УДК 004.934 DOI: 10.18698/2541-8009-2018-11-399

ПОИСК И РАСПОЗНАВАНИЕ ЗАДАННЫХ КЛЮЧЕВЫХ ФРАЗ В СЛИТНОЙ РЕЧИ

А.Т. Левинский adam.levinskiy@yandex.ru
SPIN-код: 2301-6960

М.К. Быковский 1sherhan@mail.ru
SPIN-код: 6527-5828

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Рассмотрена проблема распознавания русской речи и нехватки программного обеспечения для удовлетворения нужд потребителей. Системы распознавания речи — это вычислительные системы, позволяющие выделять слитную речь говорящего из общего потока звуков и шумов. Эти системы основаны на технологии распознавания речи, которая преобразует произнесенные слова из звуков в цифровые тактовые сигналы с помощью процесса распознавания. В качестве решения предложено программное обеспечение для распознавания речи и поиска требуемой информации в полученном тексте. Проанализирована предметная область, выявлены заинтересованные лица и их цели, на основе которых составлены сценарии использования системы, переработанные в основные функции. Выполнен обзор существующих программных средств и алгоритмов. Проведено сравнение систем распознавания.

Ключевые слова

Распознавание, погрешность, интерфейс, тестирование, метод Левенштейна, метод шинглов, сравнение, функция

Поступила в редакцию 08.10.2018

Целью данной работы является создание инструмента, выделяющего в непрерывном потоке речи заранее заданные в текстовой форме слова и фразы и сигнализирующего об их обнаружении. Основная задача — уменьшение погрешности распознавания фраз. Поиск ключевых слов в непрерывном речевом потоке является одной из наиболее сложных задач в области распознавания. Основная проблема заключается в сложности верификации найденных слов. В данный момент одной из основных задач данной области является снижение частоты ложных распознаваний и, соответственно, ложных срабатываний. Применение данной технологии в коммерческих целях для систем реального времени стало возможным лишь с недавних пор. На практике такие системы распространены достаточно широко: начиная от применения в системах национальной безопасности, телефонных сервисах, системах контроля качества, системах речевых фильтров, заканчивая простыми устройствами, активизирующимися голосовой командой [1].

На данный момент существует множество программ и алгоритмов распознавания, но большинство из них находятся на стадиях зарождения и не спо-

собны обеспечить хороший функционал. Системы поиска фраз в речи можно использовать во многих направлениях: от систем прослушивания разговоров, где требуется распознавать угрозу, исходящую от говорящего, до систем голосового управления, где требуется из общего потока речи вычленить фразу-команду, относящуюся к системе. Важен также и экономический аспект, поскольку при значительных затратах, вкладываемых в эту область, отдача получается минимальной, в частности, для систем распознавания русского языка [2].

Речь — это последовательность языковых конструкций, создаваемых по определенным правилам построения. Процесс речи или общения предполагает формирование с одной стороны языковых форм и конструкций, а с другой — восприятие этих конструкций и их понимание [3].

Общение людей осуществляется с помощью голоса. Голос — это аппарат, реализующий способность человека издавать звуки, которые выражают наши эмоции и формируют нашу речь. Основными характеристиками голоса являются такие параметры, как частота, тембр, сила и длительность. Голосовые колебания формируются с помощью голосовых складок, которые вибрируют и создают колебания проходящего через них воздуха. Именно они отвечают за параметры нашего голоса, и у каждого человека эти параметры варьируются в определенных пределах. Так, частота голоса колеблется в диапазоне 64…1300 Гц, при этом разговорная речь находится примерно на отметке в 110 Гц [4].

Классификация систем распознавания. Системы распознавания классифицируются по нескольким признакам.

1. *Размер словаря*. Частота ошибок при распознавании напрямую зависит от размера словаря. Чем больше словарь, тем, соответственно, сложнее система распознавания и тем больше возникает ошибок при распознавании.

