняет структуру, но допускает ошибки в паузах; 3) для диалогов требуется дополнительная настройка сервисов или использование специализированных инструментов для разделения речи по спикерам.

В целом, наши данные о точности автоматического распознавания речи (ASR) в разных языках, акцентах и типах текстов позволяют сделать обоснованные выводы о применимости современных технологий ASR и машинного перевода (МТ) для специфических задач, таких как автоматическое распознавание и перевод радиообмена в авиации – область, где высокая точность, многозначность языковой среды и четкость передачи информации имеют критическое значение.

В исследовании показано, что точность распознавания зависит от языка:

- русский язык: ~ 93% точности (ChatGPT), ~ 87– 90% (Speech2Text);
- английский язык: ~ 93% точности;
- казахский язык: ~78% (ChatGPT), ~ 61% (Speech2Text);

Полученные результаты показывают, что в англоязычной среде точность распознавания выше, что может быть связано с большей распространенностью и количеством обучающих данных. При работе с менее распространенными языками (например, казахским) точность снижается, что указывает на необходимость развития моделей, адаптированных под разные конкретные языки мира.

В авиации для систем автоматического распознавания и перевода авиационного радиообмена необходимо: а) использовать модели, обученные на разнообразных диалектах и акцентах английского; б) разрабатывать специализированные модели для национальных языков, если они могут использоваться в определенных регионах; в) обеспечивать гибкую смену языковых пар в режиме реального времени.

В текстах с диалогами на русском языке (Аудиофайл 4) сервисы показали высокую точность (96%), но не справились с разделением спикеров и имен собственных (например, искажения в

Әлеуметтік-гуманитарлық ғылымдар-Социально-гуманитарные науки-Social and humanities sciences
именах казахского происхождения). В юридических текстах (Аудиофайл 1) ChatGPT лучше сохраняет формальную структуру, тогда как Speech2Text допускает больше ошибок в терминах и пунктуации.

Авиационная речь имеет ряд уникальных характеристик:

- формализованная структура сообщений (например, «Speedbird 456, descend to FL240");
- специализированная терминология (метеоинформация, координаты, инструкции по снижению/набору высоты);
- использование кодов и сокращений (например, QNH, IFR, VFR, etc.);
- четкая дикция и регламентированное произношение (в идеале).

Авиационная связь часто осуществляется через радиочастотные каналы с низкой пропускной способностью, это – средства связи с фоновым шумом (двигатель, турбулентность, помехи) и устройства с различным качеством микрофонов и динамики. Такие факторы влияют на четкость и внятность речи, что напрямую влияет на эффективность ASR.

Таким образом, при низком качестве записи (возможно, имитируемом в тестах) количество ошибок увеличивается. Особенно это заметно в казахском языке, где точность падает до 61%, что может быть связано с недостатком данных и сложностью фонематики. В английском и русском языках ошибки чаще всего связаны с неправильным воспроизведением имен собственных, искажением глагольных форм (например, wrote вместо gowed), потерей пауз и интонаций, важных для понимания структуры предложения.

Если рассмотреть применения в авиации, для автоматического распознавания и перевода языка радиообмена нужны системы, способные распознавать формальные конструкции и стандартизированные фразы, корректно обрабатывать специфические термины и числовые значения (часто критичные для безопасности). А также поддерживать мультимодальное восприятие: сочетание голоса, текстового дисплея и возможной графической интерпретации команд. Помимо того, важно учитывать ограниченное время реакции, данная система должна работать в реальном времени без задержек. Необходимы системы предварительной обработки сигнала: шумоподавление, эквалаизация частот, усиление вокальных элементов. Требуется обучение моделей ASR на авиационных корпусах, включающих записи радиообмена с различными акцентами; ситуации с плохим качеством связи; примеры стандартных фраз и их вариаций; системы должны быть обучены на разговорной, но формализованной речи, чтобы учитывать особенности произношения при сохранении структурной строгости.

Авиационный радиообмен – это диалоговая коммуникация, где каждый запрос и ответ содержат контекст, который должен правильно интерпретироваться системой.

Данное исследование определяет, что оба сервиса плохо справлялись с разделением на спикеров и передачей пауз. В диалогах (Аудиофайл 4) точность распознавания была высока (~96%), но потеря структуры делала текст менее удобным для восприятия.

В случае применения в авиации, для автоматического перевода требуется специальная модель обработки диалогов: выявление начала и конца сообщения; идентификация участников (пилот, диспетчер); сохранение временной последовательности и контекста. Такая специальная модель особенно важна в условиях одновременного общения нескольких участников: прерываний, повторений, коррекций, нестандартных формулировок. Также важно учитывать следующие факторы: интеграция машинного перевода (MT) с системой распознавания речи (ASR); использование конвейерных решений, где выход ASR сразу передается в MT-модель с учетом контекста. Безопасностью и надежностью является критичность ошибок.

