

ОСОБЕННОСТИ ПРИМЕНЕНИЯ ГОЛОСОВЫХ ИНТЕРФЕЙСОВ НА ПРОИЗВОДСТВЕ

А. В. КАТРИДИ, Ю. В. ФЕДОСОВ

*Университет ИТМО, 197101, Санкт-Петербург, Россия
E-mail: katridi@yandex.ru*

За последнее десятилетие качество обработки естественного языка значительно улучшилось благодаря усовершенствованиям как программного, так и аппаратного обеспечения. Современные технологии демонстрируют высокие результаты во всех областях — от автоматического распознавания речи до преобразования текста в речь и понимания естественного языка. Это позволило большому количеству компаний значительно повысить скорость предоставления услуг, качество и удобство использования продуктов в целом. Однако существует множество других, еще не используемых, приложений голосовых интерфейсов, и одно из них — применение таких технологий на производстве в рамках современных предприятий. При использовании голосовых технологий необходимо решить ряд задач. Представлены базовые компоненты обработки естественного языка, особенности их применения, а также подходы к созданию эффективного и в то же время простого голосового пользовательского интерфейса. Рассмотрены некоторые особенности промышленного проектирования и производства, которые следует учитывать при разработке таких систем.

***Ключевые слова:** машинное обучение, обработка естественного языка, Индустрия 4.0, промышленная инженерия, голосовой пользовательский интерфейс, голосовое программное управление*

Введение. Чтобы обеспечить простой и удобный человеко-машинный интерфейс, в частности голосовое управление, многие ИТ-компании решают задачи, связанные с человеко-машинными интерфейсами (языковая обработка, распознавание эмоций и намерений, преобразование речи в текст, преобразование текста в речь и др.), а также преодолевают технические проблемы (снижение времени отклика, повышение производительности, совершенствование архитектуры). Область обработки естественного языка быстро растет. Хотя существует множество проблем, связанных, например, с удержанием клиента и монетизацией [1], эта индустрия постоянно меняется, а также улучшается качество виртуальных ассистентов. Интересно, что некоторые непредвиденные события, такие как пандемия COVID-19, могут увеличить количество пользователей виртуальных ассистентов [2]. Почти все сценарии их использования ориентированы на широкую аудиторию, тем не менее остаются незанятыми ниши, в которых могут быть применены те же технологии [3].

Основные направления развития в Индустрии 4.0. Термин „Четвертая промышленная революция“ можно определить так: наполовину или полностью автоматизированный процесс принятия решений и общее усовершенствование производственного цикла с использованием новых технологий [4]. В [4] утверждается, что производственные системы станут на 30 % быстрее и на 25 % эффективнее: хотя цифры могут различаться, отражена общая тен-

денция, согласно которой использование новейших технологий и различных программ может значительно повысить скорость и качество производства.

Исследователи выделяют девять основных направлений развития Индустрии 4.0: большие данные и аналитика, автономные роботы, симуляция, горизонтальная и вертикальная системная интеграция, промышленный интернет вещей, кибербезопасность, облачные технологии, аддитивное производство, дополненная реальность [4].

Большинство предложений по улучшению производственного процесса делятся на две категории: решения либо исключают необходимость вмешательства человека, либо улучшают человеко-машинные интерфейсы.

Есть исследования, посвященные задачам автономного функционирования, или тем ситуациям, где требуется небольшое вмешательство человека [5, 6]. Предлагаемые там подходы предназначены для повышения уровня автономности функционирования. Однако эти подходы не могут быть применены в каждом конкретном случае в связи с увеличением затрат и ограничений, поэтому работнику по-прежнему необходим простой и эффективный способ взаимодействия с машинами и принятия решений на основе предоставленных данных и прогнозов.

Оператор промышленного оборудования сталкивается с новыми проблемами, такими как растущий спрос на индивидуальную настройку и постоянно меняющаяся среда. Необходимо уметь действовать быстро и точно, чтобы стать „оператором 4.0“, потому что так или иначе интерфейсы взаимодействия все еще ориентированы на человека [7, 8].

Некоторые исследователи предложили адаптировать подходы компьютерного зрения для использования преимуществ обработки визуальных данных [9], но из-за ограничений и неинтуитивного способа взаимодействия с машинами это может применяться только в редких случаях.

