УДК 004.522

СОВРЕМЕННЫЕ ТЕНДЕНЦИИ В РАСПОЗНАВАНИЯ РЕЧИ

М. Т. Мырадов
Заведующий кафедры «Информационные системы» Институт Телекоммуникаций и информатики Туркменистана,
maksat.myradow.92@mail.ru

Р. Б. Хыдыров
Начальник Научного отдела института Инженерно-технических и транспортных коммуникаций Туркменистана,
hyd.row@yandex.ru

М. Т. Мырадов
Заведующий кафедры «Информационные системы» Институт Телекоммуникаций и информатики Туркменистана.

Р. Б. Хыдыров
Начальник Научного отдела института Инженерно-технических и транспортных коммуникаций Туркменистана

Аннотация. Распознавание речи (*speech recognition*) – это процесс преобразования аудио сигнала, содержащего речь, в текстовую форму. Эта технология используется для различных целей, включая автоматическое распознавание команд в голосовых интерфейсах, транскрибирование аудиозаписей, преобразование речи в текст в системах диктовки и многое другое.

Ключевые слова: распознавания речи, преобразования аудио сигнала, аудио сигнал, *BIG DATA* в распознавании речи, цифровая обработка аудиосигнала.

Введение. Распознавание речи (*speech recognition*) – это процесс преобразования аудио сигнала, содержащего речь, в текстовую форму. Эта технология используется для различных целей, включая автоматическое распознавание команд в голосовых интерфейсах, транскрибирование аудиозаписей, преобразование речи в текст в системах диктовки и многое другое.

Распознавание речи основывается на алгоритмах машинного обучения и обработки сигналов. Сначала аудио сигнал разбивается на небольшие фрагменты, называемые фреймами. Затем фреймы анализируются с помощью методов обработки сигналов, например, выделение характеристик (например, спектральных коэффициентов) и алгоритмов классификации (например, скрытой марковской модели или нейронных сетей), чтобы определить, какие звуки или слова присутствуют в каждом фрейме.

После этого, преобразованные фреймы объединяются и используются для построения последовательности слов, которая представляет собой распознанный текст. Этот текст может быть дальше использован для различных задач, таких как автоматическая транскрипция, поиск информации или управление устройствами с помощью голосовых команд [1].

Сегодня сформировались 4 основных направления, в которых технология распознавания речи с машинным обучением смогла себя проявить:

1 Распознавание для систем голосового обслуживания и интерактивных автоответчиков. Они распространены в колл-центрах, сервисах самообслуживания, онлайн-банкинге. К их приветствиям и голосовым меню уже давно все привыкли.

Распознавание речи играет важную роль в системах голосового обслуживания и интерактивных автоответчиках. Оно позволяет компьютерной программе "понимать" и интерпретировать речь, произносимую пользователем, и предоставлять соответствующие ответы или выполнять требуемые действия.

Существует несколько типов технологий распознавания речи, которые применяются в таких системах:

1 Статистическое распознавание речи (*Statistical Speech Recognition*): Эта технология основана на использовании больших наборов данных для обучения моделей распознавания речи. Алгоритмы статистического распознавания речи анализируют акустические характеристики и последовательности звуков, чтобы определить, какое слово или фраза были произнесены.

2 Глубокое обучение (*Deep Learning*): Глубокое обучение – это раздел машинного обучения, основанный на моделях нейронных сетей с множеством слоев. Эти модели способны изучать и идентифицировать сложные образцы или шаблоны в данных, включая речь и звуковые сигналы. Технологии глубокого обучения все чаще используются для распознавания речи в системах голосового обслуживания.

3 Комбинированные модели: Некоторые системы используют комбинацию различных технологий распознавания речи для повышения точности и производительности. Например, можно использовать статистическое распознавание речи для предварительной обработки звуковых данных, а затем применить глубокое обучение для более точного и детализированного распознавания.

Однако, несмотря на значительные достижения в области распознавания речи, оно все еще имеет свои ограничения. Например, шумные и плохо записанные аудиофайлы могут затруднить точное распознавание речи. Кроме того, характеристики и акценты говорящего могут создавать сложности для систем распознавания.

В целом, технологии распознавания речи продолжают развиваться и улучшаться, что позволяет системам голосового обслуживания и интерактивным автоответчикам становиться все более точными и эффективными в общении с пользователями.

