

ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ИДЕНТИФИКАЦИИ ГОЛОСА С УЧЕТОМ КАЗАХСКОГО АКЦЕНТА

^{1,2} Т.С. Шорманов, ² А.Т. Мазакова, ¹ М.С. Алиаскар, ² Ш.А. Джомартова, ^{1,2} Т.Ж. Мазаков ✉

¹Международный инженерно-технологический университет, Алматы, Казахстан,

²Казахский национальный университет имени аль-Фараби, Алматы, Казахстан

✉ Корреспондент-автор: tmazakov@mail.ru

Статья посвящена использованию нейронных сетей для идентификации человека по голосу.

Идентификация и распознавание речи с учетом акцента — это одна из сложных задач в области обработки естественного языка (NLP) и автоматического распознавания речи (ASR). Казахский акцент, как и другие региональные особенности речи, представляет собой уникальные лингвистические и акустические характеристики, которые могут влиять на эффективность традиционных моделей распознавания речи. В последние годы нейронные сети, особенно глубокие нейронные сети (DNN), становятся основными инструментами для решения таких задач, включая идентификацию голоса с учетом акцентов.

Несмотря на достижения в области нейронных сетей, задачи распознавания речи с учетом акцента остаются достаточно сложными. Вот несколько основных проблем: 1) Многообразие акцентов. В Казахстане существует несколько региональных акцентов, что требует учета разнообразных вариантов произношения. 2) Проблемы с трансляцией специфических казахских звуков. Некоторые звуки казахского языка могут быть трудны для стандартных моделей распознавания речи. 3) Отсутствие данных. Как уже было сказано, для качественного обучения нейронных сетей для казахского языка часто не хватает доступных и разнообразных наборов данных с акцентами.

В работе рассматриваются возможные подходы к биометрической идентификации: мел-частотные кепстральные коэффициенты (MFCC), языковая модель, а также анализируются преимущества и недостатки использования нейронных сетей. В ходе исследования выявлены основные проблемы, которые могут возникнуть в ходе ее решения. Дано общее описание проблемы и формулировка целей исследования.

Ключевые слова: биометрия, нейронная сеть, мел-частотные кепстральные коэффициенты, идентификация голоса, казахский акцент.

ҚАЗАҚ ТІЛІНІҢ АКЦЕНТІН ЕСКЕРТПЕ ДАУЫСТЫ АНЫҚТАУ ҮШІН НЕЙРЛІК ЖЕЛІЛЕРДІ ҚОЛДАНУ

^{1,2} Т.С. Шорманов, ² Ә.Т. Мазақова, ¹ М.С. Әлиасқар, ² Ш.А. Джомартова, ^{1,2} Т.Ж. Мазақов ✉

¹Халықаралық инженерлік және технология университеті,

² Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,
e-mail: tmazakov@mail.ru

Мақала адамды дауыс арқылы анықтау үшін нейрондық желілерді пайдалануға арналған.

Екпінді ескеретін сөйлеуді анықтау және тану табиғи тілді өңдеу (NLP) және автоматты түрде сөйлеуді тану (ASR) саласындағы күрделі тапсырмалардың бірі болып табылады. Қазақ екпіні басқа аймақтық сөйлеу ерекшеліктері сияқты дәстүрлі сөйлеуді тану үлгілерінің орындалуына әсер ете алатын бірегей тілдік және акустикалық сипаттарды ұсынады. Соңғы жылдары нейрондық желілер, әсіресе терең нейрондық желілер (DNN), акцентті білетін дауысты сәйкестендіруді қоса алғанда, мұндай тапсырмалар үшін негізгі құралдарға айналды. Нейрондық желілердегі жетістіктерге қарамастан, акцентті білетін сөйлеуді тану тапсырмалары өте қиын болып қала береді. Міне, бірнеше негізгі мәселелер: 1) Акценттердің әртүрлілігі. Қазақстанда бірнеше аймақтық екпіндер бар, олар

айтылудың әртүрлі нұсқаларын қарастыруды талап етеді. 2) Нақты қазақ дыбыстарын таратудағы мәселелер. Кейбір қазақ дыбыстарын стандартты сөйлеуді тану үлгілері үшін түсіну қиын болуы мүмкін. 3) Мәліметтердің жетіспеушілігі. Жоғарыда айтылғандай, қазақ тіліне арналған нейрондық желілерді сапалы дайындау үшін акценттері бар қолжетімді және алуан түрлі деректер жинақтары жиі жеткіліксіз.