Технологии виртуальной, дополненной реальности и цифровых двойников в целом достаточно развиты, чтобы дополнять окружающую среду некоторой визуальной информацией, моделируя физическое пространство максимально близко к реальности и совершенствуя процесс диагностики и принятия решений [10—12]. Кроме того, предиктивная аналитика помогает улучшить процесс принятия решений и упростить диагностику проблем еще на ранних стадиях [13—15].

Утверждается [16], что использование голосовых помощников в коммерческом секторе быстро растет, но они не рассматриваются как канал взаимодействия в промышленной среде, хотя голосовые технологии больше не сводятся только к чат-ботам для ответа на часто задаваемые вопросы. Такие технологии позволяют выполнять сложные задачи, например, заказывать столик в ресторане [17] или помогать арендаторам в отелях [18]. Тем не менее эти примеры сценарно-ориентированы и специфичны для предметной области. Группа исследователей [19] недавно представила обученную на огромном корпусе текстов модель естественного языка, которая способна генерировать ответы и понимать смысл специфичной лексики. Однако, несмотря на приемлемую способность точно распознавать предложение и связанное с ним намерение, ход диалога и сценарии по-прежнему необходимо реализовывать с учетом специфики предметной области, например, производства. Именно проблемы безопасности, помимо экономических, влияют на скорость адаптации таких технологий и использование роботов и коботов [20].

На сегодняшний день не существует универсального интерфейса, который бы широко применялся разными производителями и был стандартизован. Помимо этого, трудно измерить экономический эффект от улучшения интерфейсов, в отличие от алгоритмов прогнозирования или автономных линий роботов, которые работают без вмешательства человека [21]. Тем не менее понятный и унифицированный стандартный интерфейс для всех отраслей может способствовать более быстрой, менее подверженной ошибкам и более эффективной ра-

боте. Это преимущество не следует недооценивать. Скорость внедрения новых технологий в потребительской сфере также может повлиять на промышленную сферу, поскольку человек ожидает, что устройства будут иметь возможность общаться через голосовой канал. Таким образом, последним элементом в направлениях развития Индустрии 4.0 является улучшенное взаимодействие человека с машиной (рис. 1) [22].

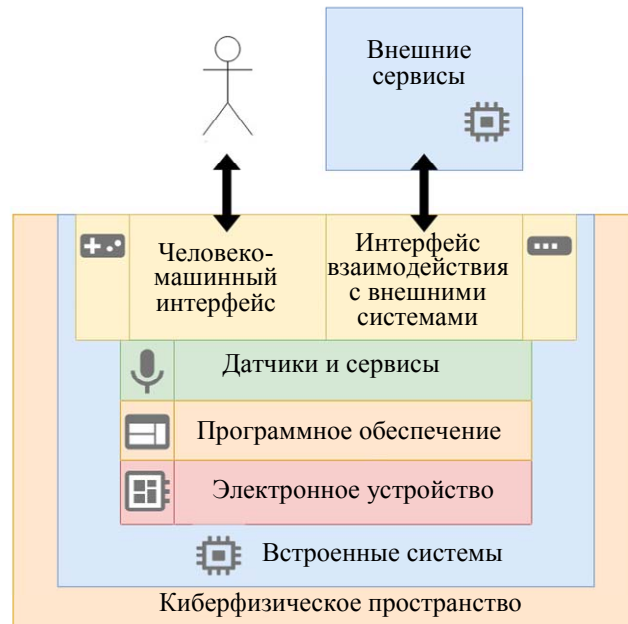


Рис. 1

Голосовые технологии на производстве могут обеспечить не только более комфортный пользовательский интерфейс, но и большую безопасность для пользователя, поскольку не требуют дополнительных интерфейсов и контроля для ручных операций. Для того чтобы подобные технологии эффективно адаптировать, нужно понимать, какие их компоненты являются ключевыми и какие проблемы существуют при адаптации таких технологий.

Ключевые компоненты технологий. Способность современных приложений понимать речь пользователя и затем предпринимать соответствующие действия обеспечивается комбинацией автоматического распознавания речи (ASR), преобразования текста в речь (TTS) и понимания естественного языка (NLU).

Основной проблемой использования ASR-компонента является правильное восприятие звука и перевод его в текст, а также разрешение спорных случаев с достаточно высоким процентом правильного распознавания. Помимо лингвистических существуют сложности, связанные с шумами среды, в которой произносится текст. Важно, чтобы модель распознавания была достаточно устойчива и к шумам.

