ПРИМЕНЕНИЕ ГОЛОСОВЫХ ПОМОЩНИКОВ И ПРОБЛЕМЫ ИХ ИСПОЛЬЗОВАНИЯ В АВТОМАТИЗИРОВАННОМ ПРОИЗВОДСТВЕ

APPLICATION OF VOICE ASSISTANTS AND PROBLEMS OF THEIR USE IN AUTOMATED PRODUCTION

УДК 004.522

Миронов С.Б.

аспирант

2 курс, факультет «Систем управления и робототехники»

Университет ИТМО

Россия, г. Санкт-Петербург

Mironov Semyon Borisovich

Graduate student

2nd year, faculty of control systems and robotics»

ITMO University

Russia, Saint Petersburg

mobrick@bk.ru

Аннотация

В статье рассматривается тема использования голосовых ассистентов в автоматизированном производстве. Описаны основные проблемы, связанные с использованием таких помощников, в том числе проблемы использования оборудования необходимого для их работы. Исследованы способы их решения. Описан план разработки устройства и программного обеспечения для проведения экспериментов по решению таких проблем. На данном этапе работы были исследованы и проанализированы проблемы использования голосовых помощников на производстве, решение этих проблем может быть достигнуто различными способами, но часто эти способы требуют дополнительных исследований в конкретном производстве, что требует некоторых средств. Следовательно, в дальнейшей работе

необходимо исследовать решения, которые упростят интеграцию голосовых помощников.

Annotation

The article discusses the use of voice assistants in automated production. The main problems associated with the use of such assistants are described, including problems with using the equipment necessary for their operation. Ways to solve them are investigated. A plan for developing a device and software for conducting experiments to solve such problems is described. At this stage of the work, the problems of using voice assistants in production were investigated and analyzed. the solution to these problems can be achieved in various ways, but often these methods require additional research in a specific production, which requires some funds. Therefore, in future work, it is necessary to explore solutions that will simplify the integration of voice assistants.

Ключевые слова: Интернет вещей, Голосовые помощники, Программное обеспечение, Микрофон, Автоматизированное производство.

Keywords: Internet of things, Voice assistants, Software, Microphone, Automated manufacturing

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.

По данным литературных источников, визуальный канал дает 80–90% всей информации [2], а второй наиболее важный, слуховой, практически не задействованный в производстве. Например, если коллега стоит в мастерской рядом с желаемым дисплеем, тогда гораздо проще спросить его, чем идти и искать нужный параметр самостоятельно. Кроме того, операторы часто глубоко вовлечены в рабочий процесс, и голосовое управление остается единственным возможным способом для взаимодействия с информационной системой.

Преимущества голосовых помощников были оценены еще в 70-е годы, когда появились первые прототипы устройств распознавания речи. Почти все ведущие IT-компании разработали или разрабатывают голосовых ассистентов: голосовой управление от Google, Алекса от Amazon, Siri от Apple, Cortana от Microsoft, Алиса от Яндекса и др. На данный момент их возможности значительно расширены от простого поиска информации до полного голосового управления различными бытовыми устройствами. Конечно, использование голосовых помощников в производственной среде имеет свои собственные нюансы, такие как шум в производственных помещениях, одновременное взаимодействие с несколькими операторами и проблемы безопасности.

Голосовые помощники используют во многих областях человеческой деятельности. Однако большинство исследований применимы только к бытовому использованию, например, установка напоминания о важной дате или времени, или установка конечной цели поездки. Напротив, практически нет исследований по использованию этой технологии в производственной сфере несмотря на то, что идея получения информации таким способом может быть успешным решением и для автоматизированного производства, где, например, вместо поиска нужного графического дисплея визуально, быстрее использовать голос.

В этой работе стоят задачи разработки не только программного обеспечения для обработки звуковой информации, но и решение проблем

путём применения аппаратных средств, для попытки достижения наиболее эффективного результата.

Кроме рассматриваемых и решаемых на этом этапе проблем стоит отметить, что 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. [3]

Так как суть работы на этом этапе заключается во многом в установке эффективного взаимодействия оператора на производстве и голосового помощника, при исследовании происходил поиск литературы в том числе и по термину *умный дом* так как в этой сфере уже найдено множество эффективных.

1 Проблемы использования голосовых помощников.

На производстве при работе в цехах вокруг оборудования, в которое можно интегрировать голосовых ассистентов могут быть плотно расположены рабочие места, в такой ситуации голосовой помощник может ошибочно начать считывать голосовые команды, которые ему не предназначались. Так как эта проблема существует не всегда, а только в определенных цехах с большой плотностью рабочих мест, было рассмотрено два возможных решения:

Решение проблемы на уровне программного обеспечения. Использование одной из форм биометрической аутентификации, позволяющей идентифицировать личность человека по совокупности уникальных характеристик голоса. Относится к динамическим методам биометрии. Однако, поскольку голос человека может меняться в зависимости от возраста, эмоционального состояния, здоровья, гормонального фона и ряда других факторов, не является абсолютно точным [4]. Также шум на производстве если он присутствует затрудняет идентификацию личности.