Мақалада биометриялық сәйкестендірудің ықтимал тәсілдері қарастырылады: мел-жиілік цестральды коэффициенттер (MFCC), тілдік модель, сондай-ақ нейрондық желілерді пайдаланудың артықшылықтары мен кемшіліктері талданады. Зерттеу барысында оны шешу барысында туындауы мүмкін негізгі проблемалар анықталды. Проблеманың жалпы сипаттамасы және зерттеу мақсаттарының тұжырымы берілген.

Түйін сөздер: биометрия, нейрондық желі, мел-жиілік цестральды коэффициенттер, дауысты анықтау, қазақша акцент.

APPLICATION OF NEURAL NETWORKS FOR VOICE IDENTIFICATION TAKING INTO ACCOUNT THE KAZAKH ACCENT

^{1,2} T.S. Shormanov, ²A.T. Mazakova, ¹M.S. Aliaskar, ²Sh.A. Jomartova, ^{1,2}T.Zh. Mazakov 

¹International Engineering and Technology University, Almaty, Kazakhstan,

²Kazakh National University named after Al-Farabi, Almaty, Kazakhstan,
e-mail: tmazakov@mail.ru

The article is devoted to the use of neural networks for human voice identification.

Accent-aware speech identification and recognition is one of the challenging tasks in the field of natural language processing (NLP) and automatic speech recognition (ASR). The Kazakh accent, like other regional speech features, presents unique linguistic and acoustic characteristics that can affect the performance of traditional speech recognition models. In recent years, neural networks, especially deep neural networks (DNNs), have become the main tools for solving such problems, including accent-aware voice identification.

Despite the advances in neural networks, accent-aware speech recognition tasks remain quite challenging. Here are some of the main challenges: 1) Diversity of accents. There are several regional accents in Kazakhstan, which requires taking into account a variety of pronunciation options. 2) Problems with translating specific Kazakh sounds. Some sounds of the Kazakh language can be difficult for standard speech recognition models. 3) Lack of data. As already mentioned, there is often a lack of accessible and diverse datasets with accents for high-quality training of neural networks for the Kazakh language.

The paper considers possible approaches to biometric identification: mel-frequency cepstral coefficients (MFCC), language model, and analyzes the advantages and disadvantages of using neural networks. The study identified the main problems that may arise in the course of its solution. A general description of the problem and formulation of the research objectives are given..

Keywords: biometrics, neural network, mel-frequency cepstral coefficients, voice identification, Kazakh accent.

Введение. Информационная безопасность является одной из важных задач в мировом сообществе, и это связано с резко возросшим влиянием глобальных информационных технологий на большинство сфер деятельности современного общества. Современные интернет-угрозы требуют эффективных методов обнаружения и защиты данных. Учитывая современную тенденцию к более активному использованию удаленных сервисов при проведении финансовых транзакций, обеспечение кибербезопасности при проведении удаленных финансовых транзакций стало обязательно, так как при проведении удаленных финансовых транзакций появляется возможность компрометации данных, что привлекает киберпреступников. Предлагае-