2. *Зависимость от диктора*. Существуют системы распознавания, зависящие от диктора и не зависящие. Зависящие от диктора системы обучаются одним пользователем и работают, соответственно, только с ним. Не зависящие от диктора системы, соответственно, работают с любым пользователем.

3. *Слитная или раздельная речь*. Речь, поступающую на вход системы, можно условно разделить на два вида: слитную и раздельную. В раздельной речи каждое слово отделяется друг от друга некой паузой; соответственно, слитная речь — это нормально произносимые предложения.

4. *Структурные единицы*. По использованию структурных единиц системы распознавания можно подразделить на два типа. В системах первого типа в качестве структурных единиц используются слова и фразы. Такой тип называют системой распознавания по шаблону. Второй тип строится на базе выделения лексических элементов, в нем в качестве структурных единиц используются фонемы, дифоны, аллофоны.

5. *Алгоритмы распознавания*. После того как речевой сигнал разбивается на определенные части, происходит вероятностная оценка принадлежности этих частей к тому или иному элементу распознаваемого словаря. Это осуществляется посредством одного из алгоритмов распознавания [5].

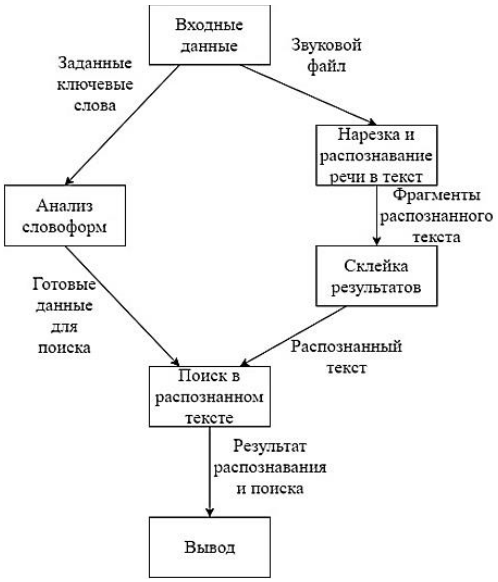

Рис. 1. Общая схема работы программы

Алгоритм и архитектура программы. Был разработан и предложен алгоритм работы специального программного обеспечения (рис. 2). Он состоит из основных блоков, подробное описание каждого из них представлено ниже.

В поступающем звуковом файле проверяется структура заголовка и целостность файла. Далее, исходя из таких параметров файла, как битрейт и размер, происходит нарезка файла на блоки (обоснование параметров нарезки приведено ниже). Используется рекурсивный вызов функций для работы с каждым фрагментом. При разбиении исходный файл уменьшается на длину фрагмента, а конец файла отслеживается специальной функцией.

При передачи фрагмента в функцию распознавания он отправляется на сервер Яндекса. Полученный ответ интерпретируется функцией распознавания. После проверки наличия предыдущих фрагментов в зависимости от

Общая схема работы программы. Схема работы программного обеспечения представлена на рис. 1.

На вход программы поступают два вида данных: звуковой файл формата *.WAV и данные, по которым необходимо выполнить поиск в данном файле. Сначала осуществляется отдельная обработка этих данных. Для звукового файла происходит его нарезка на фрагменты, распознавание (преобразование речевого сигнала в текст) и последующая склейка результатов распознавания, а для ключевых слов — анализ на словоформы, позволяющий выполнять поиск без учета склонений и падежей.