Вторым является NLU-компонент. Термин NLU трактуется очень широко, нас будет интересовать несколько разделов в этой области понимания:

- распознавание намерения: по входящей фразе (уже в текстовом виде) нужно понять, что именно хочет получить пользователь от системы – это нетривиальная задача, когда у системы достаточно широкий спектр выбора, и нужно правильно выбрать между вариантами;

- выделение сущностей, которые могут быть: именованными (Named Entity Recognition, NER), например „Отправь это Петровичу“, или содержать какие-то необходимые данные (Slot Filling), чтобы потом передать эти данные для выбранного сценария, например: „Перейди на координаты 20 20“.

Естественный язык сложно формализовать: как только растет список задач, которые должны быть решены, возрастает и сложность реализации этого компонента.

Еще один компонент — диспетчер, который будет на основе информации из предыдущих шагов — намерения, сущностей и т.д. — выбирать необходимую команду для отправки на устройство. В таком случае существует сложность с выбором правильного сценария и формированием нужных команд для передачи на устройство. Можно разработать универсальный адаптер, который бы отвечал за преобразование команд в специфичные для конкретного устройства, чтобы выполнялся принцип „написать один раз и запускать, где угодно“ [23]. В этом случае открывается возможность работать с широким кругом устройств и платформ.

Наконец, последний компонент — преобразование текста в речь (TTS) — необходим, чтобы ответить пользователю (задачи похожи на задачи ASR). В промышленной сфере, возможно, стоит немного пренебречь живостью речи и тем, насколько естественно она звучит, в пользу четкости и корректности произносимой информации, поскольку высока цена ошибки.

Базовая сквозная схема взаимодействия представлена на рис. 2.

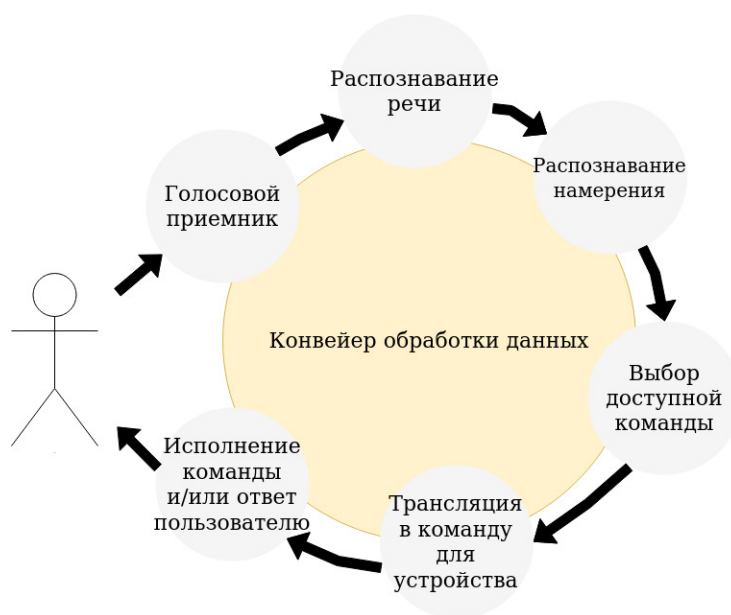


Рис. 2

Голосовой приемник получает вводимые пользователем данные и отправляет их через Wi-Fi или bluetooth-компонент, звуковые волны преобразуются в аудиоформат, а затем в текстовый формат. Затем текст отправляется в модуль NLU с контекстом: датой, предыдущим стеком команд и всей потенциально полезной информацией, необходимой для совершенствования процесса принятия решений для системы. Модуль NLU на основе общих входных данных извлекает сущности, предсказывает намерения и отправляет их диспетчеру. Диспетчер решает, какую команду выбрать, а затем отправляет ее на заданное устройство после преобразования в соответствующую команду.

В такой системе (помимо диалоговых стратегий, способа представления данных пользователю и общего дизайна) внутренний механизм должен отвечать некоторым критическим требованиям. Один из них — скорость ответа. Обработка данных — от перевода звуковых волн в аудиоформат, аудио в текст, текста в сущности и обратно — довольно затратная по времени и вычислительным мощностям операция.