мое в данной статье решение позволяет повысить эффективность биометрической голосовой аутентификации. Человек может быть аутентифицирован тремя способами: через знание, владение или наследование. Фактор знания относится к тому, что человек должен иметь для получения доступа. Аутентификация на основе предопределенного значения (пароля) является самым распространенным методом аутентификации. Право собственности относится к учетным данным аутентификации пользователя на основе предметов, которыми владеет пользователь, как правило, аппаратных устройств, таких как мобильный телефон пользователя или токен безопасности. С точки зрения внутренних факторов наиболее распространенным методом является аутентификация на основе биометрии, которая использует отпечатки пальцев, голос или распознавание лиц для аутентификации пользователей. Среди этих биометрических методов речь является основным способом общения между людьми. Пример возросшего использования голоса в мае 2021 года HSBC сообщил, что HSBC Voice ID защитил своих клиентов от телефонных мошенников в 2020 году. По данным банка, благодаря новой биометрической системе количество случаев мошенничества сократилось более чем на 50%. В целом HSBC считает, что его система голосовой биометрии предотвратила кражу примерно 249 миллионов фунтов стерлингов (346,5 миллиона долларов США) денег клиентов [1]. Индивидуальные характеристики голоса являются результатом влияния нескольких характеристик: анатомо-физиологических особенностей (особенности строения рта говорящего, например, наличие или отсутствие зубов и т. д.); артикуляционных особенностей говорящего (индивидуальные особенности произношения, например, картавость или шепелявость). Сейчас существуют много сервисов по созданию поддельных голосовых копий, таких как Elevenlabs.io и другие. Исходя из вышеперечисленных фактов, для повышения качества биометрической идентификации по голосу необходимо анализировать не только акустические параметры голоса, но и анализировать и распознавать индивидуальные характеристики речи, такие как

пол, возраст, акцент; такой подход позволит повысить устойчивость системы распознавания к попыткам фальсификации голоса, что особенно важно в сфере финансовых услуг.

Обзор литературы. В статье [2] авторы рассматривают четыре схемы взвешивания терминов и используют десять существующих схем в сочетании с вариантами векторов tf и tf-idf для исследования их производительности на коротких текстовых наборах данных. Этот подход дает хорошие результаты для классификации небольших текстов, таких как сообщения Twitter. Текст в пространстве терминов можно смоделировать как $d=(w_1, \dots, w_n)$, где n обозначает пространство терминов или размер признаков. Эффективность модели часто повышается с использованием N-грамм или схем взвешивания терминов, таких как частота терминов-обратная частота документа. Недостатком метода является, что в коротких текстах таких как твиты или заголовки новостей, содержится ограниченное количество терминов, что делает их представление и взвешивание в модели векторного пространства неэффективными для классификации текста.

В статье [3] представлен подход к использованию модели векторного пространства (VSM), представление текста – это задача преобразования содержимого текстового документа в вектор в пространстве терминов, чтобы документ можно было распознать и классифицировать. Представленный подход к использованию контролируемого метода взвешивания терминов, tf.rf, показывает лучшую производительность, чем другие методы взвешивания терминов. Хотя этот подход имеет некоторые недостатки, он не предназначен для работы с большими наборами данных и методами взвешивания терминов для неструктурированных текстовых данных или использования внешних знаний для улучшения оценки важности терминов.

Статья [4] представляет альтернативный подход, основанный на векторном представлении слов. Преимуществом метода «встраивания слов» (word embeddings) заключается в способности улавливать сходства в классификации значений слов. Точность процедуры оценивается

с помощью кодированных обучающих предложений; валидация проводилась через изучение негативности в речах австрийского парламента. Результаты показали потенциал подхода «встраивания слов» для анализа настроений для небольших текстов, но в больших текстах метод улавливает семантические сходства между словами, но он не может полностью учитывать контекст и порядок слов в предложении. Это ограничивает возможности моделей для анализа настроений и нюансов, особенно для больших текстов.

В статье [5] представлен алгоритм кодирования пар байтов (BPE алгоритм), это способ получения токенов, поскольку ранее языковая модель использовала только слова из заранее представленного словаря, но в представленной статье использовался другой подход к методам сегментации слов, который заключается в разделении слов на более мелкие подслова или слоги. Представленный подход помогает решить проблемы с редко используемыми словами а также с ограничением по лексике, чем использование языковой моделью обученной только на основании слов, которые были в обучающем корпусе. Однако этот подход имеет трудности со словами, которые не встречались в обучающей выборке данных. Это может ограничить практическую применимость моделей машинного перевода, особенно для языков с богатым словарным запасом или с новыми словами и текстами.