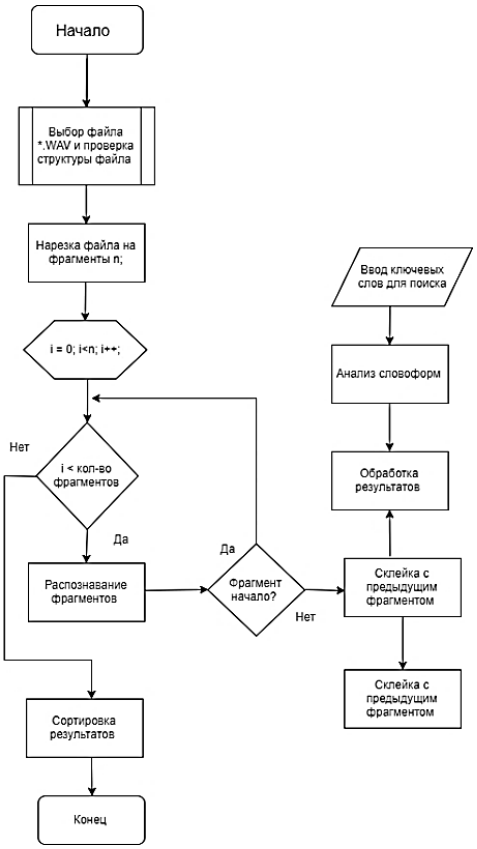

Рис. 2. Алгоритм работы программы

результата отправляем распознанный текст в функцию обработки. В функции обработки происходит поиск слов внутри обработанного текста. Но предварительно введенные слова обрабатываются с помощью функции анализа словоформ. Данная функция отбрасывает аффиксы у искомого слова, позволяя осуществлять поиск без учета склонения и падежа. Также присутствует возможность явно указать, что слово уже укорочено с помощью знака «=» (например, «распознава=» или «=познава=»). Обработанные данные выводятся на экран функцией вывода, затем запускается функция проверки конца исходного файла. Если обнаружен конец файла, запускается функция конечной сортировки результатов, позволяя расставить найденные слова в порядке вхождения в исходный файл. После этого обработка заканчивается [6].

Пользовательский интерфейс. Для комфортного взаимодействия пользователей с программой был разработан пользовательский интерфейс. В данной работе интерфейс представлен в упрощенном варианте (рис. 3), при реализации программы он будет требовать доработок.

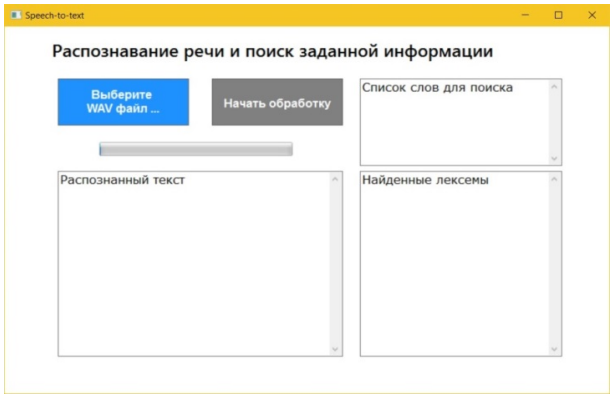

Рис. 3. Пользовательский интерфейс

Далее после начала обработки программное обеспечение выдаст распознанный текст и информацию, найденную в нем. Интерфейс после выполнения обработки представлен на рис. 4.

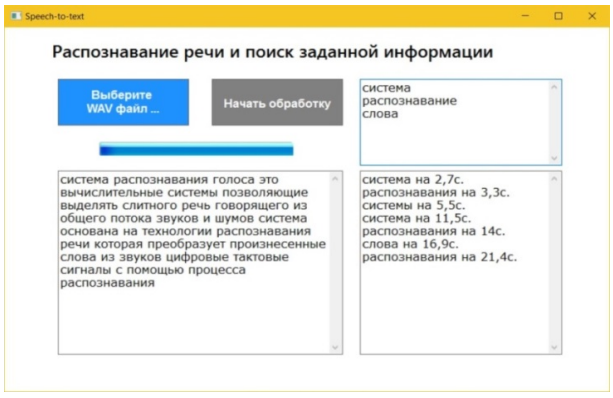

Рис. 4. Интерфейс после выполнения обработок

Алгоритм нарезки, описание и тестирование. Для улучшения эффективности работы программы и уменьшения ресурсоемкости, а также реализации попытки улучшить качество распознавания используется алгоритм нарезки исходной звуковой дорожки на фрагменты. Для данного случая были проведены эксперименты работы с различными параметрами, в частности, для файла размером 6 Мб было выполнено тестирование без разделения, с разделением по 3, 2, 1 Мб и 500 Кб. Результаты тестирования приведены в табл. 1 [7].

