

УДК 004.522

doi: 10.17586/2226-1494-2019-19-4-714-721

ВЗАИМОДЕЙСТВИЕ С УСТРОЙСТВАМИ ИНТЕРНЕТА ВЕЩЕЙ С ИСПОЛЬЗОВАНИЕМ ГОЛОСОВОГО ИНТЕРФЕЙСА

В.Н. Шматков, П. Бонковски, Д.С. Медведев, С.В. Корзухин,
 Д.В. Голендухин, С.Ф. Спыну, Д.И. Муромцев

Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
 Адрес для переписки: shmatkovvlad@gmail.com

Информация о статье

Поступила в редакцию 23.04.19, принята к печати 01.06.19

Язык статьи — русский

Ссылка для цитирования: Шматков В.Н., Бонковски П., Медведев Д.С., Корзухин С.В., Голендухин Д.В., Спыну С.Ф., Муромцев Д.И. Взаимодействие с устройствами интернета вещей с использованием голосового интерфейса // Научно-технический вестник информационных технологий, механики и оптики. 2019. Т. 19. № 4. С. 714–721. doi: 10.17586/2226-1494-2019-19-4-714-721

Аннотация

Предмет исследования. Рассмотрены современные наиболее популярные голосовые ассистенты для управления устройствами интернета вещей, представленные на рынке, такие как Google Cloud Speech-to-Text, Amazon Transcribe, IBM Speech to Text, Yandex SpeechKit. Выявлены их достоинства и недостатки. Для работы голосовых ассистентов необходимо подключение к сети Интернет с целью обработки получаемых данных в «облаке», а также для синхронизации и управления устройствами пользователя. Большой практический интерес могут представлять голосовые ассистенты, которым не требуется наличие подключения к сети Интернет. **Метод.** В данной статье предложена модель организации локального (без использования сети Интернет) распознавания речи с использованием мобильных устройств. В качестве системы распознавания спонтанной речи используется программное обеспечение CMU Sphinx, осуществляющее распознавание речи с одновременным применением акустической и языковой моделей, а также трансляцию распознанной речи в команды, понятные системе, построенной на базе открытой платформы управления устройствами «OpenHab». Предложен подход к описанию грамматик и формированию словаря для распознавания речи. Приведен пример словаря и описания грамматики для речевой системы управления подключенными устройствами. Для тестирования описываемого в статье подхода создан демонстрационный стенд, основой которого является одноплатный персональный компьютер Raspberry Pi с установленным программным обеспечением openHab, и построены устройства интернета вещей на базе микроконтроллеров ESP8266. **Основные результаты.** Реализовано голосовое управление устройствами интернета вещей, которое осуществляется с применением протокола MQTT. Реализовано взаимодействие с сервером по протоколу HTTP. Проведено тестирование распознавания голосовых команд. Показана возможность практического применения предложенного в статье подхода к распознаванию спонтанной речи. **Практическая значимость.** Данная модель позволяет описать и подключить к системе значительное число устройств интернета вещей, представленных на рынке. Используя данную модель можно минимизировать или исключить влияние работоспособности сторонних внешних сервисов на работу системы голосового управления устройствами интернета вещей.

Ключевые слова

человечно-компьютерное взаимодействие, интернет вещей, голосовое управление, умный дом, управление устройствами

doi: 10.17586/2226-1494-2019-19-4-714-721

INTERACTION WITH INTERNET OF THINGS DEVICES BY VOICE CONTROL

V.N. Shmatkov, P. Bąkowski, D.S. Medvedev, S.V. Korzukhin,
 D.V. Golendukhin, S.F. Spynu, D.I. Mouromtsev

ITMO University, Saint Petersburg, 197101, Russian Federation
 Corresponding author: shmatkovvlad@gmail.com

Article info

Received 23.04.19, accepted 01.06.19

Article in Russian

For citation: Shmatkov V.N., Bąkowski P., Medvedev D.S., Korzukhin S.V., Golendukhin D.V., Spynu S.F., Mouromtsev D.I. Interaction with Internet of Things devices by voice control. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 4, pp. 714–721 (in Russian). doi: 10.17586/2226-1494-2019-19-4-714-721