В статье [6] представлен алгоритм WordPiece для обучения токенов на разбиение слов на обучающими подслова или слоги с добавлением специальных символов в начале каждого токена, чтобы показать, что это часть большого слова. Этот процесс помогает анализировать отдельные символы и контекст больших слов, данный подход позволяет лучше обобщать новые слова и уменьшить размер обучающей выборки слов. Алгоритм может быть применен к различным языкам, что делает его универсальным для многоязычных приложений. Алгоритм WordPiece имеет недостаток связанный со значительными вычислительными требованиями, это может быть ограничивающим фактором для практического

использования, особенно для задач реального времени или очень больших текстов.

В статье [7] BERT представлен как многослойный двунаправленный кодер Transformer. Эта языковая модель, основанная на архитектуре трансформера, использует двунаправленный внутренний анализ. Модель используется совместно с классификатором, входные данные которого являются результатом работы BERT – векторным представлением входных данных. В основе обучения модели лежат две идеи. Первая – заменить 15 % слов масками и обучить сеть предсказывать эти слова. Вторая – дополнительно обучить BERT определять, может ли одно предложение следовать за другим. Архитектура Transformer состоит из различных компонентов, которые были описаны в [5]: Токенизаторы, которые преобразуют текст в токены. – один слой внедрения, который преобразует токены и позиции токенов в векторные представления. – слои преобразователя, которые выполняют повторяющиеся преобразования векторных представлений, извлекая все больше и больше лингвистической информации. Они состоят из чередующихся слоев внимания и прямой связи. – (необязательно) слой деинсталляции, который преобразует окончательные векторные представления обратно в распределение вероятностей по токенам. Предварительно обученные модели, такие как BERT, имеют ограничения, связанные с обучением на огромных корпусах данных, который могут содержать смещения и ограничения, отражающие данные, на которых они обучались. Это может привести к тому, что модели примут те же смещения и ограничения. Пример таких смещений и ограничений представлен в статье [8] как результат обучения многоязычной языковой модели BERT, анализ для греческого и испанского языков и обнаружено, что в некоторых случаях языковая модель предпочитает использовать более англоязычную настройку (явные местоимения и порядок субъект-глагол-объект) по сравнению с одноязычной моделью языка управления. Это исследование показывает, что многоязычная модель BERT имеет недостатки, связанные с доминированием английского язы-

ка, поскольку модель изучалась на данных из нескольких языков, но данные из английского языка доминировали, и это повлияло на результаты на других языках.

Для решения вопросов, связанных с голосовой идентификацией, необходимо использовать не только акустический голос для идентификации, а проводить распознавание речи с учетом особенностей конкретного человека. Помимо физиологических особенностей голоса существует фонетическая особенность речи, которая указывает на региональную, этническую или иноязычную принадлежность человека. Каждый человек обладает определенным словарным запасом, этот словарный запас определяется его социальной и ментальной средой. Особенности речи, голоса, интонации, а также манера речи, сформированные в юности, сохраняются на протяжении всей жизни и имеют набор определенных характеристик, присущих только им [9]. Анализируя отдельные элементы речи, можно определить индивидуальный стиль речи человека. Однако такой подход имеет ряд недостатков, для анализа индивидуальных голосовых характеристик речи необходимо собрать достаточно большую базу данных речи конкретного человека.

Поскольку в Республике Казахстан два основных языка общения – казахский и русский, то при разработке необходимо учитывать особенности и акценты, используемые в Республике Казахстан. Акцент – фонетическая особенность речи, связанная с региональной, этнической или иноязычной принадлежностью человека. Человек с акцентом может неправильно произносить некоторые звуки или изменять их непосредственно в словах. Особенности казахского акцента в русском языке. Акцент – это изменение произношения звуков, распространенное искажение. Акцент может привести к замене или искажению определенных звуков. К таким звукам относятся, например, «ы» и «и» в русском языке. Еще одной распространенной проблемой является использование звука «у», так как его произношение в казахском языке отличается. Интонация – еще один признак акцента. Казахский язык имеет свою собственную интонацию, которая меняется

в зависимости от культурного контекста, и эти интонации появляются в других языках. Ударение в русском языке – это выделение гласного звука в слове; оно может быть как в середине слова, так и в конце; в казахском ударении чаще всего используется ударение на конце слова, по аналогии с казахским языком. Фонетические особенности ударения: измененное произношение звуков; часто изменения в произношении звуков; в первую очередь, это безударные гласные после твердых шипящих «ш», «ж» и «ц». Использование собственных казахских звуков. В казахском языке есть специфические звуки, которых нет в русском языке, это 9 звуков, которые часто используются в качестве замен русских звуков и существенно влияют на произношение [10]. Таким образом, целью данного исследования является совершенствование способов анализа коротких аудиозаписей говорящего с целью выявления наличия казахского акцента для повышения качества биометрической аутентификации.