Таблица 1

Результаты тестирования

Параметр распознавания	Размер фрагментов для файла размером 6 Мб				
	6 Мб	3 Мб	2 Мб	1 Мб	500 Кб
Время распознавания, с	27	28	25	35	40
Верность, %	92,27	92,39	92,5	90,14	89,87

Запуск и использование программы. Работа программы не зависит от наличия на компьютере библиотек и прочего. Единственное условие ее эффективного функционирования — наличие стабильного соединения с Интернетом. Запуск программы выполняется с помощью исполняемого файла SpeechProg.exe. После запуска будет доступен пользовательский интерфейс. Для пользования программой необходимо иметь данные двух видов: звуковую дорожку в формате *.WAV и список ключевых слов для поиска.

Алгоритм использования программного обеспечения следующий:

1) нажать кнопку «Выберите WAV файл…»;

2) загрузить в текстовое поле «Список слов для поиска» необходимые данные;

3) нажать кнопку «Начать обработку». Распознанный текст будет отображаться в текстовом поле «Распознанный текст», а найденные ключевые слова в поле «Найденные лексемы».

Ход выполнения программы будет отображаться в информационной строке под кнопками управления. Распознавание завершится, когда будет заполнена вся строка.

Тестирование. Тестирование системы распознавания целесообразнее проводить с использованием данных от нескольких дикторов, записанных в разных условиях. Для тестирования программы одинаковый текст был проговорен четырьмя дикторами в шумном и тихом местах. В тихом месте уровень шума, измеренный с помощью любительского шумометра на промежутке в 37 с, составляет 38,4 дБ, при этом в условиях проводимого тестирования помещение характеризуется как тихое, если уровень шума в нем находится в пределах 36,6…44,7 дБ (рис. 5). В зашумленном месте уровень шума составляет 74,4 дБ, соответственно, шумное место в условиях данного тестирования характеризуется диапазоном шума 60,6…84,1 дБ. Результаты сравнения исходного текста и полученного в ходе работы программы представлены на рис. 5, 6 [8].

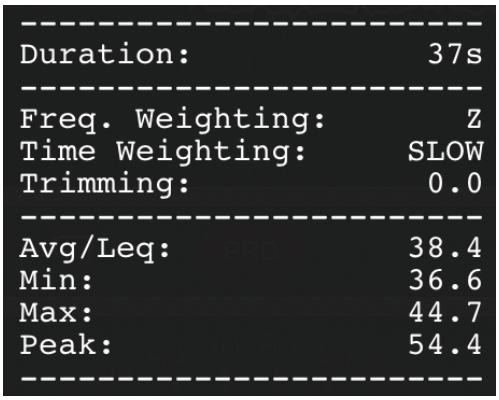

Рис. 5. Уровень шума в тихом месте

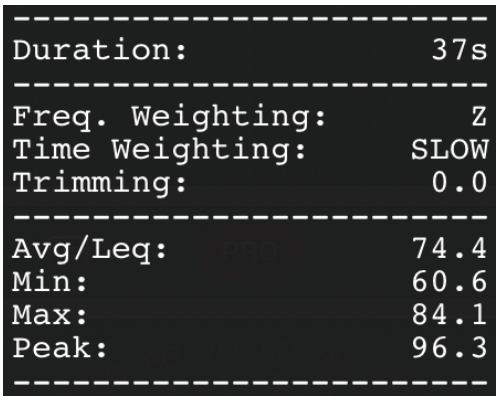

Рис. 6. Уровень шума в зашумленном месте

Данные сравнения распознанных текстов с оригинальными, а также процент найденных ключевых слов в данных текстах приведены в табл. 2.