Abstract

Subject of Research. The paper considers actual popular voice assistants for voice control of Internet of Things devices, such as Google Cloud Speech-to-Text, Amazon Transcribe, IBM Speech-to-Text, Yandex SpeechKit. Their pros and cons are identified. Internet connection is necessary for operation of voice assistants aimed at data processing in the cloud and synchronization and control of user’s mobile devices. Voice assistants, which can operate without Internet connection, can have significant practical value. **Method.** Architectural model for on-site speech recognition (without the Internet) with the use of mobile devices is proposed. CMU Sphinx software is used as a base for spontaneous speech recognition system. The software uses both acoustic and speech models for spontaneous speech recognition and also translates voice commands into such ones that can be processed by the system based on OpenHab open platform for device control. The approaches to grammar and dictionary creation for speech recognition are proposed. Example of dictionary and grammar description for voice control of attached devices are given. In order to test the described approach, the demonstration stand was created based on single-board Raspberry Pi computer with OpenHab software installed. In addition, devices of the Internet of Things based on ESP8266 microcontroller were built. **Main Results.** Control of the Internet of Things devices and interaction with the server are implemented with the use of MQTT protocol. Testing of voice commands recognition is carried out. The possibility of practical application of the proposed approach to spontaneous speech recognition is shown. **Practical Relevance.** Proposed model describes and integrates into control system a significant part of the Internet of Things devices represented on market. By applying the model, it is possible to minimize or even remove the impact of external third-party services on working capacity of voice control system for the Internet of Things devices.

Keywords

human-computer interaction, IoT, Internet of Things, voice control, smart home, device control

Введение

Интернет вещей — это концепция, появившаяся в начале 2000-х годов и с тех пор надежно закрепившаяся среди наиболее популярных направлений в области информационных технологий. В ее основе лежит идея о вычислительной сети устройств, способных взаимодействовать друг с другом и с внешним миром [1].

Решения интернета вещей предлагают различные способы взаимодействия пользователя и системы устройств. Успешные разработки в области распознавания и синтеза речи способствуют появлению нового вида человеко-машинных интерфейсов: голосовых ассистентов, способных вести с пользователем диалог на естественном языке и выполнять различные голосовые команды.

Основной проблемой голосовых ассистентов является облачная обработка данных и зависимость от качества интернет-подключения. Каждая из систем, представленных на рынке, имеет собственную экосистему и список поддерживаемых устройств. Часто подключить и настроить устройства других производителей невозможно или они имеют урезанный функционал.

В данной работе мы провели краткий обзор существующих решений области и предложили концепцию голосового управления системой интернета вещей, включающую в себя разбор спонтанной речи и формирование команды взаимодействия с устройствами. С упором на предложенную концепцию мы спроектировали стенд для тестирования разработанной системы интернета вещей. Данная работа основана на следующих технологиях: CMU Sphinx — открытое программное обеспечение (ПО) [2] для распознавания речи, OpenHAB — открытое ПО для координации интернет вещей¹.

Прототип голосовой платформы

Данная платформа состоит из двух подсистем: подсистемы голосового управления и подсистемы управления интернет вещами, которая отвечает за формирование и выполнение сценариев интернет вещами. В ходе реализации предложенной схемы голосового управления был создан стенд, имитирующий систему «Умный дом». Общая схема стенда отражена на рис. 1. Далее подробнее рассмотрим каждую из подсистем.