Материалы и методы. При работе использованы аудиозаписи. При проведении исследования в статье использованы методы машинного обучения, основанный на применении нейронных сетей.

Результаты и обсуждение. Для решения данной задачи предлагается использовать двух-этапный процесс анализа аудиозаписей с целью выявления наличия акцентов. Первый этап получение акустической модели говорящей персоны, второй этап получение языковой модели говорящей персоны.

Первый этап получения голосовой модели – получение голосовой модели, на основании анализа вышперечисленной литературы наиболее оптимальный метод для получения голосовой информации является Мел-частотные кепстральные коэффициенты (MFCC). Это метод, который позволяет анализировать частотный контент аудиозаписи путем разложения звука на мел-частотные компоненты.

Таким образом, можно анализировать как частотные, так и временные характеристики звука. Акустическая модель – преобразование звука

в удобный для использования формат Мел-частотные кепстральные коэффициенты (MFCC) являются одним из наиболее распространенных и эффективных способов представления речевых данных. Этот подход с использованием мел-кепстральных коэффициентов был представлен в работе [11]. Поскольку разница в тембрах разных голосов описывается различными частотными спектрами. Математический аппарат для анализа частотного спектра – это преобразование Фурье, как способ описания сложной звуковой волны с помощью спектрограммы. Обработка данных, представляющих числовые амплитудно-временные зависимости. Учитывая особенности человеческого слуха, а именно его нелинейную природу по отношению к восприятию звуковых частот, иными словами, люди гораздо лучше различают небольшие изменения высоты звука на низких частотах, чем на высоких.

Для этой задачи используется преобразование из шкалы Герца в шкалу мел (мел-психофизическая единица высоты тона). Для нашей задачи [12] идентификации человека по голосу очень важен охват всех частот, поскольку, как было сказано во введении, для задач идентификации очень важно выявить все характеристики человека, связанные с вокальностью и высотой тона речи конкретного человека. Также преимуществом алгоритма MFCC стало то, что он уже имеет реализацию во многих языках программирования.

Второй этап: Получение языковой модели. Для обучения модели языка целесообразно использовать модель языка BERT. Модель языка BERT была представлена [13] на основе архитектуры transformer и предназначена для предобучения языковых представлений (для задач NLP). Главной особенностью этой модели языка является ее двунаправленность; эта модель языка также была доработана и может послужить основой для других алгоритмов. BERT использует четыре основных типа специальных токенов для обозначения начала и конца предложения или продолжения.

- 1) [CLS]: обозначает начало предложения.
- 2) [SEP]: обозначает конец предложения.

3) [UNK]: обозначает неизвестную часть/слово в предложении.

4) [PAD]: используется для заполнения.

Ввод данных состоит из трех этапов:

- 1) Реализации токенов
- 2) Реализация сегментов
- 3) Реализация позиционирования.

Токенизатор подслов BERT: BERT использует токенизатор подслов, который разбивает слова на более мелкие единицы, чем слова (слоги). Токенизатор, используемый BERT, – это токенизатор WordPiece, который представляет собой алгоритм, процесс создания окончательного набора слов (словаря) путем слияния слогов (подслов) из букв.

Обучение BERT проводится в два этапа: этапы предварительной подготовки и тонкой настройки. Этап предварительной подготовки выполняется одновременно по двум задачам: Next Sentence Piece (NSP) и Masked Language Model (MLM). Masked Language Model (MLM) Для предварительной подготовки BERT случайным образом маскирует 15 % слов из входного текста с помощью токена [MASK]. Этот подход используется только для предварительной подготовки, а не для тонкой настройки. Выполняя задачу сопоставления токенов [MASK], BERT развивает способность понимать контекст.