2 Распознавание и идентификация по голосу. Крупные банки используют его для идентификации клиентов по голосовому отпечатку, для голосовой подписи, а также в системах безопасности.

Распознавание и идентификация по голосу – это технологии, которые позволяют определить и проверить личность человека на основе его голоса.

Распознавание по голосу использует алгоритмы и модели машинного обучения для анализа уникальных характеристик голоса, таких как тембр, частота, интонация и ритм. Эти характеристики помогают создать уникальный голосовой профиль для каждого человека [2].

Идентификация по голосу осуществляется путем сравнения голосового профиля с заранее сохраненными данными в базе данных. Если голосовой профиль совпадает с профилем в базе данных, то система идентифицирует личность.

Технологии распознавания и идентификации по голосу широко используются в различных областях, включая банковское дело, безопасность, телефонию и медицину. Например, системы голосового банкинга позволяют клиентам осуществлять операции по телефону, используя только свой голос для идентификации. Также голосовые системы безопасности могут использоваться для контроля доступа в здания или для аутентификации пользователей в компьютерных системах.

Однако, стоит отметить, что распознавание и идентификация по голосу не являются 100% надежными и могут быть обмануты при использовании поддельного голоса или при наличии физических изменений голоса у человека.

3 Речевая аналитика звонков и переговоров. Предназначена для оценки отзывов и удовлетворённости клиентов, повышения качества работы операторов, выявления трендов при обращениях в службы поддержки и отделы продаж.

Речевая аналитика звонков и переговоров – это процесс анализа и интерпретации речи в телефонных разговорах и переговорах с помощью технологий обработки естественного языка и машинного обучения.

Речевая аналитика звонков может использоваться для различных целей, таких как контроль качества обслуживания клиентов, обнаружение мошенничества, анализ эмоционального состояния или настроения собеседника, анализ эффективности продаж и других бизнес-процессов.

Технологии речевой аналитики звонков включают в себя распознавание речи, извлечение ключевых слов и фраз, определение тональности и эмоционального состояния говорящего, классификацию и категоризацию разговоров, а также построение статистических моделей и прогнозов на основе данных из разговоров.

Речевая аналитика звонков может быть полезной для компаний, чтобы улучшить качество обслуживания клиентов, повысить эффективность продаж, выявить проблемы в бизнес-процессах и принять меры для их улучшения. Она также может быть использована в области безопасности для обнаружения мошенничества или незаконной деятельности.

Однако, при использовании речевой аналитики звонков следует учитывать проблемы конфиденциальности и соблюдения законодательства о защите персональных данных. Компании должны обеспечить соответствие своей деятельности требованиям закона и уведомлять своих клиентов о том, что их разговоры могут быть записаны и проанализированы.

4 Голосовое управление. Применяется во многих сферах, например: в быту для управления «умным» домом, электронными приборами, даже имейлом и браузерами; в автопромышленности для привычных навигаторов, а скоро и для управления беспилотным автотранспортом.

Голосовое управление – это технология, которая позволяет людям взаимодействовать с компьютерами, устройствами и системами с помощью голосовых команд. Она основана на распознавании и обработке речи с целью выполнения определенных задач или операций.

Голосовое управление может быть использовано в различных сферах, включая домашнюю автоматизацию, мобильные устройства, автомобили, медицину, образование и бизнес. С помощью голосового управления люди могут выполнять такие задачи, как отправка сообщений, поиск информации в Интернете, управление устройствами в доме (например, светом, температурой), навигация по маршруту в автомобиле и многое другое.

Основными технологиями голосового управления являются распознавание и синтез речи. Распознавание речи позволяет компьютерам интерпретировать голосовые команды и преобразовывать их в текст или операции. Синтез речи, в свою очередь, позволяет компьютерам генерировать голосовые ответы или инструкции для пользователя.

Голосовое управление имеет ряд преимуществ, таких как удобство использования, повышение эффективности и производительности, а также доступность для людей с ограниченными возможностями. Однако, оно также имеет свои ограничения, такие как возможность неправильного распознавания речи или недостаточной точности в определенных условиях (например, шумном окружении).

В целом, голосовое управление является одной из важных технологий, которая продолжает развиваться и находить все большее применение в различных областях нашей жизни.