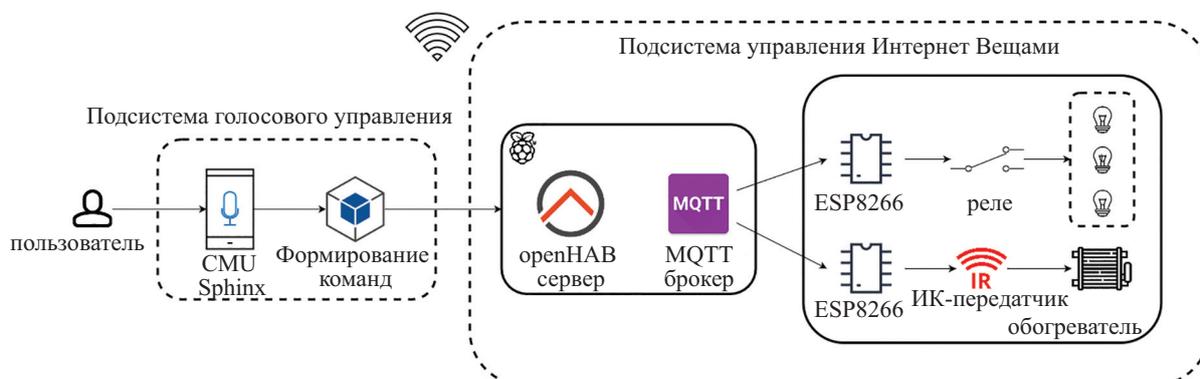


Рис. 1. Схема прототипа голосового управления интернетом вещей

¹ Документация openHAB [Электронный ресурс]. Режим доступа: <https://www.openhab.org/docs>, свободный. Яз. англ. (дата обращения: 15.05.2019).

Обзор голосовых ассистентов

На сегодняшний день существует большое разнообразие коммерческих систем автоматического распознавания речи: Google Cloud Speech-to-Text¹, Amazon Transcribe², IBM Speech to Text³, Yandex SpeechKit⁴. Данные системы решают большое количество задач и освобождают разработчика от необходимости создавать такую комплексную и сложную систему, как систему автоматического распознавания речи. Например, распознавание речи с помощью проверенных временем алгоритмов, демонстрирующих высокие результаты [3], можно реализовать с использованием высокопроизводительных вычислительных ресурсов облачных систем. Облачные платформы предоставляют отказоустойчивый программный интерфейс приложения (API, Application Programming Interface) для распознавания речи, а также имеют экосистемы с большим количеством пользователей, куда легко встраивать собственные приложения с голосовым интерфейсом.

Кроме несомненных достоинств, вышеупомянутые системы имеют следующие недостатки.

1. Перечисленные выше системы являются облачными системами, поэтому они зависят от их доступности, пропускной способности и прочих факторов, влияющих на скорость распознавания.

2. Нет возможности настроить систему распознавания речи для решения конкретной задачи. Качество распознавания зависит от используемой языковой модели. В разных прикладных областях разная вероятность у разных слов. Стандартные системы используют усредненную модель языка, либо модель, разработанную для решения проблем, поставленных перед создателями платформы, далеко не всегда совпадающими с задачами пользователей системы.

3. Обычно работа с облачными системами требует дополнительных финансовых затрат, заданных правилами использования таких систем.

4. Ресурсозатратность голосовой активации. При постоянной записи звука дополнительно расходуется интернет-трафик.

Современные системы локального распознавания речи можно разделить на две группы: основанные на последовательном использовании нескольких моделей — акустических и языковых; и end-to-end подходы, основанные на нейронных сетях и не требующие использования отдельных моделей. К первой группе можно отнести, например, Kaldi и CMU Sphinx, которые опираются в основном на HMM (анг. Hidden Markov Model) и GMM (анг. Gaussian Mixture Model) [4–6]. Представителями второй группы являются: wav2letter++, OpenSeq2Seq, ESPNet. Последние используют глубокое обучение, в частности, опираются на модели seq2seq (англ. sequence to sequence) с функцией потерь CTC (англ. Connectionist Temporal Classification) [7].

В последние десятилетия используют end-to-end системы за счет упрощения работы системы, а также сокращения разрыва в точности по сравнению с системами с отдельными моделями. Тем не менее в области локального распознавания речи для интернета вещей часто применяются именно подходы из первой группы. Например, некоторые работы демонстрируют успешное совместное применение GMM, HMM и нейронные сети с глубинным обучением (DNN, Deep Neural Network) [8].

Голосовое управление

Современные решения в области интернета вещей сложно представить без голосового управления и умных ассистентов. Более того благодаря технологическим достижениям в области голосового управления и искусственного интеллекта голосовой ассистент теперь является центром управления автоматизированной киберфизической системой — например, умным домом.

Подсистема распознавания речи разрабатываемой системы состоит из двух блоков.

1. Блок ASR (анг. Automatic Speech Recognition).

2. Блок формирования команды. ASR осуществляет перевод речевого сигнала в текст [9]. Таким речевым сигналом будут служить команды вида:

— «Умный дом, включи свет на кухне»;

— «Умный дом, проветри квартиру»;

— «Умный дом, нагрей комнаты до 25 градусов».