Таблица 2

Приближение распознанного текста к оригинальному, %

Используемый метод	Текст 1		Текст 2		Текст 3		Текст 4	
	Шум	Тихо	Шум	Тихо	Шум	Тихо	Шум	Тихо
Левенштейна	93,97	96,25	89,07	95,88	92,22	95,17	92,41	95,10
Шинглов	63,64	78,02	53,85	75,27	73,40	69,47	64,95	75,27
Найденные слова, %	100	100	92	100	83	100	92	100

Результаты сравнения текстов двумя представленными методами оформлены в графическом виде и приведены на рис. 7 для метода Левенштейна и на рис. 8 для метода шинглов. Результат поиска ключевой информации в распознанных текстах представлен на рис. 9.

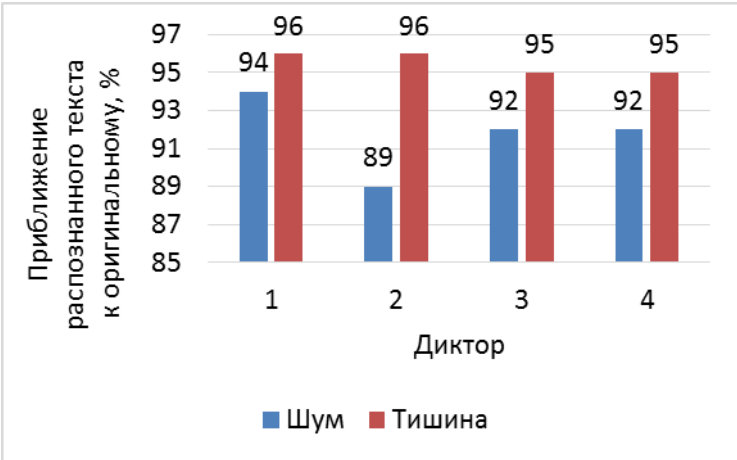

Рис. 7. Результаты, полученные методом Левенштейна

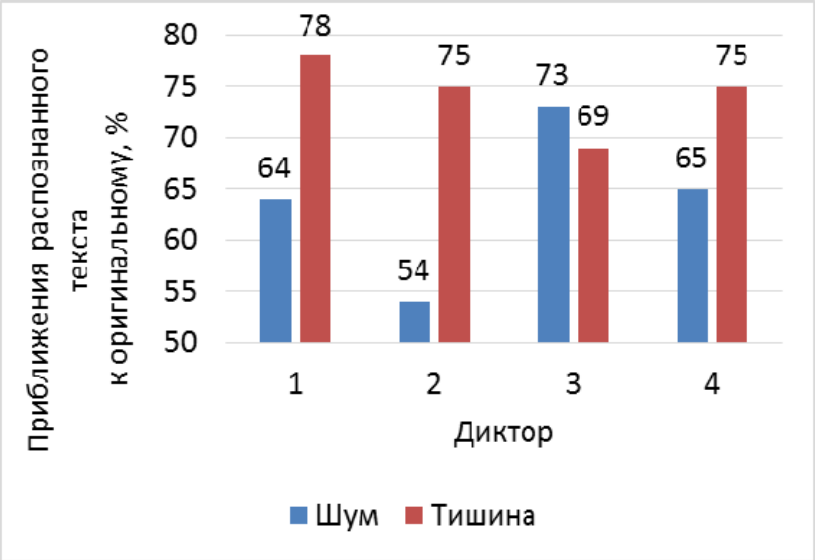

Рис. 8. Результаты, полученные методом шинглов

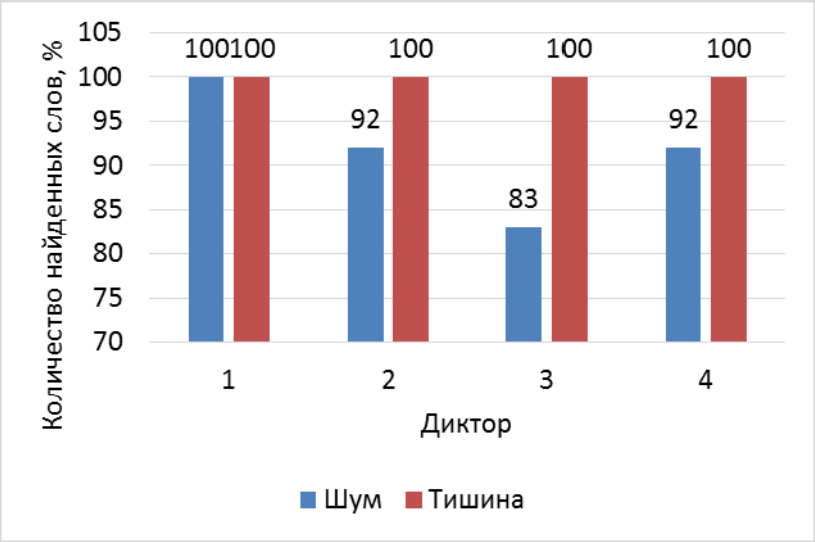

Рис. 9. Результаты поиска информации в текстах

Анализируя текущие результаты, можно сделать вывод, что при смысловом анализе текста методом Левенштейна, среднее значение верности распознавания составляет 94 %, средняя разница в верности распознавания зашумленного и чистого текста составляет 4 %. В то время как при анализе методом шинглов, показывающим, насколько тексты различаются между собой, выявлено, что среднее значение верности распознавания составляет 70 %, средняя разница в верности распознавания зашумленного и чистого текста составляет 11 %. Результаты анализа тестирования верности системы распознавания представлены в табл. 3 [9].

Таблица 3

Результаты анализа тестирования системы

Используемый метод	Среднее значение верности распознавания, %	Средняя разница верности распознавания двух текстов, %
Левенштейна	94	4
Шинглов	70	11