В графических приложениях пользователь ожидает, когда приложение загружается. Пока приложение не произвело обработку данных, ничего сделать невозможно. Эквивалентное времени загрузки в графическом приложении время ожидания в голосовом интерфейсе — это время, затраченное на ожидание ответа приложения, распознавание того, что сказал пользователь, и понимание того, что он имел в виду. Эти процессы требуют времени, и время ожидания должно быть как можно меньше. Задержки обычно вызваны: недостаточной пропуск-

ной способностью канала связи; ресурсоемкими вычислительными операциями; ожиданием ответа от внешних служб.

Таким образом, требуемые преобразования могут быть ограничены либо скоростью выполнения вычислительных операций, либо ожиданием ввода-вывода.

Подходы к дизайну диалоговых систем. Одним из наиболее важных аспектов правильного проектирования голосовых технологий является использование известных принципов для построения диалога [23].

Настоящее время можно охарактеризовать как вторую эру VUI. Мобильные приложения, такие как Siri, Google Assistant, Alexa и др., сочетающие визуальную и слуховую информацию, а также голосовые устройства, например, Amazon Echo и Google Home, сейчас массово используются. Согласно отчету Google, 20 % поисковых запросов теперь выполняются с помощью голоса [24].

Голосовое управление имеет несколько достоинств:

- диктовать текстовые сообщения быстрее, чем печатать их [25];
- в некоторых случаях требуется, чтобы руки были свободны, например, при вождении или приготовлении пищи;
- этот интерфейс доступен широкому кругу пользователей;
- тон, громкость, интонация и скорость речи передают большое количество информации.

Однако бывают ситуации, когда голос может быть менее подходящим для использования:

- производственные шумы могут нарушить взаимодействие с машиной;
- не каждому оператору удобно разговаривать с устройством;
- текстовые сообщения могут быть более привычным способом общения, а люди обычно неохотно принимают изменения;
- нарушается конфиденциальность сообщения [26].

При разработке VUI необходимо учитывать одношаговые задачи, такие как ответ на поисковый запрос, настройка встречи в календаре, выполнение телефонного звонка, воспроизведение песни и т.д. Если сценарии обработки описаны достаточно полно, они могут иметь древовидную структуру, и в ходе обработки такого сценария будет рассматриваться следующий шаг с учетом контекста [26].

При проектировании подобных систем необходимо учитывать, как будет происходить подтверждение ввода [26]:

— трехуровневая уверенность. В этом случае система отклонит текст с первым, самым низким, порогом вероятности и неявным образом подтвердит все, что превышает второй порог вероятности. При этом будет явно запрошено подтверждение информации, которая находится между первым и вторым порогами. Особенно важно явно подтверждать информацию, если цена нераспознавания высока;

— неявное подтверждение: подтверждать команды только неявно, не требуя от пользователя каких-либо действий. Этот способ включает воспроизведение части исходного вопроса, чтобы пользователь мог убедиться, что система правильно распознает ввод;

— подтверждение отсутствия речи — основано на завершении действия, не требующего голосового ответа, например, с помощью визуальных индикаторов. Иным типом подтверждения является характерный звуковой сигнал;

— общее подтверждение. В некоторых диалоговых системах может быть уместно не подтверждать слова пользователя. Это больше относится к диалоговым системам, которые основаны на открытом чате с пользователем. Например, система спрашивает, как кто-то себя чувствует, но не обязательно реагирует на полученный ответ;

— визуальное подтверждение. На мобильных устройствах часто используется визуальное подтверждение. Например, задавая вопрос Google, можно получить как звуковое, так и

визуальное подтверждение. Работа со списком элементов на экране намного эффективнее;

— явное подтверждение. Хотя это может показаться избыточным, в некоторых случаях требуется явное подтверждение команды. Чтобы продолжить взаимодействие, пользователь должен произнести фразу соглашения.

Возможные причины несрабатывания VUI:

— речь не обнаружена;

— речь обнаружена, но не распознана;

— что-то распознано неправильно;

— все распознано правильно, но система неправильно обработала команду.

В промышленных приложениях цена отказа очень высока: не только машины, но и человеческие жизни могут пострадать из-за неправильной схемы обработки ошибок. Таким образом, обработка ошибок — важная часть разработки такой схемы взаимодействия.

Специфика применения голосового интерфейса на предприятии. К предприятиям оборонно-промышленного комплекса предъявляются высокие требования не только с точки зрения производства как такового (к качеству, цене и срокам исполнения), но и с точки зрения безопасности. Такого рода требования рождают свою специфику и свои ограничения. Недавний случай с компанией Garmin [27], которая занимается авиацией, навигацией и носимыми устройствами с использованием навигации, показал, как могут быть уязвимы компании к подобным хакерским атакам. Помимо проверки самих вендоров, которые предоставляют какие-либо технологии, нужно производить независимый аудит, чтобы минимизировать подобные риски.