Блок формирования команды в свою очередь преобразует текстовые данные в структурированный объект, который бы характеризовал команду пользователя. Далее этот объект передается подсистеме формирования сценариев, которая на основе его и результатов работы других подсистем формирует соответствующие сценарии.

¹ Главная страница Google Cloud Speech-to-Text [Электронный ресурс]. Режим доступа: <https://cloud.google.com/speech-to-text>, свободный. Яз. англ. (дата обращения: 15.05.2019).

² Главная страница Amazon Transcribet [Электронный ресурс]. Режим доступа: <https://aws.amazon.com/ru/transcribe>, свободный. Яз. рус. (дата обращения: 15.05.2019).

³ Главная страница Watson Speech to Text [Электронный ресурс]. Режим доступа: <https://www.ibm.com/cloud/watson-speech-to-text>, свободный. Яз. англ. (дата обращения: 15.05.2019).

⁴ Главная страница Yandex SpeechKit [Электронный ресурс]. Режим доступа: <https://cloud.yandex.ru/docs/speechkit>, свободный. Яз. рус. (дата обращения: 15.05.2019).

Прототип подсистемы голосового управления был реализован в виде мобильного приложения на базе операционной системы (ОС) Android, в котором для распознавания использовалась библиотека pocketSphinx из CMU Sphinx. Распознавание речи происходило на самом телефоне без подключения к интернету [10].

Распознавание речи

Для разработки блока автоматического распознавания речи был использован проект CMU Sphinx. CMU Sphinx является одним из самых крупных проектов по распознаванию речи, а также обладает реализацией для ОС Android, что и стало решающим при выборе программного средства для распознавания речи.

Для распознавания речи в CMU Sphinx используются две модели: акустическая и языковая. Акустическая модель позволяет оценить распознавание речевого сегмента с точки зрения схожести на звуковом уровне. Для каждого звука описывается его произношение в речи. Языковая же модель позволяет определить наиболее вероятные последовательности слов. Для связи между языковой и акустической моделями задается словарь, в котором сопоставлены лексемы и фонемы (слово и его звучание). А также для описания языка задается грамматика — формальные правила, которые описывают простые правила построения предложений.

В качестве примера настроек используемой ASR продемонстрируем некоторые настройки разработанного прототипа системы голосового управления умным домом. В роли языковой и акустической моделей русского языка были использованы модели¹, подготовленные разработчиками CMU Sphinx. Словарь и грамматика создавались самостоятельно. Словарь состоит из слов, относящихся к управлению умным домом: «умный», «дом», «обогреватель», «свет» и т. д. Ввиду того что словарь состоит только из слов, связанных с прикладной областью, вероятность распознавания посторонних фраз становится очень низкой. Для каждого устройства описываются собственные правила грамматики. Так, если слова «включи» и «выключи» могут использоваться для нескольких устройств, то слово «горячее» может использоваться только в предложениях, относящихся к работе с обогревателем.

Пример описания грамматики для CMU Sphinx представлен в листинге 1.

```
<heaterExpression> = <heater> <intermediateAction> <heaterParameter>;
<heater> = (обогреватель);
<intermediateAction> = (сделай);
<heaterParameter> = (горячее|холоднее);
<lampExpressionOnOff> = <action><direction> <lamp>;
<direction> = (правую|левую|центральную|настольную);
<lamp> = (лампу);
<action> = (включи|выключи);
```

Листинг 1. Пример описания грамматики для CMU Sphinx

Учитывая, что предполагается использовать онтологические описания устройств, то словарь можно формировать из слов, относящихся к устройствам: к осуществляемым ими действиям, расположению, описанию и т. д.

Стоит отметить, что система распознавания работает в режиме постоянной обработки сигналов, записанных микрофоном. Для перехода в режим распознавания команд используется активационная фраза, услышав которую подсистема начинает анализировать сигнал с микрофона с целью сопоставить его с описанными ранее грамматиками.

Формирование команд

В блоке формирования команд необходимо на основании распознанной речи сформировать команду. В общем случае блок ASR возвращает строку, синтаксическое дерево или другие структурированные данные, из которых необходимо извлечь информацию об объекте, действии, месте действия и т. д. В случае использования CMU Sphinx можно использовать описанные ранее грамматики для извлечения необходимой информации.