Таким образом, можно сделать вывод, что для распознавания слитной речи приемлемую вероятность распознавания можно получить по данным двух методов анализа, что позволяет с высокой эффективностью использовать данное программное обеспечение [10].

Установлено также, что вероятность верного нахождения заданной информации в распознанном тексте зависит лишь от качества распознавания и наличия адекватно распознанной информации в тексте в том или ином виде. Данные исследования показали, что в не зашумленных записях вероятность нахождения требуемой информации составляет почти 100 %, а в зашумленных записях — 91,75 %. Данный результат свидетельствует о том, что система эффективна и выполняет поставленные перед ней задачи.

Литература

[1] Галунов В.И., Соловьев А.Н. Современные проблемы в области распознавания речи. *Информационные технологии и вычислительные системы*, 2004, № 2, с. 42–45.

[2] Запрягаев С.А., Коновалов А.Ю. Распознавания речевых сигналов. *Вестник воронежского государственного университета. Сер. Системный анализ и информационные технологии*, 2009, № 2, с. 39–48.

[3] Центр речевых технологий (ЦРТ). URL: http://www.speechpro.ru/innovation/ (дата обращения 12.05.2017).

[4] Леонтьев А.А. *Язык, речь, речевая деятельность*. Москва, URSS, 2007, 212 с.

[5] Гусев М.Н. *Методы и модели распознавания русской речи в информационных системах*. Санкт-Петербург, СПбГУТ им. проф. М.А. Бонч-Бруевича, 2014, 378 с.

[6] Речевые технологии SpeechKit. Cloud. URL: https://tech.yandex.ru/speechkit/cloud/ (дата обращения 24.05.2017).

[7] Пальшина Д.А. Темп речи как одна из причин возникновения аллегровых форм русских слов в повседневной коммуникации. *Вестник Пермского университета. Российская и зарубежная филология*, 2013, № 2(22), с. 18–24.

[8] Мазуренко И.Л. *Компьютерные системы распознавания р*ечи. Москва, URSS, 2017, 56 с.