Распознавание речи в шумах предъявляет повышенные требования к качеству и достоверности методов и критериев распознавания. Наличие шумов в речевом сигнале снижает эффективность многих методов, традиционно используемых при распознавании, например, методов сегментации, выделения пауз, разделения слов и т. д. [5]. Наличие изменяющегося во времени шума в паузах может приниматься за полезный речевой сигнал и приводить к отказу ряда (особенно простых, первых) методов распознавания [6]. Для получения достоверного распознавания в шумах необходимо использовать подход, основанный на распознавании не по одному, а по комплексу критериев или характеристик речевого сигнала, а в сложных случаях – по результатам применения нескольких методов.

Нужно много исследований конкретного производства, чтобы предложить эффективное средство борьбы со случайными шумами. Для их

постоянного измерения, очевидно, потребуются специальные датчики, размещенные вблизи источников шумов и на значительном расстоянии от источника речевого сигнала, такие технические решения известны [7]. Проблема состоит в их сложной технической реализации и недостаточно высокой эффективности. Однако если провести исследования и анализ случайных шумов в конкретной обстановке, то полезная информация может быть накоплена и использована для их подавления, что будет более эффективным методом борьбы с шумами. При не случайных, то есть стабильных шумах, может быть исследование шума, анализ основных характеристик сигналов шума: амплитуды, частотного диапазона, временных характеристик (периода повторения – при периодических шумах) и т. д. Полученные результаты анализа можно использовать при разработке и реализации систем подавления шумов на начальных этапах обработки речевого сигнала при его распознавании.

При наличии стабильных и случайных шумов распознавание человека по голосу становится ещё более не предсказуемым что повышает риски при выполнении задач на производстве.

Как было отмечено во введении существуют и другие проблемы голосовых помощников. 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 , 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.

Но эти проблемы разрешаются работой в направлении создания автономных голосовых помощников, что может использоваться не только для автоматизированного производства, тогда как задачей этой работы будет решение проблем приёма звуковых команд на производстве на стороне «клиента» - оператора. Далее будут рассматриваться проблемы шумоподавления и работе голосового помощника при речи нескольких человек.

2 Программное обеспечение, используемое в качестве голосового помощника

Для выбора сервиса, который будет использован в качестве голосового ассистента необходимо выделить требования к этому программному обеспечению:

1. Необходима возможность использования русского языка для общения с помощником.
2. Необходимо наличие продвинутых технологий в этой области, чтобы минимизировать риски непонимания речи при использовании этого сервиса.

По этим требованиям больше всего подходит сервис Алиса, который обладает так называемыми «навыками», которые с технической точки зрения являются веб-сервисами. Такой веб-сервис ожидает реплики пользователя, которые поступают в него через Яндекс.Диалоги – платформу для разработчиков, которая позволяет добавлять сценарии общения с пользователями. Преимущество навыков в том, что они могут быть написаны на любом языке программирования или веб-фреймворке. Далее необходимо определить тип разрабатываемого навыка.

Чтобы определить какие команды будут реализованы в навыке необходимо рассмотреть, какие действия будут доступны путём использования голосовых команд. Так как воспринимаемая информация должна быть однозначно понятна программному обеспечению, иначе есть риск испортить изделие или даже травмировать оператора, слишком сложные и длинные команды необходимо либо разделить на короткие и простые, либо их исполнение оставить оператору. Существует множество типов оборудования, к которому потенциально эффективно подключить управление голосом, но при этом, доверить звуковым сенсорам можно только чёткие и короткие команды. Для тестирования таких команд необходимо разработать тестовый стенд - устройство.

3 Устройство, используемое для взаимодействия с голосовым ассистентом.

Для определения того, какое оборудование для распознавания речи безопасно и эффективно использовать на производстве необходимо разделить его на два типа:

Мобильное (динамическое) – не прикрепленное к рабочему месту, но к работнику на автоматизированном производстве. Пример такого оборудования – гарнитура Bluetooth или проводная подключенная к смартфону, серверу, рабочему компьютеру или другому оборудованию, поддерживающему данные устройства.

Статическое – оборудование которое работник не может использовать из любой точки на автоматизированном производстве. Прикрепленное к рабочему месту или мобильному роботу.