В результате работы блока формирования команд создается объект, в котором указаны объект команды и действие, например, «лампа» и «включить». Далее объект передается в подсистему управления интернет вещами.

В рамках реализации прототипа в роли узла, формирующего сценарии, использовалось ПО openHAB — средство автоматизации управления интернетом вещей. Взаимодействие с openHAB осуществляется с помощью REST API.

¹ Модели CMU Sphinx для русского языка [Электронный ресурс]. Режим доступа: <https://is.gd/PiEI5J>, свободный. Яз. англ. (дата обращения: 15.05.2019).

Описание тестовой платформы IoT

Подсистема управления интернетом вещей в рамках стенда построена на базе одноплатного мини-ПК Raspberry Pi и микроконтроллеров. На стенде реализованы управление освещением (6 «умных» источников света), отоплением («умный» отопитель — конвектор, имеющий 2 уровня мощности), а также набор сенсоров. Каждое «умное» устройство строилось с использованием платформы Arduino на основе микроконтроллера ESP 8266, объединенного в сеть Wi-Fi с сервером выполнения команд. Сервер выполнения команд был развернут на платформе Raspberry Pi с установленным ПО openHAB. К этой же сети Wi-Fi был подключен смартфон, работающий на ОС Android с приложением, реализующим распознавание речи и формирование команды. Для отправки команд на сервер выполнения команд использовался протокол HTTP. Для взаимодействия сервера выполнения команд с Internet of Things (IoT)-устройств был использован протокол MQTT, работающий поверх протокола TCP. Выбор протокола обусловлен его гибкостью, простотой использования, простотой адресации к устройствам, поддержкой адресации к группам устройств посредством подписки на рассылку категории (или «темы» в терминологии MQTT) сообщений [11].

Устройство стенда схематически отобразено на рис. 1, фото стенда показано на рис. 2.

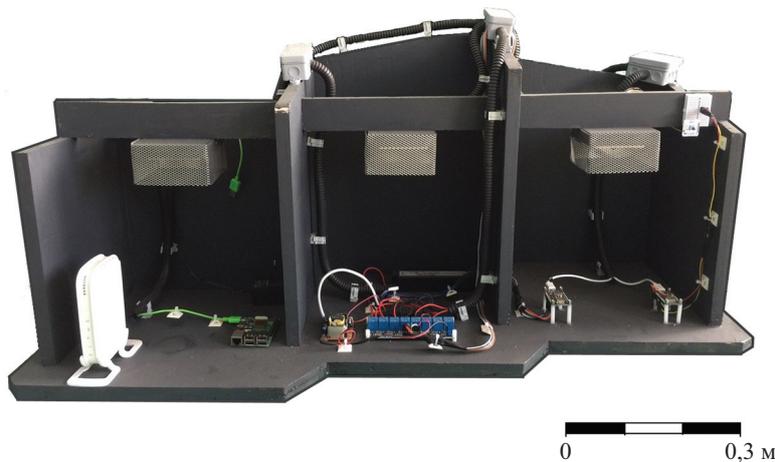


Рис. 2. Прототип платформы

Были собраны следующие IoT-устройства.

1. «Умный» светильник. Представляет собой контроллер, реализующий включение и выключение светодиодного светильника при помощи подачи управляющего напряжения на реле посредством выставления логического уровня на одном из цифровых выходов.

2. «Умный» обогреватель. Представляет собой контроллер, соединенный переходной цепью с ИК-передатчиком и отправляющий команды бытовому конвектору по ИК-каналу. В стенде использован бытовой конвектор De'Longhi DCH 4590 ER, имеющий два уровня мощности и управляемый по ИК-каналу.

3. Сенсор температуры и влажности. Собран на основе цифрового датчика DHT-21, передающего показания контроллеру по интерфейсу I2C.

4. Сенсор присутствия. Собран на основе цифрового пассивного инфракрасного датчика движения HC-SR501. Датчик автоматически срабатывает, если человек находится в поле видимости модуля и совершает движения.

5. Сенсор освещенности на основе аналогового датчика (фоторезистора). Аналого-цифровое преобразование сигнала датчика выполняется контроллером.

Схемы реализованных IoT-устройств показаны на рис. 3.

