

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 004.934

DOI 10.21685/2072-3059-2018-3-1

И. В. Алексеев, М. А. Митрохин, Е. А. Кольчугина

ПРОГРАММНОЕ СРЕДСТВО ОЦЕНКИ ЭФФЕКТИВНОСТИ ТЕХНОЛОГИЙ РАСПОЗНАВАНИЯ РЕЧИ

Аннотация.

Актуальность и цели. Объектом исследования являются современные технологии распознавания речи. Предмет исследования – оценка эффективности современных систем распознавания речи. Целью работы является определение основных показателей эффективности современных технологий распознавания речи на примере некоторых систем для определения возможности их применения в речевом интерфейсе систем специального назначения.

Материалы и методы. Исследования выполнены с использованием методов распознавания образов и методов математической статистики.

Результаты. Проведена оценка эффективности систем распознавания речи, получены количественные показатели точности и вероятности ошибок при распознавании произнесенных команд управления.

Выводы. Существующие системы распознавания речи общего назначения обладают достаточно высокой эффективностью, но ни одна из рассмотренных систем не может на данном этапе использоваться в специальных системах.

Ключевые слова: пользовательский интерфейс, речевой интерфейс, скрытые марковские модели, нейронные сети, распознавание речи

I. V. Alekseev, M. A. Mitrokhin, E. A. Kol'chugina

EFFICIENCY ESTIMATING SOFTWARE FOR SPEECH RECOGNITION TECHNOLOGIES

Abstract.

Background. The object of research is modern technologies of speech recognition. The subject of the study is the evaluation of the effectiveness of modern speech recognition systems. The purpose of the work is to determine the main performance indicators of modern speech recognition technologies by the example of some systems to determine the possibility of their use in the speech interface of special purpose systems.

© Алексеев И. В., Митрохин М. А., Кольчугина Е. А., 2018. Данная статья доступна по условиям всемирной лицензии Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), которая дает разрешение на неограниченное использование, копирование на любые носители при условии указания авторства, источника и ссылки на лицензию Creative Commons, а также изменений, если таковые имеют место.

Materials and methods. The research was carried out using the methods of pattern recognition and methods of mathematical statistics.

Results. The efficiency of speech recognition systems was evaluated, quantitative indicators of accuracy and error probability were obtained in recognition of spoken control commands.

Conclusions. The existing systems of speech recognition of general purpose have quite high efficiency, but none of the systems considered cannot be used in special systems at this stage.

Keywords: user interface, speech interface, hidden markov model, neural networks, speech recognition.

Введение

С момента появления электронно-вычислительных машин их пытались применять в различных областях, в том числе для использования в человеко-машинных системах специального назначения, к которым относятся автоматизированные системы управления, бортовые системы различных технических средств, а также системы массового обслуживания [1]. Такие системы находят применение в управлении транспортными потоками, при обеспечении безопасности важных объектов, для управления технологическим и иным оборудованием. К этим системам всегда предъявлялись повышенные требования по надежности и безопасности.

На сегодня при разработке систем специального назначения с большим числом задач, порядка нескольких десятков или сотен, и высоким уровнем автоматизации особое внимание уделяют пользовательскому интерфейсу в целях обеспечения максимального удобства и эффективности пользования им. С учетом особенностей функционирования систем специального назначения и требований, предъявляемых к их пользовательским интерфейсам, можно сделать вывод, что дальнейшее повышение эффективности взаимодействия оператора и вычислительной системы возможно при использовании наиболее естественных для человека способов обмена информацией – жестов, письма и речи. Так, уже обеспечивается поддержка ввода с сенсорных панелей. Однако наиболее привлекательным для использования в таких системах и наиболее перспективным является голосовой интерфейс, потому что это наиболее естественный и устойчивый способ общения. Крупнейшие мировые компании в своих разработках предлагают речевой интерфейс как альтернативу графическому. Более того, в системах специального назначения также давно применяются различные системы голосового оповещения оператора о различных событиях для повышения эффективности обратной связи. Ввиду этого для ускорения работы оператора в системе специального назначения, повышения надежности и эффективности его работы целесообразно применение прямой связи оператора с ЭВМ посредством голоса и получения в итоге полноценного речевого интерфейса.

В настоящее время существует несколько систем распознавания речи общего назначения. Для анализа с целью определения возможности внедрения в качестве подсистемы распознавания речи в интерфейс управления системы специального назначения были выбраны три системы распознавания речи: CMU Sphinx, DeepSpeech и Yandex ASR.