Существует несколько новых технологий, которые используются для распознавания речи:

1 Глубокое обучение: Применение глубокого обучения позволяет создавать более точные модели распознавания речи. Глубокие нейронные сети могут изучать сложные шаблоны и обнаруживать скрытые закономерности в речевых данных, что повышает качество и точность распознавания.

2 Рекуррентные нейронные сети (*RNN*): *RNN* – это тип нейронной сети, который способен обрабатывать и анализировать последовательные данные, такие как речь. Они сохраняют информацию о предыдущих состояниях, что помогает учесть контекст и зависимости между звуками и словами.

3. Рекуррентные сверточные нейронные сети (*CRNN*): *CRNN* являются комбинацией сверточных и рекуррентных нейронных сетей. Они позволяют совмещать преимущества сверточных слоев, которые помогают распознавать низкоуровневые аудиофункции, с возможностью учета контекста и последовательности, предоставляемой рекуррентными слоями.

4. Методы передаточного обучения: Передаточное обучение позволяет использовать предварительно обученные модели на больших исходных данных для улучшения производительности распознавания речи. Модели, обученные на больших наборах данных, могут быть использованы для инициализации новых моделей и последующего дообучения на более специфических данных.

5. Техники улучшения качества данных: Дополнительные техники, такие как улучшение качества аудио, устранение шума и реверберации, могут помочь улучшить точность распознавания речи. Это может включать применение алгоритмов шумоподавления, цифровой фильтрации и других методов обработки сигнала.

Эти новые технологии и методы способствуют значительному улучшению производительности систем распознавания речи и широкому применению в различных областях, включая автоматический диктовки текста, разговорные ассистенты и системы управления голосом.

BIG DATA в распознавании речи является важной исследовательской областью, которая имеет огромный потенциал для улучшения точности и эффективности систем распознавания речи. С использованием больших объемов данных, мы можем применять анализ данных и различные алгоритмы машинного обучения, чтобы достичь высокой степени распознавания и понимания человеческой речи. Это имеет широкий спектр применений, включая разработку голосовых помощников, создание систем автоматического распознавания речи, определение настроения и эмоций из речи, а также многое другое.

Например, благодаря большим объемам данных, мы можем обучить систему распознавать различные акценты, диалекты и речевые особенности, что делает ее более гибкой и адаптивной к различным пользовательским потребностям. Это особенно полезно для разработки голосовых помощников, которые могут работать с людьми разных культур и национальностей. Также, большие объемы данных способствуют улучшению процесса обучения системы, что позволяет снизить количество ошибок и создать более надежные и точные системы распознавания речи.

Одним из интересных применений *BIG DATA* в распознавании речи является определение настроения и эмоций из речи. Благодаря анализу больших объемов данных, можно выявить характеристики и особенности речи, которые свидетельствуют о

настроении человека. Это может быть полезно в различных областях, включая маркетинг, клиентское обслуживание и психологию.

В целом, *BIG DATA* в распознавании речи открывает множество возможностей для улучшения коммуникации и взаимодействия с помощью голосовых технологий. Большие объемы данных позволяют создавать более интеллектуальные и адаптивные системы, которые способны лучше понимать и отвечать на потребности пользователей. С каждым днем мы узнаем все больше о том, как использование *BIG DATA* может преобразить наш опыт работы с голосовыми интерфейсами и распознаванием речи.

Цифровая обработка аудиосигнала: проблемы и решения. Цифровая обработка аудиосигналов привела к существенным изменениям в области записи и управления звуком. Однако, несмотря на все преимущества цифрового звука, существуют некоторые проблемы, связанные с чрезмерным усилением аудиосигнала в цифровой среде. В этой статье мы рассмотрим причины чрезмерного шума и возможные пути решения этой проблемы.

Ограничение, также известное как ограничение, происходит, когда амплитуда аудиосигнала слишком высока для отображения на цифровом носителе. При этом пиковый уровень сигнала отсекается, что приводит к искажению звука и потере динамического диапазона. Эта проблема часто возникает при записи и мастеринге музыкальных композиций.

Могут быть различные причины усиления аудиосигнала в цифровой среде. Одна из них – неправильная настройка уровней записи. Если уровень записи слишком высок, сигнал может превысить максимально допустимое значение и может произойти клиппирование. Это также может произойти во время микширования и мастеринга, если аудиосигнал не обработан должным образом.