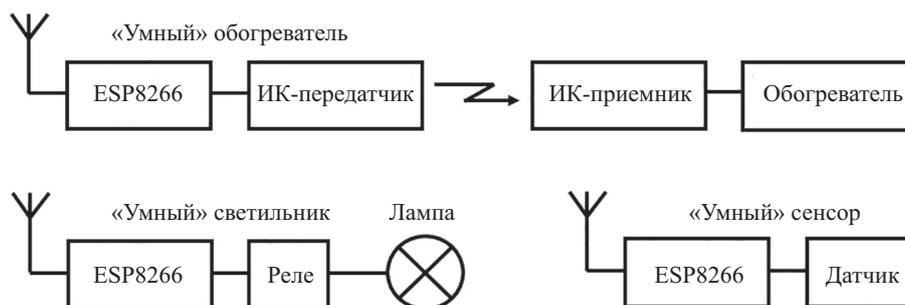


Рис. 3. Схемы IoT-устройств

Тестирование голосового управления

Работа тестового стенда происходила следующим образом.

1. Анализ сигнала микрофона подсистемой распознавания речи.
2. Переход в режим ожидания команды после обнаружения активационной фразы.
3. Распознавание речи. На данном этапе происходит сопоставление распознанных слов с заданной грамматикой.
4. Формирование команды. На основе распознанной речи формируются параметры и отправляется HTTP-запрос к серверу, осуществляющему взаимодействие с интернет вещами.
5. Выполнение команды. Сервер, реализующий взаимодействие с интернет-вещами, обрабатывает http запрос, после в зависимости от сценария, указанного запросом, отправляет необходимые команды IoT-устройствам.

В качестве примера голосовой команды рассмотрим следующее предложение: «Умный дом, сделай обогреватель горячее». В данном случае активационной фразой является «Умный дом», а командой — «сделай обогреватель горячее».

В рамках тестирования тремя разными дикторами было произнесено 30 голосовых команд, попадающих под правила грамматик, и 30 не попадающих. Суммарно система получила на вход 180 голосовых команд. Общее количество команд было разделено на две части по N команд в каждой. В таблице приведены результаты тестирования. Гипотезе H_0 соответствует ситуация, когда система распознает голосовую команду, а гипотезе H_1 — наоборот, система не распознает команду. Из всего множества команд первая часть команд относилась к H_0 , вторая — к H_1 .

Таблица. Ошибки первого и второго рода работы прототипа

		Верная гипотеза		N
		H_0	H_1	
Результат работы прототипа	H_0	76	5	81
	H_1	14	85	99

Оценка работы системы распознавания осуществлялась с помощью F -меры, учитывающей ошибки первого и второго рода:

$$F = 2 \frac{Precision \times Recall}{Precision + Recall},$$

где $Precision$ — точность классификации, определяется отношением числа верно распознанных голосовых команд к общему числу распознанных. $Recall$ — полнота классификации, определяется отношением числа верно распознанных голосовых к числу команд, которые должны были быть распознаны. В результате прототип показал $F = 88,89\%$.

На созданном стенде была протестирована предложенная схема реализации голосового управления. Установлено, что такой подход к реализации жизнеспособен.

Заключение

В ходе исследования был разработан прототип стенда «умного дома» с интерфейсом голосового управления с реализацией на локальном сервере без необходимости к подключению сети Интернет. По результатам работы выявлены недостатки в виде недостаточного показателя эффективности системы, что скорее связано с тем, что прототип разрабатывался с целью доказать возможность в небольшие сроки создать подобную систему. На данный момент недостатком можно считать сложность добавления новых умных устройств в систему. При добавлении устройства необходимо вручную обновлять словарь ASR, а затем дописывать обработчики новых команд, что является трудозатратным при частом обновлении списка устройств.

Разработанный стенд позволяет решить задачу тестирования и автоматизации помещений с применением различных архитектур и интерфейсов взаимодействия с пользователем.

Дальнейшее развитие платформы предусматривает использование онтологий, как способ описания правил взаимодействия устройств и логику работы системы в целом [12]. Планируется реализация автоматического добавления устройств в систему, а также использование логического вывода для динамического формирования сценариев управления интернет вещами. Для создания онтологии будет использоваться язык описания онтологий OWL, который будет интерпретироваться системой посредством модуля, написанного на Java с использованием библиотеки Apache Jena. Данный модуль будет участвовать в динамическом формировании сценариев взаимодействия устройств системы. Также рассматривается внедрение системы аутентификации пользователя по голосу на основе нейронных сетей [13–16].