Система CMU Sphinx является дикторонезависимой системой распознавания слитной речи, использует в качестве признаков мел-частотные кеп-

стральные коэффициенты, из входного сигнала выделяет сеноны (устойчивые и переходные фрагменты фонем) при помощи скрытых Марковских моделей, при помощи детерминированного конечного автомата выбирает фразу из числа возможных с максимальной вероятностью правдоподобия [2].

DeepSpeech использует открытый исходный код, алгоритмы и инструментарий TensorFlow, где спектрограммы обрабатываются как изображения с помощью последовательности сверточных слоев нейронной сети, соединяющихся с последовательностью рекуррентных блоков [3]. На выходе нейронной сети появляется результат распознавания – последовательность символов [4]. В основу DeepSpeech положена система Sphinx4.

Система распознавания речи от Яндекс является дикторонезависимой системой и предназначена для распознавания русской речи. Акустический модуль реализован с помощью глубокой рекуррентной нейронной сети [5], в остальном система напоминает предыдущие рассмотренные, так как использует схожие механизмы функционирования. Данная система активно развивается и внедряется в новые области и отрасли в составе голосового помощника «Алиса».

1. Методика тестирования

Для учета особенностей работы операторов систем специального назначения предложена следующая методика тестирования [1]:

- Запускается программа из командной строки, тестируемая система и метод распознавания (формализованные команды по грамматике или неформализованная речь по расширенному словарю) указываются в качестве аргументов. При этом активируется программный модуль работы с указанной системой, которая загружает выбранную языковую модель.
- Перед началом диктовки диктор указывает произносимую фразу.
- С помощью графического интерфейса диктор запускает модуль записи речевого сигнала.
- В обстановке, имитирующей шум на автоматизированном рабочем месте оператора системы специального назначения, записывается речевой сигнал, сохраняется в файл.
- Записанные данные передаются в систему распознавания.
- По окончании распознавания результаты предоставляются пользователю, результат сравнивается с ответом, собирается статистика.

2. Разработка алгоритма работы программного средства тестирования

В соответствии с данной методикой разработан алгоритм тестирования систем распознавания речи (рис. 1).

Алгоритм работает следующим образом. Для начала диктовки пользователь нажимает кнопку на экранной форме, при этом программа начинает записывать речевой сигнал и выводит соответствующее уведомление пользователю, а тестируемая система освобождает ресурсы для распознавания данного речевого фрагмента. Пользователь, произнеся фразу, отпускает кнопку, программа останавливает запись, передает ее системе распознавания и выводит соответствующее уведомление. По окончании распознавания пользователю передаются результаты распознавания и после подтверждения правильности распознавания исполнительная команда передается на другие системы.