В авиации даже одна ошибка в распознавании или переводе может иметь катастрофические последствия. Поэтому системы должны обеспечивать высокую степень достоверности (более 99% точности); механизмы самопроверки и подтверждения; резервирование и возможность ручного ввода. Данное исследование показывает, что даже при высокой точности (до 96%)

Әлеуметтік-гуманитарлық ғылымдар-Социально-гуманитарные науки-Social and humanities sciences
остаются ряд ошибок, требующих редактирования: 5-7 манипуляций на 100 слов. Особенно проблемными оказались имена собственные, глаголы и термины. Это говорит о том, что полностью автоматизировать процесс нельзя без надежной системы проверки.

Для применения в авиации необходимы гибридные системы, сочетающие автоматическое распознавание и перевод, проверяемый человеком (например, дублирование текста на экране диспетчера или пилота); голосовое подтверждение ключевых команд; разработку моделей с уверенностью (confidence score), которые могли бы выделять потенциально ошибочные фрагменты и предлагать альтернативные варианты; интеграцию с системами управления полетом (FMS), чтобы информация из радиообмена могла автоматически попадать в навигационные системы.

На основе анализа можно сформулировать следующие рекомендации. Технологические рекомендации:

1. Обучение мультимодальных моделей на корпусах авиационной речи с акцентами, шумами и формализованными выражениями.
2. Интеграция ASR + MT + NLU (Natural Language Understanding) для глубокой обработки смысловых конструкций.
3. Разработка спикер-специфичных моделей для повышения точности разделения диалогов.
4. Применение контекстуальных моделей (например, на базе Transformer), способных учитывать историю диалога и предсказывать наиболее вероятные команды.

Организационные рекомендации:

1. Создание международных стандартов для автоматизации радиообмена.
2. Сертификация систем искусственного интеллекта в авиационной сфере (по аналогии с сертификацией оборудования).
3. Обучение персонала работе с новыми системами, включая интерпретацию ошибок и резервные методы связи.

Исследовательские рекомендации:

1. Анализ производительности ASR в условиях экстремальных акцентов и фоновых шумов.
2. Изучение влияния автоматического перевода на ситуационную осведомленность экипажа.
3. Разработка метрик оценки точности и надежности систем в реальных условиях.

Заключение

Современные технологии ASR и MT уже достаточно развиты, чтобы применяться в авиационном радиообмене, однако они сталкиваются с рядом серьезных вызовов:

- языковое разнообразие и ограниченная поддержка малораспространенных языков;
- нужда в контекстуальной обработке и учете структуры диалога;
- необходимость работы в условиях шума и низкого качества связи;
- критичность ошибок, требующих дополнительных механизмов проверки.

Как показывает исследование, современные ASR-системы (включая ChatGPT и Speech2Text) еще не готовы к полной автоматизации радиообмена. Даже при точности распознавания выше 95%, ошибки в именах, паузах и структуре диалога делают систему ненадежной без контроля человека. Поэтому автоматизация авиационного радиообмена пока возможна только в гибридном режиме: ASR + человек-оператор, который проверяет ключевые элементы — особенно имена, команды и числовые данные.

Однако эти проблемы преодолимы при условии целенаправленной разработки специализированных моделей, интеграции с существующими авиационными системами и строгом подходе к сертификации. Таким образом, автоматизация авиационного радиообмена с помощью искусственного интеллекта – это не только перспективное, но и реально достижимое

Қ.Жұбанов атындағы Ақтөбе өңірлік университетінің хабаршысы, №3 (81), маусым 2025
Әлеуметтік-гуманитарлық ғылымдар-Социально-гуманитарные науки-Social and humanities sciences
направление, которое может значительно повысить безопасность, эффективность и доступность
международной авиации.