Литература

1. Gershenfeld N.A. *When Things Start to Think*. New York: Henry Holt and Company, 2000. 224 p.
2. Беленко М.В., Балакшин П.В. Сравнительный анализ систем распознавания речи с открытым кодом // *Международный научно-исследовательский журнал*. 2017. № 4-4. С. 13–18. doi: 10.23670/IRJ.2017.58.141
3. Dernoncourt F., Bui T., Chang W. A framework for speech recognition benchmarking // *Proc. Interspeech*. Hyderabad, India, 2018. P. 169–170.
4. Povey D., Ghoshal A., Boulianne G. The Kaldi speech recognition toolkit // *Proc. Workshop on Automatic Speech Recognition and Understanding*, 2011.
5. Lamere P. et al. The CMU SPHINX-4 speech recognition system // *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Hong Kong, 2003. V. 1. P. 2–5.
6. Balandin S., Andreev S., Koucheryavy Y. (eds) *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2015. 713 p. doi: 10.1007/978-3-319-10353-2
7. Pratap V., Hannun A., Xu Q. et al. Wav2Letter++: A fastest open-source speech recognition system // *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Brighton, UK, 2019. doi: 10.1109/icassp.2019.8683535
8. Mehrabani M., Bangalore S., Stern B. Personalized speech recognition for Internet of Things // *Proc. IEEE 2nd World Forum on Internet of Things*. Milan, Italy, 2015. P. 369–374. doi: 10.1109/WF-IoT.2015.7389082
9. Levis J., Suvorov R. Automatic speech recognition // *The Encyclopedia of Applied Linguistics*. 2012. doi: 10.1002/9781405198431.wbeal0066
10. Sharma A.S., Bhalley R. ASR – A real-time speech recognition on portable devices // *Proc. 2nd Int. Conf. on Advances in Computing, Communication, & Automation*. Bareilly, India, 2016. doi: 10.1109/ICACCAF.2016.7749004
11. Дикий Д.И., Артемьева В.Д. Протокол передачи данных MQTT в модели удаленного управления правами доступа для сетей интернета // *Научно-технический вестник информационных технологий, механики и оптики*. 2019. Т. 19. № 1. С. 109–117. doi: 10.17586/2226-1494-2019-19-1-109-117
12. Maarala A.I., Su X., Riecki J. Semantic reasoning for context-aware Internet of Things applications // *IEEE Internet of Things Journal*. 2016. V. 4. N 2. P. 461–473. doi: 10.1109/jiot.2016.2587060
13. McLaren M., Lei Y., Ferrer L. Advances in deep neural network approaches to speaker recognition // *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Brisbane, Australia, 2015. P. 4814–4818. doi: 10.1109/ICASSP.2015.7178885
14. Matejka P., Glembek O., Novotny O. et al. Analysis of DNN approaches to speaker identification // *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Shanghai, China, 2016. P. 5100–5104. doi: 10.1109/ICASSP.2016.7472649
15. Бучнева Т.И., Кудряшов М.Ю. Нейронные сети в задаче идентификации диктора по голосу // *Вестник ТвГУ. Серия: Прикладная математика*. 2015. № 2. С. 119–126.
16. Ge Z. et al. Neural network based speaker classification and verification systems with enhanced features // *Proc. Intelligent Systems Conference*. London, 2017. P. 1089–1094. doi: 10.1109/IntelliSys.2017.8324265

Авторы

Шматков Владислав Николаевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0003-1391-1455, shmatkovvlad@gmail.com
Бонковски Патрик — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-8325-1873, baski@itmo.ru
Медведев Дмитрий Сергеевич — студент, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-5596-3988, mislitel.d@gmail.com
Корзухин Сергей Владимирович — студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0003-3163-9061, Sergey.korzukhin@gmail.com