Next Sentence Piece (NSP) После отправки двух предложений BERT обучается на них, сравнивая, является ли данное предложение одним непрерывным предложением или нет. Для этого BERT использует два фактически связанных предложения и два случайно связанных предложения в соотношении 50:50. Эта маркировка использует специальный токен [CLS], который помещается в начало предложения, чтобы определить, являются ли два предложения фактически последовательными предложениями.

Тонкая настройка BERT – это этап тестирования уже предварительно обученного BERT путем его дополнительного обучения для заданных задач, которые мы хотим решить. Для наших задач, тонкая настройка языковой модели BERT

включает использование готовых уже обученных моделей BERT на основе русского языка с расширением словаря с учетом казахских слов и казахского акцента. Для анализа мы сравниваем две языковые модели BERTmultilingual и kazakhBERTmulti [14] для целей распознавания речи на 15 коротких аудиозаписей. По результатам анализа для определения казахского акцента для более точной оценки необходимо исключить проверку орфографии, так как казахский акцент

в основном проявляется в виде неправильно произнесенных слов. Для определения акцента для наших задач мы убрали проверку на орфографические ошибки, а именно неправильно произнесенные слова, чем больше таких слов было выявлено, тем более выраженным был акцент в аудиозаписи. Результаты, полученные с многоязычной моделью BERT и моделью kazakhBERTmulti, приведены в таблице 1.

Таблица 1- Сравнение результатов

Model	Accuracy	Recall	Precision	F1-score
BERTmultilingual	0.92	0.77	0.85	0.64
kazakhBERTmulti	0.92	0.84	0.92	0.68

Для multilingual BERT Russian F1 Score равен 0.64, для kazakhBERTmulti F1 Score равен 0.68 что показывает, что языковая модель работает достаточно хорошо на тестовой выборке.

Преимущество. Предлагаемый двухэтапный подход к извлечению голосовых признаков с помощью сбора аудиоданных MFCC представляет собой набор аудиопризнаков, извлеченных путем математической обработки частот. Это обеспечивает удобство входящего сигнала с потерей неважной информации. С их помощью создается упрощенная модель звуков речи, она также похожа на подслова или слоги, которые используются для языковых моделей BERT.

Недостатки. Использованный подход показывает приемлемые результаты по распознаванию для коротких аудиозаписей, использованием данного подхода для анализа больших аудио записей требует использование другим алгоритмов. Ограничения голосовой модели. Алгоритм MFCC плохо работает с тоновыми языками,

а особенностью казахского акцента является то, что ударения расставляются по-другому, в отличие от русского языка, и такие особенности плохо различаются в выходных файлах.

Выводы. Использование предложенного метода анализа позволяет получить четкую информацию о тексте речи. Все аудиозаписи были правильно распознаны для одного голоса. Применение нейронных сетей для биометрической идентификации по голосу дает значительное преимущество, поскольку позволяет проводить идентификацию не только физиологической характеристики такой как голос человека, но и дополнительно проверять индивидуальные голосовые поведенческие характеристики.

Финансирование. Работа выполнена за счет средств НИИ математики и механики при КазНУ имени аль-Фараби и грантового финансирования научных исследований на 2023–2025 годы по проекту AP19678157.

Литература

- 1.HSBC' s Voice ID prevents £249 million of attempted fraud. [Электронный ресурс] URL: <https://www.finextra.com>
2. Samant S.S., Bhanu Murthy N.L., Malapati A. Improving term weighting schemes for short text classification in vector space model.//IEEE Access.-2019.- Vol.7. - P.166578 -166592.

DOI 10.1109/ACCESS.2019.2953918

3. Lan M., Tan C.L., Su J., et al. Supervised and traditional term weighting methods for automatic text categorization. // IEEE Trans Pattern Anal Mach Intell.- 2009. - Vol.31(4). - P.721 - 735.