В цифровой среде существует несколько подходов к решению проблемы передискретизации аудиосигнала. Во-первых, вам необходимо правильно настроить уровни записи. Ключевым моментом здесь является достижение баланса между достаточно высокими уровнями сигнала, чтобы сохранить детализацию, и достаточно низкими уровнями, чтобы избежать клиппирования [1].

Во-вторых, вы можете использовать специальные алгоритмы сжатия и ограничения, которые позволяют контролировать динамический диапазон аудиосигнала. Сжатие позволяет уменьшить разницу между самым громким и самым тихим звуками, а лимитирование не позволяет сигналу превысить максимально допустимый уровень.

Также важно учитывать характеристики конкретного цифрового формата, используемого для записи и воспроизведения аудиосигнала. Некоторые форматы могут иметь ограничения на максимальный уровень сигнала, поэтому выбирайте формат, учитывающий необходимый вам динамический диапазон.

Следует отметить, что выбросы аудиосигнала могут быть нежелательным явлением в цифровой среде. Это особенно важно в таких профессиональных областях, как студийная запись и режиссура звука, где качество звука имеет решающее значение.

Туркменская буква А отличается от других букв тем, что она обозначает звук [a], который является открытым передним гласным. В туркменском языке этот звук может быть долгим или коротким, в зависимости от положения в слове и наличия ударения. Например, в слове алма (яблоко) буква А обозначает долгий звук [a:], а в слове ат (лошадь) – короткий звук [a]. Также, туркменская буква А может быть частью диграфа Аа, который обозначает звук [ɑ], который является открытым задним гласным. Например, в слове аалам (мир) буква Аа обозначает звук [ɑ]

В разные периоды истории туркменского языка буква А писалась по-разному. В арабской письменности, которая использовалась до 1928 года, буква А писалась с помощью арабской буквы алиф. В латинице-яналифе, которая использовалась с 1928 по

1940 год, буква А писалась так же, как и в современной латинице. В кириллице, которая использовалась с 1940 по 1993 год, буква А писалась так же, как и в русском алфавите. В современной латинице, которая используется с 1993 года по настоящее время, буква А пишется так же, как и в латинице-яналифе

Туркменская буква А отличается от русской буквы А тем, что она может быть частью диграфа Аа, который обозначает звук [ɑ], который является открытым задним гласным. Например, в слове аалам (мир) буква Аа обозначает звук [ɑ]. Также, туркменская буква А может быть надстрочным знаком «циркумфлекс» (ˆ), который служит для обозначения долготы гласных в словах арабского и персидского происхождения. В русском языке таких знаков нет.

Русская буква А отличается от туркменской буквы А тем, что она всегда обозначает звук [а], который является открытым передним гласным. В русском языке этот звук может быть ударным или безударным, в зависимости от положения в слове. Например, в слове ра́зный буква А обозначает ударный звук [*а*], а в слове ро́знь — безударный звук [*а*]. Также, русская буква А может быть частью разных буквосочетаний, таких как ау, ао, аэ, которые обозначают соответствующие дифтонги. Например, в слове авто́бус буква А обозначает дифтонг [*аu*].

Туркменская модель распознавания голоса отличается от других тем, что она специально адаптирована для туркменского языка, который имеет свои особенности фонетики, грамматики и лексики. Также туркменская модель учитывает различные диалекты и акценты, которые могут встречаться в речи носителей туркменского языка. Туркменская модель использует современные методы машинного обучения и обработки естественного языка, чтобы достичь высокой точности и скорости распознавания голоса. Туркменская модель может применяться в разных сферах, таких как образование, здравоохранение, бизнес, туризм и др.

Другие модели распознавания голоса могут отличаться от туркменской по ряду параметров, таких как:

– Язык или языки, которые они поддерживают. Некоторые модели могут распознавать только один язык, другие могут распознавать несколько языков или даже переводить речь с одного языка на другой.

– Методы и алгоритмы распознавания речи, которые они используют. Некоторые модели могут основываться на статистических или нейронных сетях, другие могут использовать правила или шаблоны.

– Архитектура системы распознавания речи, которая определяет, как происходит обработка и передача речевого сигнала. Некоторые модели могут работать в режиме онлайн или офлайн, другие могут работать на локальном устройстве или на удаленном сервере.