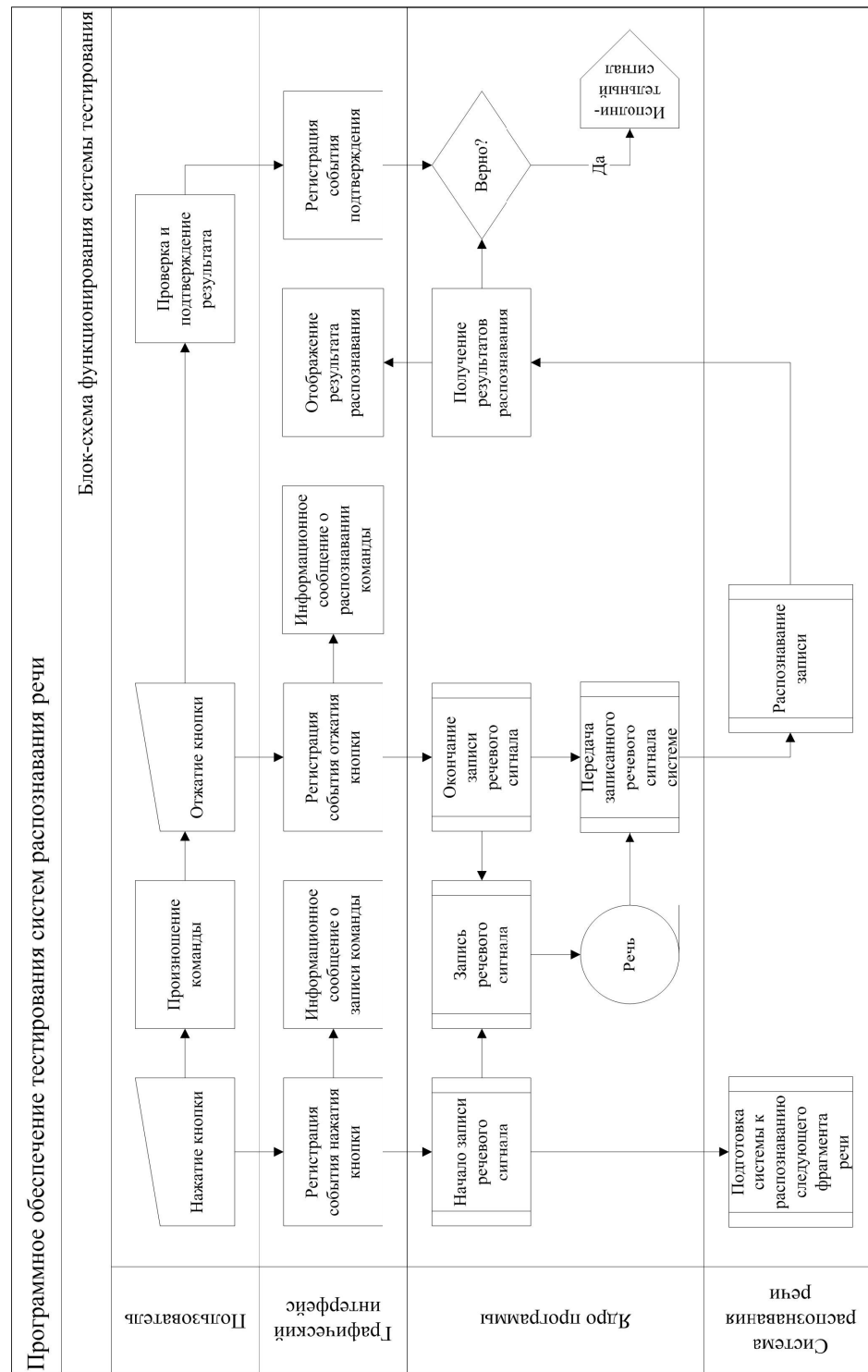


Рис. 1. Функциональная схема программного обеспечения тестирования систем распознавания речи

3. Разработка программного обеспечения тестирования

Для автоматизированного тестирования систем распознавания речи разработана программа, реализующая данную методику и работающая по описанному алгоритму. Программа состоит из модулей, представленных на рис. 2.

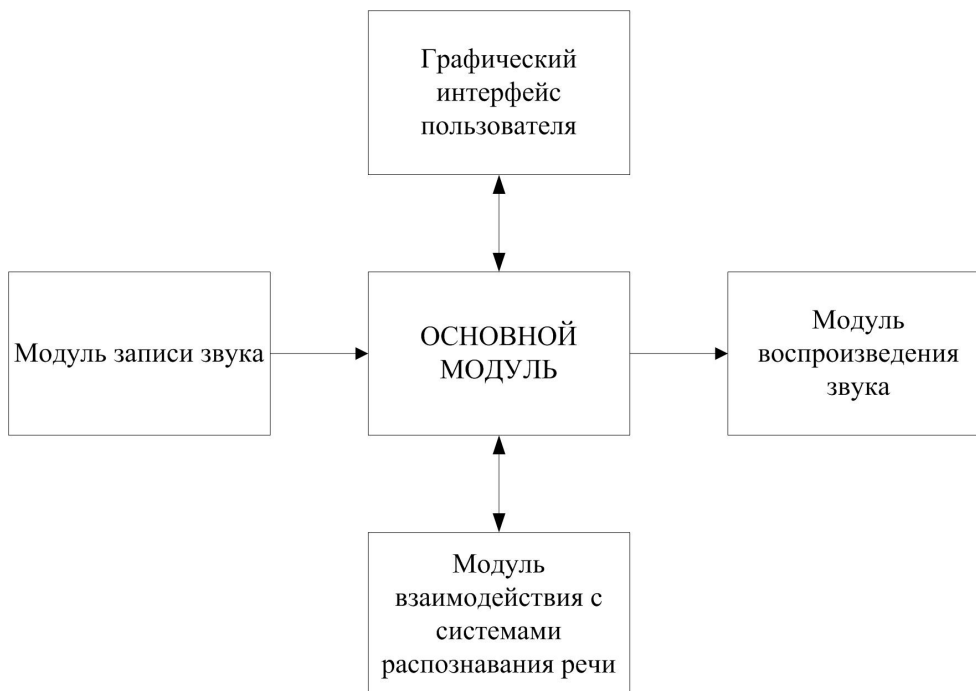


Рис. 2. Структура программы тестирования систем распознавания речи

Графический интерфейс пользователя предназначен для взаимодействия программы с оператором и содержит в себе средства работы с графическим интерфейсом и средства обработки событий.