DOI 10.1109/TPAMI.2008.110

4. Rudkowsky E., Haselmayer M., Wastian M., Jenny M., Emrich Š., & Sedlmair M. More than Bags of Words: Sentiment Analysis with Word Embeddings. // Communication Methods and Measures. – 2018.- Vol. 12(5).- P.140 - 157. DOI 10.1080/19312458.2018.1455817

5. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units. // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany. Association for Computational Linguistics.- 2016.-Vol.1.- P.1715-1725. DOI 10.18653/v1/P16-1162

6. Wu Y., Schuster M., Chen Z., Le Q. V., Norouzi M., Macherey W., ... & Dean J. Google' s neural machine translation system: Bridging the gap between human and machine translation. // arXiv preprint arXiv:1609.08144.2016. DOI 10.48550/arXiv.1609.08144

7. Jacob D., Chang M., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. // North American Chapter of the Association for Computational Linguistics. - 2019. - P. 4171-4186. DOI 10.18653/v1/N19-1423

8. Papadimitriou I., Lopez K., Jurafsky D. Multilingual BERT has an accent: Evaluating English influences on fluency in multilingual models. // arXiv preprint arXiv:2210.05619.- 2023.- P.1194-1200. DOI 10.18653/v1/2023.findings-eacl.89

9. Россинская Е.Р. Профессия – эксперт. Введение в юридическую специальность. – М.: Юрист, 1999. - 192 с. Rossinskaya, E. (1999). Profesiyy – expert. Vvedenie v uridicheskuu spechialnoct. <http://lawlibrary.ru>. [in Russian]

10. Основные черты казахского акцента в русской речи в области вокализма и суперсегментной фонетики. [Электронный ресурс] URL: <https://rcdo.kz>. [in Russian]

11. Davis S., Mermelstein P. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences // IEEE Transactions on Acoustics, Speech, and Signal Processing. -1980. -Vol. 28, N4.- P. 357- 366. DOI 10.1109/TASSP.1980.1163420

12. Aliaskar M., Mazakov T., Mazakova A., Jomartova S., Shormanov T. Human voice identification based on the detection of fundamental harmonics // IEEE 7th International Energy Conference (ENERGYCON). - 2022.- P. 1-4. DOI 10.1109/ENERGYCON53164.2022.9830471

13. Li F., Jin Y., Liu W., et al. Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: an empirical study.// JMIR Med Inform. - 2019. -Vol. 7(3) - P.1-13. DOI 10.2196/14830

14. Mazakov T.Zh., Jomartova Sh. A., Shormanov T. S., Ziyatbekova G. Z., Amirkhanov B. S., Kisala P . The image processing algorithms for biometric identification // News of the national academy of sciences of the Republic of Kazakhstan. Series of Geology and Technical Sciences.- 2020.-Vol.1(439). – P.14-22. DOI 10.32014/2020.2518-170X.2

Сведения об авторах

Шорманов Т.С. – докторант КазНУ имени аль-Фараби, старший преподаватель МИТУ, Алматы, Казахстан, e-mail: shormanov@gmail.com;

Мазакова А.Т. – докторант КазНУ им.аль-Фараби, e-mail: aigerym97@mail.ru;

Әлиасқар М.С. - старший преподаватель МИТУ, Алматы, Казахстан, e-mail: m.alyasqar@gmail.ru;

Джомартова Ш.А. - доктор технических наук, доцент, КазНУ им. аль-Фараби, Алматы, Казахстан, e-mail: jomartova@mail.ru;
Мазакон Т.Ж. – доктор физико-математических наук, профессор, КазНУ им. аль-Фараби, Алматы, Казахстан, e-mail: tmazakov@mail.ru.

Information about the authors

Shormanov T.S. - PhD student of the Al-Farabi Kazakh National University, Lecturer at the International University of Engineering and Technology, Almaty, Kazakhstan, e-mail: shormanov@gmail.com;

Mazakova A.T. - PhD student of the Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: aigerym97@mail.ru;

Aliaskar M.S. - Lecturer at the International University of Engineering and Technology, Almaty, Kazakhstan,
e-mail:m.alyasqar@gmail.ru;

Jomartova Sh.A. - Doctor of Technical Sciences, Associate Professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan,
e-mail: jomartova@mail.ru;

Mazakov T.Zh. – Doctor of Physical and mathematical sciences, professor, Al-Farabi Kazakh National University, Almaty, Kazakhstan, e-mail: tmazakov@mail.ru