– Приложения и цели, для которых они созданы. Некоторые модели могут быть специализированы для определенных задач или доменов, другие могут быть универсальными или адаптивными.

Теперь посмотрим на общение туркменского языка с другими языками:

1 Galyň

[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.8372901e-06, 0.0021463048, -0.00065699883, 0.0011694628, 0.0005780997, 0.00090964924, 0.0023731706, 0.36976808, -0.09002273, -0.18889533, 0.013359557, 0.029345099, -0.07667064, -0.022750532, -0.09647286, 0.01741997, 0.006388563, -0.009117429, 0.0011819443, -0.0041174744, -0.001189196, -0.00086708, -0.0026997093, 0.0023305258, -0.0013775838, 0.00049126043, 0.00048636287,

-0.0005392069, -0.0005627414, 0.001556347, -0.0026535941, 0.00065200834, 0.0008123335, -0.003130231]

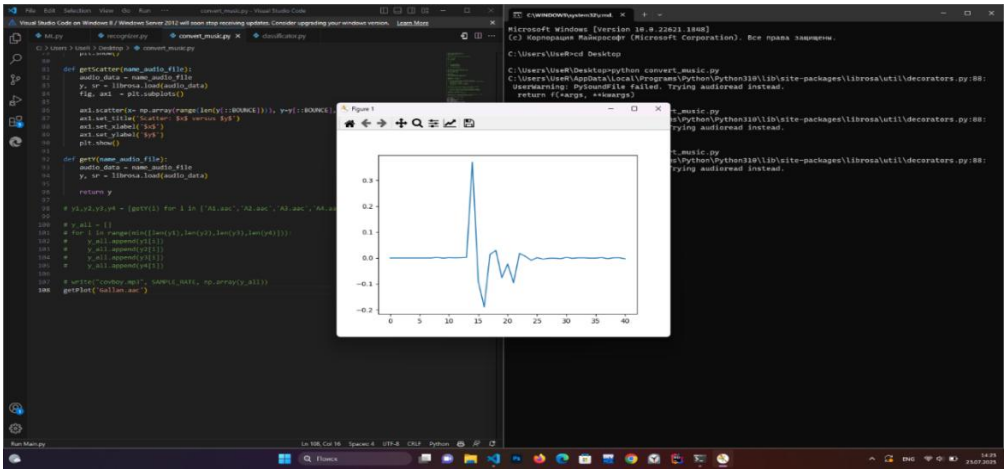

Рисунок 1. Спектограмма слова «*Galyň*»

2 Хлеб

[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -7.070387e-05, -0.0007108364, 0.0013004278, -0.00020348698, 0.003172811, 0.0017472361, -0.0056980136, 0.001056354, -0.00040601083, -0.00027398035, 0.0040313196, 6.3943444e-05, 0.0034262876, 0.004594876, -0.0005809597, 0.01720382, 0.16767077, 0.091864705, -0.11550929, 0.0210597, -0.00052442326, 0.010957343, -0.0012078126, 0.0054858434, -0.0024814683, -2.8397384e-05, 0.0019544598, -0.00022891922, 0.002090315, -0.001053638, -0.0024105515, -9.216722e-05, -0.0010083001, 0.0035980109, -0.0012208977, 0.0019506499, -0.0008992838, -0.0011744745]