Модули записи и воспроизведения звука предназначены для записи и воспроизведения речи и взаимодействуют с аудиосистемой ЭВМ через ее программный интерфейс

Модуль взаимодействия с системами распознавания речи предназначен для передачи системам записанной речи и получения результатов распознавания, а также для получения информации о скорости распознавания для сбора статистики.

Основной модуль представляет собой ядро программы и выполняет все функции по обеспечению записи и воспроизведения речи, поддержания приложения в работоспособном состоянии и выполнения всех задач приложения.

Графический интерфейс разработанного приложения представлен на рис. 3.

4. Тестирование систем распознавания речи

Тестирование проводилось на моделях, предоставляемых с данными системами, различными дикторами в условиях незначительной и средней за-

шумленности. Общий объем тестовых данных составил более 2 ч речи. Статистика результатов тестирования представлена в табл. 1 и 2.

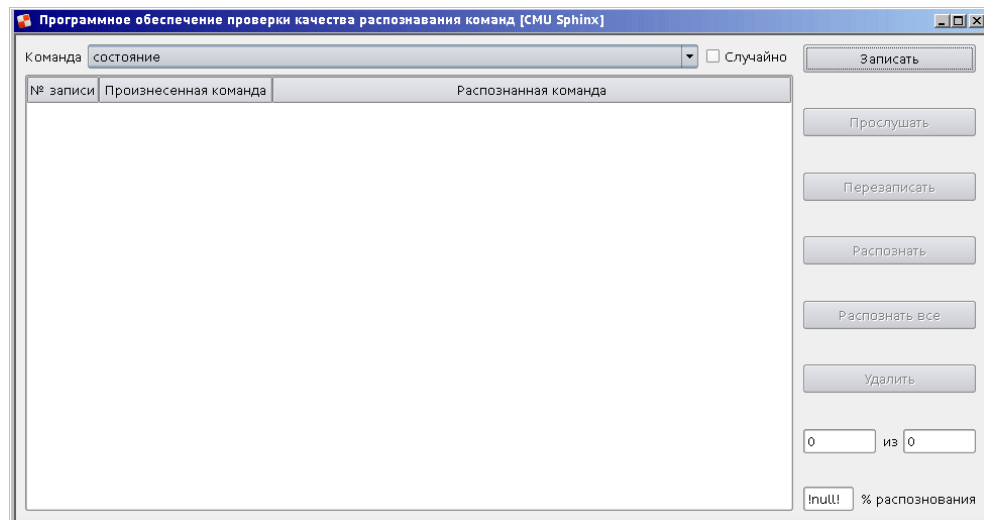


Рис. 3. Графический интерфейс пользователя разработанного приложения

Таблица 1

Средняя точность распознавания речи (%)

Система	CMU Sphinx	DeepSpeech	Yandex ASR
Формализованные команды, незначительное зашумление	95	96	—
Формализованные команды, значительное зашумление	84	85	—
Неформализованная речь, незначительное зашумление	58	70	88
Неформализованная речь, значительное зашумление	36	41	65

Таблица 2

Вероятности ошибочного распознавания команд (%)

Система	CMU Sphinx	DeepSpeech	Yandex ASR
Формализованные команды, незначительное зашумление	11	8	—
Формализованные команды, значительное зашумление	34	19	—

Результаты тестирования показали, что при распознавании ограниченного набора команд точность достаточно высокая, однако слишком велика вероятность ошибки при произнесении команды, отсутствующей в грамматике, что недопустимо.

Также было выявлено, что скрытые марковские модели не дают высокой точности распознавания ввиду использования статистических моделей, а также ввиду главного своего ограничения – зависимости последующего со-

стояния только от одного предыдущего. Особенно плохо такие технологии работают при распознавании спонтанной речи, сопровождаемой аграмматизмами и речевым «мусором», и в системах диктовки, как показано в работе [6] и подтверждено проведенными тестами.