References

1. Gershenfeld N.A. *When Things Start to Think*. New York, Henry Holt and Company, 2000, 224 p.
2. Belenko M.V., Balakshin P.V. Comparative analysis of speech recognition systems with open code. *International Research Journal*, 2017, no. 4-4, pp. 13–18. (in Russian) doi: 10.23670/IRJ.2017.58.141
3. Dernoncourt F., Bui T., Chang W. A framework for speech recognition benchmarking. *Proc. Interspeech*. Hyderabad, India, 2018, pp. 169–170.
4. Povey D., Ghoshal A., Boulianne G. The Kaldi speech recognition toolkit. *Proc. Workshop on Automatic Speech Recognition and Understanding*, 2011.
5. Lamere P. et al. The CMU SPHINX-4 speech recognition system. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Hong Kong, 2003, vol. 1, pp. 2–5.
6. Balandin S., Andreev S., Koucheryavy Y. (eds) *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*. Springer, 2015, 713 p. doi: 10.1007/978-3-319-10353-2
7. Pratap V., Hannun A., Xu Q. et al. Wav2Letter++: A fastest open-source speech recognition system. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Brighton, UK, 2019. doi: 10.1109/icassp.2019.8683535
8. Mehrabani M., Bangalore S., Stern B. Personalized speech recognition for Internet of Things. *Proc. IEEE 2nd World Forum on Internet of Things*. Milan, Italy, 2015, pp. 369–374. doi: 10.1109/WF-IoT.2015.7389082
9. Levis J., Suvorov R. Automatic speech recognition. *The Encyclopedia of Applied Linguistics*, 2012. doi: 10.1002/9781405198431.wbeal0066
10. Sharma A.S., Bhalley R. ASR – A real-time speech recognition on portable devices. *Proc. 2nd Int. Conf. on Advances in Computing, Communication, & Automation*. Bareilly, India, 2016. doi: 10.1109/ICACCAF.2016.7749004
11. Dikii D.I., Artemeva V.D. MQTT data protocol in remote access control management model for Internet networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 1, pp. 109–117 (in Russian). doi: 10.17586/2226-1494-2019-19-1-109-117
12. Maarala A.I., Su X., Riecki J. Semantic reasoning for context-aware Internet of Things applications. *IEEE Internet of Things Journal*, 2016, vol. 4, no. 2, pp. 461–473. doi: 10.1109/jiot.2016.2587060
13. McLaren M., Lei Y., Ferrer L. Advances in deep neural network approaches to speaker recognition. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Brisbane, Australia, 2015, pp. 4814–4818. doi: 10.1109/ICASSP.2015.7178885
14. Matejka P., Glembek O., Novotny O. et al. Analysis of DNN approaches to speaker identification. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Shanghai, China, 2016, pp. 5100–5104. doi: 10.1109/ICASSP.2016.7472649
15. Buchneva T.I., Kudryashov M.Yu. Neural network in the task of speaker identification by voice. *Herald of Tver State University. Series: Applied Mathematics*, 2015, no. 2, pp. 119–126. (in Russian)
16. Ge Z. et al. Neural network based speaker classification and verification systems with enhanced features. *Proc. Intelligent Systems Conference*. London, 2017, pp. 1089–1094. doi: 10.1109/IntelliSys.2017.8324265

Authors

Vladislav N. Shmatkov — engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0003-1391-1455, shmatkovvlad@gmail.com
Patryk Bąkowski — engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-8325-1873, baski@itmo.ru
Dmitry S. Medvedev — student, engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-5596-3988, mislitel.d@gmail.com
Sergey V. Korzukhin — student, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0003-3163-9061, Sergey.korzukhin@gmail.com

Голендухин Денис Валерьевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-5572-8902, evilkrauzer@gmail.com

Спыну Сергей Федорович — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, ORCID ID: 0000-0001-6123-8272, vadervega@gmail.com

Муромцев Дмитрий Ильич — кандидат технических наук, доцент, доцент, руководитель лаборатории, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, Scopus ID: 55575780100, ORCID ID: 0000-0002-0644-9242, d.muromtsev@gmail.com

Denis V. Golendukhin — engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-5572-8902, evilkrauzer@gmail.com

Sergey F. Spynu — engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, ORCID ID: 0000-0001-6123-8272, vadervega@gmail.com

Dmitry I. Mouromtsev — PhD, Associate Professor, Associate Professor, Laboratory head, ITMO University, Saint Petersburg, 197101, Russian Federation, Scopus ID: 55575780100, ORCID ID: 0000-0002-0644-9242, d.muromtsev@gmail.com