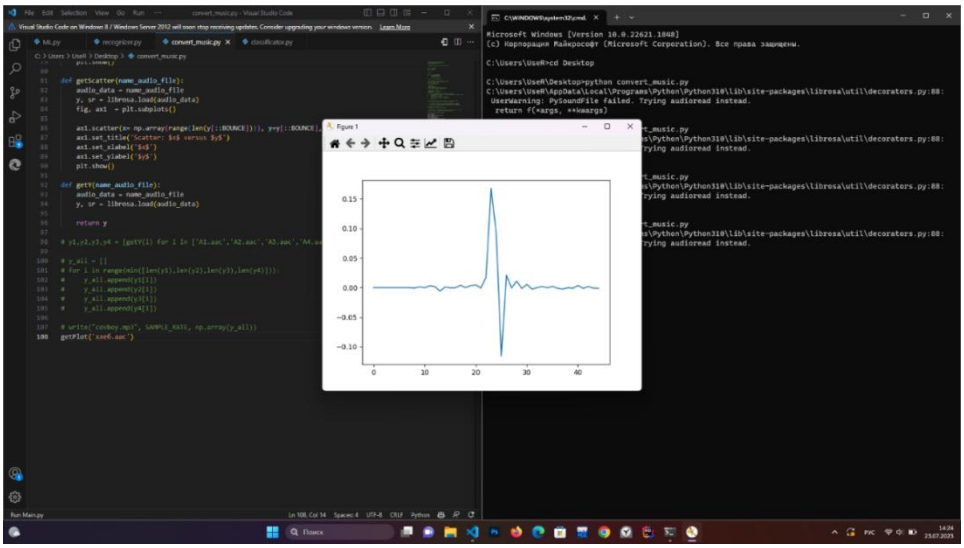

Рисунок 2. Спектограмма слова «Хлеб»

3 Car

[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 5.265972e-05, -0.0012408087, 0.0017302174, 0.0022520719, 0.0010935117, -0.00021780284, -0.0011149977, 0.0028004688, 0.016202055, 0.2799708, 0.029356796, 0.06212503, -0.16084166, -0.004478919, -0.0924315, -0.020218002, 0.023306465, -0.0009708171, -0.0070251557, 0.0035064754, -0.0026175305, 0.0024730577,

-0.0006635411, 0.0013180777, 0.00061744143, -0.0005611405, -0.00045245292, -0.0007941218, -0.0019001411, 0.0044609047, 0.0012176798]

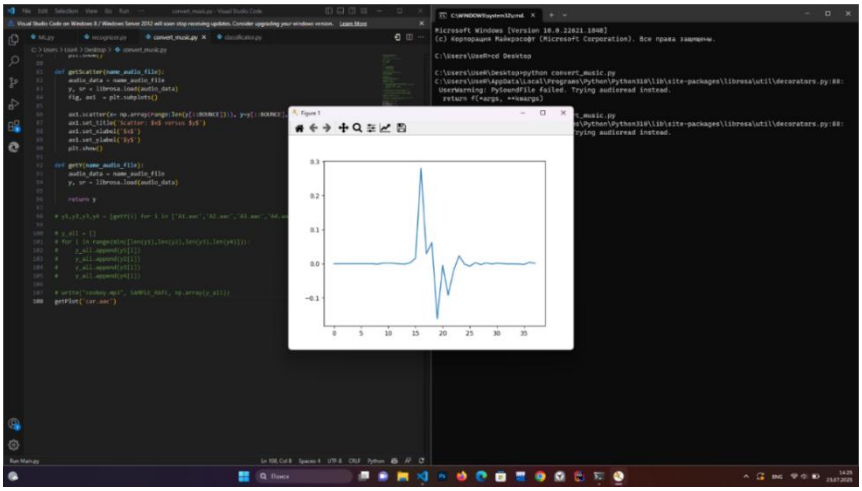

Рисунок 3. Спектограмма слова «*Car*»

Заключение. В результате перегрузка аудиосигнала в цифровой среде становится проблемой, с которой сталкиваются как звукорежиссеры, так и музыканты. Однако эту проблему можно преодолеть, правильно настроив уровни записи, используя специализированные алгоритмы и учитывая требования цифровых форматов. Управление уровнями и динамическим диапазоном имеет решающее значение для поддержания качества звука и достижения оптимального звучания при обработке цифровых аудиозаписей.

Список литературы

[1] Богданов Д.С – «Методы создания и использования речевых баз данных и инструментальных средств анализа и исследования речи для развития речевых технологий» автореферат кандидата технических наук, 2013 г.

[2]Чекмарев А. Речевые технологии - проблемы и перспективы. // Компьютерра, 26-43, 1997 г.

Авторский вклад

Авторы внесли равноценный вклад.

MODERN TRENDS IN SPEECH RECOGNITION

Myradov M. T.
Head of the Department of Information Systems, The Institute of Telecommunications and Informatics of Turkmenistan

Hydyrov R. B.
Head of the Scientific Department of The Institute of Engineering-technical and transport communications of Turkmenistan

Annotation. Speech recognition is the process of converting an audio signal containing speech into text form. This technology is used for a variety of purposes, including automatically recognizing commands in voice interfaces, transcribing audio recordings, converting speech to text in dictation systems, and much more.

Keywords: speech recognition, audio signal conversion, audio signal, BIG DATA in speech recognition, digital audio signal processing.