Список литературы

1. Iderson J. C. The politics of aviation English testing. Language Assessment Quarterly, 2011 8(4), 386-403.
2. Automatic Speech Recognition for ATC. Sesar Digital Academy: 11/05/2021
3. Barshi I., Farris C. Misunderstandings in ATC communication. Farnham, UK: Ashgate Publishing Ltd, 2013
4. Covelli J. M., Rolland J. P., Proctor M., Kincaid J. P., Hancock P. A., Field of view effects on pilot performance in flight. International Journal of Aviation Psychology, 2010, 20 (2), 197-219.
5. Eurofighter Typhoon, Technology: Cockpit. 2014 URL: <http://typhoon.starstreak.net/Eurofighter/cockpit.html>. Last accessed: 29 March 2014.
6. IATA, Pilots and air traffic controllers phraseology study. Montreal: International Air Transport Association, 2011
7. ICAO, Manual of radiotelephony Doc 9432 AN/925. Montreal: International Civil Aviation Organization, 2007
8. Jurafsky D., Martin J. H., Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2nd international edition. Upper Saddle River, NJ: Pearson Prentice Hall, 2009
9. LDACS – The future of ATC communications. Canso. 13/01/2023 URL: <https://canso.org/ldacs-the-future-of-atc-communications/#>
10. Шелестюк Е.В., Суюнбаева А.Ж., Торайғыров университетінің Хабаршысы, ISSN 2710-3439 Гуманитарлық сериясы. № 4, 2024.

References

1. Iderson J. C. The politics of aviation English testing. Language Assessment Quarterly, 2011 8(4), 386-403.
2. Automatic Speech Recognition for ATC. Sesar Digital Academy: 11/05/2021
3. Barshi I., Farris C. Misunderstandings in ATC communication. Farnham, UK: Ashgate Publishing Ltd, 2013
4. Covelli J. M., Rolland J. P., Proctor M., Kincaid J. P., Hancock P. A., Field of view effects on pilot performance in flight. International Journal of Aviation Psychology, 2010, 20 (2), 197-219.
5. Eurofighter Typhoon, Technology: Cockpit. 2014 URL: <http://typhoon.starstreak.net/Eurofighter/cockpit.html>. Last accessed: 29 March 2014.
6. IATA, Pilots and air traffic controllers phraseology study. Montreal: International Air Transport Association, 2011
7. ICAO, Manual of radiotelephony Doc 9432 AN/925. Montreal: International Civil Aviation Organization, 2007
8. Jurafsky D., Martin J. H., Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2nd international edition. Upper Saddle River, NJ: Pearson Prentice Hall, 2009
9. LDACS – The future of ATC communications. Canso. 13/01/2023 URL: <https://canso.org/ldacs-the-future-of-atc-communications/#>
10. Şelестюк Е.В., Suiunbaeva A.J. Toraiğyrov universitetiñ Habarşysy, ISSN 2710-3439 Gumanitaryq seriasy. № 4. 2024.

СҮЮНБАЕВА А.Ж.^{1*} , ШЕЛЕСТИЮК Е.В.² 

***Суюнбаева Алтынгүл Жакиповна¹** - Филология ғылымдарының кандидаты, доцент, Әуе қорғанысы күштерінің Әскери институты, Ақтөбе қ., Қазақстан

E-mail: altin_suenbaeva@mail.ru, <https://orcid.org/0000-0002-0676-2028>

Шелестюк Елена Владимировна² - Филология ғылымдарының докторы, профессор, Челябинск мемлекеттік университеті, Челябинск қ., Ресей Федерациясы