Первое, от чего необходимо отказаться в целях безопасности — это наушники, или любое другое оборудование, ограничивающее слух работников, так как звук часто служит сигналом тревоги в случае непредвиденных и опасных ситуаций. Также для многих работников вызовет дискомфорт снимать и надевать наушники каждый раз, когда надо поговорить с коллегой или начальством.

Второе – небезопасно наличие лишних висящих проводов, которые могут зацепиться за элементы оборудования на производстве. Из-за этого могут возникнуть риски не только порчи оборудования, но и травмирование работника.

При использовании статического оборудования безопасно использовать микрофон и звуковые колонки, подключенные к рабочему месту, через которые проходит общение с голосовым ассистентом,

При использовании динамического оборудования единственное найденное решение безопасное и комфортное для работы оказалось закрепление устройства записи к рабочему костюму, например к каске в случае, если она используется на производстве, или к верхней части костюма.

А динамики, через которые воспроизводятся голосовые сообщения ассистента могут быть закреплены либо также к костюму, либо к рабочему месту, на котором происходит работа. Такое оборудование не выпускается в большом объеме поэтому цена его подключения и поддержки будет выше, чем при использовании отдельных гарнитур.

В этой работе предлагается при использовании статического оборудования использовать систему из трех направленных микрофонов для установки источника звука по разнице во времени улавливания голоса, если подключить и настроить подобную систему, то таким образом решаются такие проблемы:

1. Будет обрабатываться только речь источника, звук от которого был принят центральным микрофоном раньше по времени, чем остальными, следственно будет обрабатываться только речь работника, который находится за этим рабочим местом.
2. Появляется возможность фильтровать шумы источников, которые находятся вне диапазона действия этой установки.
3. При установке этого оборудования на несколько устройств с которыми происходит взаимодействие работника, появляется возможность использовать направление голоса, чтобы обращаться именно к тому устройству, на которое повернут работник.

У такого подхода есть и недостатки:

При работе с этим оборудованием необходимо находится на одном определенном месте и при смещении с этой точки на небольшое расстояние пропадает гарантия того, что речь будет обрабатываться.

Не решается проблема того, что при обработке речи шум всё равно будет записывать вместе с ней. В этом случае необходимо использовать методы избавления от помех, предложенные во втором разделе.

3.1 Разработка схемы устройства

Для создания тестового стенда, который будет демонстрировать работу предложенного устройства необходимо определиться с его схемой. (рис. 1).

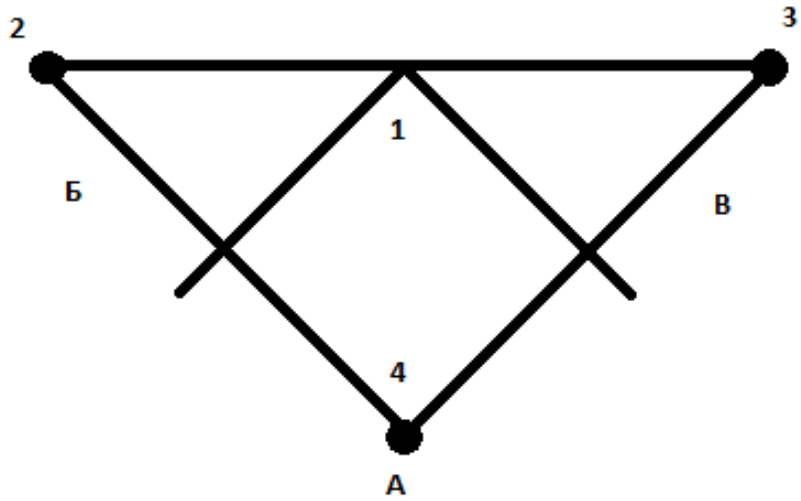

Рисунок 1. Схема устройства из трех микрофонов определения источника звука.

В точках 1, 2 и 3 кругами обозначено расположение микрофонов. А – сектор с которого происходит запись звука. В точке 4 обозначено расположение датчика, который указывает на то идёт запись звука или нет.

Угол сектора А не указан, так как его нужно настраивать отдельно для каждого рабочего места на уровне программного обеспечения, подставляя искусственные задержки считывания в код программы, либо путём передвижения микрофона 1 к или от точки 4. Расположение микрофона 1 ближе к оператору обусловлено необходимостью фильтровать звуки приходящие с противоположной от оператора стороны.

Звук обрабатывается только если звук считался на микрофоне 1 раньше, чем на остальных микрофонах. Далее определяется является ли этот звук речью. Если да, то голосовой помощник определяет команда ли это для действия. В случае если это команда, то ассистент отправляет запрос к соответствующему физическому исполнителю, который при принятии запроса выполняет необходимое действие.

3.2 Оборудование для сборки устройства.