Нейросетевые технологии показали лучшие результаты, а в условиях высокой зашумленности в проведенных тестах оказались наиболее устойчивыми к шумам. Ввиду использования рекурсивных связей в скрытых слоях нейросети преодолевался основной недостаток скрытых марковских моделей – зависимость только от одного предыдущего состояния.

Заключение

Из результатов тестирования видно, что ни одна из рассмотренных систем не может на данном этапе использоваться в специальных системах, где к точности распознавания команды предъявляются высокие требования. Следовательно, необходимо разрабатывать специальные алгоритмы и системы распознавания речи для применения в комплексах специального назначения.

Необходимым действием при построении подсистем распознавания речи в таких системах является надежная фильтрация входного сигнала от шумов, так как при различных вариантах ситуационных нарушений процессов речеобразования должна обеспечиваться высокая точность передачи смысловой информации. Кроме того, поскольку короткие фонетические сегменты практически лишены какой-либо лингвистической привязки, важным является совершенствование языковых моделей с учетом семантики и прагматики.

Библиографический список

1. Куликов, С. С. Тестирование программного обеспечения. Базовый курс : практ. пособие / С. С. Куликов. – Минск : Четыре четверти, 2015. – 294 с.
2. CMUSphinx Tutorial for Developers // GitHub. – URL: <https://cmusphinx.github.io/wiki/tutorial/> (дата обращения: 17.07.2018).
3. Speech and Machine Learning // Mozilla Research. – URL: <https://research.mozilla.org/machine-learning/> (дата обращения: 17.07.2018).
4. An Open Source Machine Learning Framework for Everyone // TensorFlow. – URL: <https://www.tensorflow.org/> (дата обращения: 17.07.2018).
5. Распознавание речи. Документация SpeechKit Mobile SDK // Yandex. – URL: <https://tech.yandex.ru/speechkit/mobilesdk/doc/common/speechkit-common-asr-overview-technology-docpage/> (дата обращения: 17.07.2018).
6. Современные проблемы в области распознавания речи // Auditech. – URL: <http://www.auditech.ru/page/darkness.html> (дата обращения: 17.07.2018).

References

1. Kulikov S. S. *Testirovanie programmnogo obespecheniya. Bazovyy kurs: prakt. posobie* [Software testing. Basic practical course]. Minsk: Chetyre chetverti, 2015, 294 p.
2. *GitHub*. Available at: <https://cmusphinx.github.io/wiki/tutorial/> (accessed Jul. 17, 2018).
3. *Mozilla Research*. Available at: <https://research.mozilla.org/machine-learning/> (accessed Jul. 17, 2018).
4. *TensorFlow*. Available at: <https://www.tensorflow.org/> (accessed Jul. 17, 2018).
5. *Yandex*. Available at: <https://tech.yandex.ru/speechkit/mobilesdk/doc/common/speechkit-common-asr-overview-technology-docpage/> (accessed Jul. 17, 2018).
6. *Auditech*. Available at: <http://www.auditech.ru/page/darkness.html> (accessed Jul. 17, 2018).

Алексеев Илья Владимирович

аспирант, Пензенский государственный университет (Россия, г. Пенза, ул. Красная, 40)

E-mail: aius@pnzgu.ru

Alekseev Ilya Vladimirovich

Postgraduate student, Penza State University (40 Krasnaya street, Penza, Russia)

Митрохин Максим Александрович

доктор технических наук, заведующий кафедрой вычислительной техники, Пензенский государственный университет (Россия, г. Пенза, ул. Красная, 40)

E-mail: vt@pnzgu.ru

Mitrokhin Maksim Aleksandrovich

Doctor of engineering sciences, head of sub-department of computer engineering, Penza State University (40 Krasnaya street, Penza, Russia)

Кольчугина Елена Анатольевна

доктор технических наук, профессор, кафедра математического обеспечения и применения ЭВМ, Пензенский государственный университет (Россия, г. Пенза, ул. Красная, 40)

E-mail: kea@pnzgu.ru

Kol'chugina Elena Anatol'evna

Doctor of engineering sciences, professor, sub-department of mathematical support and computer application, Penza State University (40 Krasnaya street, Penza, Russia)

УДК 004.934

Алексеев, И. В.

Программное средство оценки эффективности технологий распознавания речи / И. В. Алексеев, М. А. Митрохин, Е. А. Кольчугина // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – № 3 (47). – С. 5–12. – DOI 10.21685/2072-3059-2018-3-1