[9] Задача распознавания речи пока не решена. URL: https://habr.com/post/408017/ (дата обращения 23.03.2018).

[10] «Cloud, распознавание речи, технологии Яндекса» [В Интернете]. URL: https://tech.yandex.ru/speechkit/cloud/doc/guide/ (дата обращения 11.06.2018).

Левинский Адам Тагирович — студент магистратуры кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Быковский Максим Кириллович — студент магистратуры кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Тихомирова Елизавета Алексеевна, кандидат технических наук, доцент кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

DETERMINED KEY PHRASES SEARCH AND RECOGNITION IN CONTINUOUS SPEECH

A.T. Levinskiy adam.levinskiy@yandex.ru
SPIN-code: 2301-6960

M.K. Bykovskiy 1sherhan@mail.ru
SPIN-code: 6527-5828

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

The problem of recognition of Russian speech and of lack of software to meet the needs of consumers is considered. Speech recognition systems are computing systems that allow to select continuous speech of the speaker from the general flow of sounds and noises. These systems are based on speech recognition technology that converts spoken words from sounds to digital clock signals using a recognition process. As a solution, software is proposed for speech recognition and searching for the required information in the received text. The subject area is analyzed, interested parties and their objectives are identified, on the basis of which the system usage scenarios is compiled and processed into basic functions. A review of existing software and algorithms is performed. Recognition systems are compared.

Keywords

Recognition, error, interface, testing, Levenshtein method, w-shingling, comparison, function

Received 08.10.2018

References

[1] Galunov V.I., Solov'yev A.N. Modern problems in field of speech recognition. *Informatsionnye tekhnologii i vychislitel'nye sistemy*, 2004, no. 2, pp. 42–45.

[2] Zapryagaev S.A., Konovalov A.Yu. Speech signals recognition. *Vestnik voronezhskogo gosudarstvennogo universiteta. Ser. Sistemnyy analiz i informatsionnye tekhnologii* [Proceedings of Voronezh State University. Series: Systems analysis and information technologies], 2009, no. 2, pp. 39–48.

[3] Tsentr rechevykh tekhnologiy (TsRT) [Center of speech technologies (TsRT)]. Available at: http://www.speechpro.ru/innovation/ (accessed 12 May 2017).

[4] Leont'yev A.A. Yazyk, rech', rechevaya deyatel'nost' [Language, speech, speech activity]. Moscow, URSS publ., 2007, 212 p.

[5] Gusev M.N. Metody i modeli raspoznavaniya russkoy rechi v informatsionnykh sistemakh [Methods and models of Russian speech recognition in information systems]. Sankt-Petersburg, SPbSUT publ., 2014, 378 p.

[6] Rechevye tekhnologii SpeechKit. Cloud [SpeechKit speech technologies. Cloud]. Available at: https://tech.yandex.ru/speechkit/cloud/ (accessed 24 May 2017).

[7] Pal'shina D.A. Speech rate as a reason for reduced forms of Russian words in everyday communication. *Vestnik Permskogo universiteta. Rossiyskaya i zarubezhnaya filologiya* [Perm University Herald. Russian and Foreign Philology], 2013, no. 2(22), pp. 18–24.

[8] Mazurenko I.L. Komp'yuternye sistemy raspoznavaniya rechi [Computer systems of speech recognition]. Moscow, URSS publ., 2017, 56 p.
[9] Zadacha raspoznavaniya rechi poka ne reshena [Speech recognition problem is not solved yet]. Available at: https://habr.com/post/408017/ (accessed 23 March 2018).
[10] "Cloud, raspoznavanie rechi, tekhnologii Yandeksa" [V Internete]. Available at: https://tech.yandex.ru/speechkit/cloud/doc/guide/ (accessed 11 June 2018).

Levinskiy A.T. — Master's Degree student, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russian Federation.

Bykovskiy M.K. — Master's Degree student, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — E.A. Tikhomirova, PhD of Engineering, Assoc. Professor, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russian Federation.