Для тестового стенда необходимо:

1. Одноплатный компьютер Arduino или Raspberry Pi.
2. Три одинаковых микрофона, достаточно мощных для качественной записи речи.
3. Светодиод красный.
4. Светодиод зеленый.
5. Wi-fi роутер.
6. Wi-fi USB адаптер.
7. Компьютер, на который установлена Алиса.

На одноплатном компьютере планируется производить запись звуков со всех трёх микрофонов, звуковая информация с него через локальную сеть, обеспеченную Wi-fi роутером и USB адаптером подключенному к этой платформе, будет передаваться только в случае, если сигнал соответствует требованиям:

1) Звук громче нижнего порогового значения.
2) Звук на центральном микрофоне считался раньше, чем на остальных.

При передаче этой звуковой информации красный светодиод на одноплатном компьютере подает сигнал о том, что звук отправляется в локальную сеть. Далее разработанный навык Алисы определяет:

1. Является ли звук речью.
2. Если является, то соответствует ли он команде, заложенной в этом навыке.

Если оба результата оказались положительными, то Алиса отправляет сигнал обратно на одноплатный компьютер и «зажигает» желтый светодиод.

На данном этапе работы были исследованы и проанализированы проблемы использования голосовых помощников на производстве, решение этих проблем может быть достигнуто различными способами, но часто эти способы требуют дополнительных исследований в конкретном производстве,

что требует некоторых средств. Следовательно, в дальнейшей работе необходимо исследовать решения, которые упростят интеграцию голосовых помощников.

Был выбран сервис, который будет использоваться при развитии этой темы, а также разработана схема и определено необходимое оборудование, для построения стенда, на котором планируется тестирование различных решений для упрощения интеграции голосовых ассистентов в автоматизированное производство.

В дальнейшем планируется разработка навыка для Алисы и его оптимизация для наиболее комфортного использования, построение стенда для нахождения источника звука, а также эксперименты с ним.

Литература

1. Gershenfeld N.A. When Things Start to Think. New York: Henry Holt and Company, 2000. 224 p.
2. Н. Линдгрен, “Органы чувств и их электронные аналоги” Электроника, вып. 35, № 7, с. 22–27, 1962.
3. В.Н. Шматков, П. Бонковски, Д.С. Медведев [и др.] Взаимодействие с устройствами интернета вещей с использованием голосового интерфейса // Научно-технический вестник информационных технологий, механики и оптики, 2019, том 19, № 4
4. Е. К. Брагина, С. С. Соколов. Современные методы биометрической аутентификации: обзор, анализ и определение перспектив развития. // Вестник АГТУ. — 2016. — № 61. — ISSN 1812-9498.
5. Жигалов В.И. Проблемы сегментации речевых сообщений // Информационные и телекоммуникационные технологии // ООО «Литера–Принт». 2014. №24. С. 48-51.
6. Жигалов В.И. Особенности распознавания речевых сообщений в условиях шума // Концепции фундаментальных и прикладных научных исследований: сборник статей международной научно-практической

конференции (16.02.2018г., г. Тюмень.) В 2ч. Ч1/ - Уфа: ОМЕГА САЙНС, 2018, - 214с. (с.27-31).

7. Ли У.А., Нейбург Э.П., Мартин Т.Б. [и др.] Методы автоматического распознавания речи: в 2-х кн. Пер. с англ. / под ред. У. Ли. М.: Мир, 1983. Кн. 1. 328с., ил.

Literature

1. Gershenfeld N.A. When Things Start to Think. New York: Henry Holt and Company, 2000. 224 p.

2. N. Lindgren, “Sense organs and their electronic analogues” electronics, no. 35, no. 7, p. 22-27, 1962.

3. V. N. Shmatkov, P. Bonkowski, D. S. Medvedev [et al.] Interact with IOT devices using the voice interface // Scientific and technical Bulletin of information technologies, mechanics and optics, 2019, vol. 19, no. 4

4. E. K. Bragina, S. S. Sokolov. Modern methods of biometric authentication: review, analysis and determination of development prospects. // Bulletin of AGTU. — 2016. — no. 61. — ISSN 1812-9498.

5. Zhigalov V. I. problems of segmentation of speech messages // Information and telecommunication technologies // Litera– Print LLC. 2014. No. 24. P. 48-51.

6. Zhigalov V. I. features of speech message recognition in noise conditions // concepts of fundamental and applied scientific research: collection of articles of the international scientific and practical conference (16.02.2018, Tyumen.) In 2H. H1/ - Ufa: OMEGA SCIENCES, 2018, - 214s. (p. 27-31).

7. Lee U. A., Neiburg E. P., Martin T. B. [et al.] Methods of automatic speech recognition: in 2 books per. s angl. / ed. u. Li. M.: Mir, 1983. KN. 1. 328s., ill.