E-mail: shelestiuk@yandex.ru, <https://orcid.org/0000-0003-4254-4439>

Аңдатпа. Компьютерлік технологиялардың ілгерілеуі корпуслық мәтіндерді автоматты жүйелеуге, өңдеуге және пайдалануға, яғни арнайы бағдарламалардың көмегімен табиғи коммуникативтік тіл деңгейінде тілді тану процесіне мүмкіндік береді. Мақала ұшқыштар мен авиадиспетчерлер арасындағы түсініспеушілікті азайту үшін қазіргі заманғы технологияларды қолдануға арналады. Зерттеушілер бірнеше ондаған жылдар бойы осы технологияны зерттеп, қолданып келеді. Сөйлеуді автоматты тану (ASR) және машинамен аудару технологияларын енгізу қауіпсіздікті арттырып, авариялар мен оқыс оқиғалардың алдын ала отырып, осындай проблемаларды айтарлықтай төмендетуі мүмкін. Мақалада орыс, ағылшын және қазақ тілдеріндегі мәтіндерді өңдеу кезінде ауызша сөйлеуді тану бойынша екі танымал - ChatGPT (<https://chatgptchatapp.com>) және Speech2Text (<https://speech2text.ru>) бағдарламалардың тиімділігін салыстыру болып табылатын эксперимент жобасы ұсынылған. Жүргізілген зерттеу ауызша сөйлеудің түрлерін қамтыды: мәтінді формальды түрде диктовкалау, заңды мәтін, ана тіліндегі әңгіме, сондай-ақ екі қатысушы арасындағы табиғи (өздігінен) диалог. Талдау транскрипцияның дәлдігі тұрғысынан да, мәтіннің құрылымын жаңғырту, өз есімдерін беру, үзілістерді сақтау және спикерлер бойынша сөйлеуді бөлу ерекшеліктерін ескере отырып жүргізілді. Корпус құрамына «таза» ауызша сөйлеудің аудиожазбасы ғана емес, сондай-ақ телефон арқылы сөйлеу де кіреді, бұл бағдарламаға әмбебаптық береді. Телефондық сөйлесулерді транскрипциялаудың арнайы күрделі әдістемесі әзірленді, өйткені мұндай сөйлеу көбінесе түсініксіз, басқа тілдердегі ақпаратты қамтуы, айқын екпіні болуы, әртүрлі сөйлеу және тілсіз шуды қамтуы мүмкін (әртүрлі дауыс тембрі, сөйлеушінің келесі ойды айтуға дайындығының дыбыстары, түсініксіз сөздер, бірнеше сөйлеушінің сөйлеуі және т.б.). Бұл жүйе бастапқы аудио файлды машинамен оқылатын пішімге түрлендіреді. Яғни, қосымша компоненттер мен транскрипцияларды қажет етпейді. Бұл жүйе интегралдық модельді пайдаланатын сөйлеуді танудың қазіргі заманғы технологияларының нұсқасы болып табылады. Ұсынылған эксперименттік әзірлеме дауыстық командаларды түсінуге мүмкіндік беретін сөйлеуді тану технологиясын пайдаланады, оның ұқсас электрондық қосымшаларға қарағанда артықшылығы - ол төтенше жағдайларда ақпараттың кедергісіз ағынын және оның өңделуін бақылай алады, сондай-ақ табиғи диалогтарды дәл орындау қабілетіне ие.

Түйін сөздер: ASR-MT, формалды диктовка, сөзді автоматты тану, когнитивтік жүктеме, сәйкестілік, конвейерлік шешімдер.

AUTOMATIC SPEECH RECOGNITION AND ASSESSMENT OF ITS APPLICABILITY IN AVIATION AT THE PRESENT STAGE

SUYUNBAEVA A.ZH.^{1*} , SHELESTYUK E.V.² 

***Suyunbaeva Altyngul Zhakipovna¹** - Candidate of philological sciences, associate professor, Military Institute of the Air Defense Forces, Aktobe, Kazakhstan

E-mail: altin_suenbaeva@mail.ru, <https://orcid.org/0000-0002-0676-2028>

Shelestyuk Elena Vladimirovna² - Doctor of Philology, Professor, Chelyabinsk State University, Chelyabinsk, Russian Federation

E-mail: shelestiuk@yandex.ru, <https://orcid.org/0000-0003-4254-4439>

Abstract. The progress of computer technologies opens up opportunities for automatic systematization, processing and use of corpus texts, that is, the process of language recognition at the level of natural communicative speech, using special programs. The article is devoted to the use of modern technologies to reduce misunderstandings between pilots and air traffic controllers. Researchers have been studying and applying this technology for several decades. The introduction of automatic speech recognition (ASR) and machine translation technologies can significantly reduce such problems, improving safety and

Әлеуметтік-гуманитарлық ғылымдар-Социально-гуманитарные науки-Social and humanities sciences preventing accidents and incidents. The article presents a draft experiment, which is a comparison of the effectiveness of two popular programs for recognizing oral speech - ChatGPT (<https://chatgptchatapp.com>) and Speech2Text (<https://speech2text.ru>) - when processing texts in three languages: Russian, English and Kazakh. The study covered different types of oral speech: formalized text dictation, legal text, native language story, and natural (spontaneous) dialogue between the two participants. The analysis was carried out both in terms of transcription accuracy and taking into account the peculiarities of reproducing the structure of the text, transmitting proper names, preserving pauses and dividing speech into speakers. The body includes not only audio recordings of «pure» spoken speech, but also telephone speech, which gives the program versatility. A special complex technique for transcribing telephone conversations has been developed, since such speech is often indistinct, may contain information in other languages, have a pronounced accent, contain various speech and non-speech noises (various voice tones, sounds of preparing the speaker for the next thought, unclear words, superimposition of speech of several speakers, etc.). This system converts the original audio file into a machine-readable format. That is, it does not require additional components and transcriptions. This system is a variant of modern speech recognition technologies using an integrated model. The experimental development presented uses speech recognition technology to understand voice commands, its advantage over similar electronic applications is that it can control the smooth flow of information and its processing in extreme situations, and also has the ability to accurately reproduce natural dialogues.

Key words: ASR-MT, formal dictation, automatic speech recognition, cognitive load, identity, pipeline solutions.