Кроме очевидных хакерских атак есть менее очевидные проблемы при внедрении голосовых технологий. Устройства, как правило, производят ресурсоемкие операции по переводу голоса в текст, текста в смысл, смысла в команду и обратно на серверах, которые выбрал поставщик данной технологии. Кроме того, многие компании используют полученные запросы для того, чтобы их разметить для улучшения качества работы моделей машинного обучения. Это значит, что создание навыка для уже существующего голосового помощника, как было предложено в работе [3], не является достаточно безопасным, особенно для наукоемких предприятий с передовыми разработками и секретностью.

Таким образом, данные должны циркулировать внутри экосистемы предприятия (предприятий), чтобы гарантировать, что с ними работают только те, кто имеет на то основание и необходимые допуски.

Вышесказанное не отменяет того факта, что такие предприятия могут сотрудничать с частными компаниями, но в этом случае хранение и обработка данных должна осуществляться особым образом.

Помимо этого, как будет показано ниже, специфика предприятия, его шумовые и вибрационные характеристики существенно отличаются [28] от тех, на которые рассчитаны обычные акустические модели, однако этот вопрос требует дополнительного исследования.

В рамках исследования проведен эксперимент (рис. 3), цель которого — измерить качество распознавания речи при условии варьирования шумовых уровней на производстве, согласно нормативным документам [28]. Использовалось следующее оборудование: шумомер UNI-T UT353, источник шума JBL Flip 4, диктофон Olympus WS-331M, измерительная рулетка Fit DIY 17824. Такой эксперимент призван воспроизвести производственный шум, аналогичный шуму на рабочем месте оператора механообрабатывающего оборудования, и проверить, насколько возможно применение подобных технологий. Установлено расстояние, где уровень постоянного шумового давления составлял 50, 60, 65 и 75 дБА, где 50 дБА — это максимально допустимый уровень при проведении научных работ, расчетов и конструирования, а 75 дБА — максимальный уровень для сосредоточенной работы в помещениях с шумным оборудованием [29]. На каждом уровне расстояние от говорящего до записывающего устройства составляло 10 см, как если бы оператор произносил фразы при штатном размещении гарнитуры.

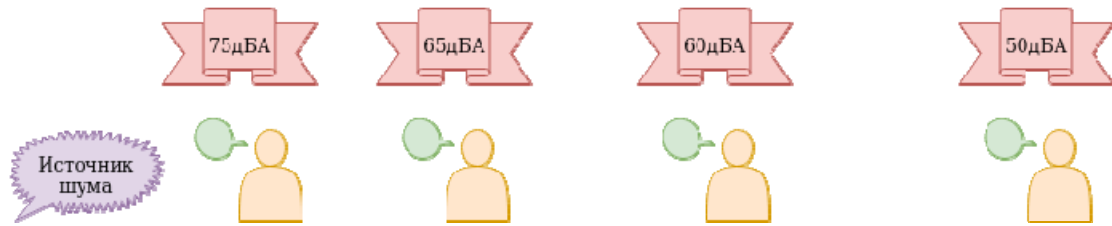


Рис. 3

Рабочая гипотеза состояла в том, что речевой движок от компании Google [30], выбранный для распознавания аудиопотока, будет с достаточно высокой точностью распознавать речь на уровнях 50 и 60 дБА, но на уровне 75 дБА качество может быть существенно хуже.

Выбран набор фраз, который мог бы гипотетически быть реализован для голосового программного управления для станков с ЧПУ, а именно:

- *возврат в референтную точку;*
- *запустить шпиндель;*
- *следующий инструмент;*
- *показать таблицу корректоров;*
- *запустить программу;*
- *следующий кадр;*
- *остановить программу.*

Таким образом, для обозначения голосовых интерфейсов, которые могут быть использованы при исполнении команд на станках с ЧПУ, предлагается применять термин Computer Voice Control (CVC), или голосовое программное управление (ГПУ). Обоснование их применимости, их классификация и назначение являются предметом отдельного изучения.

Вопреки изначальной гипотезе, на всех уровнях качество распознавания было достаточно высоким (цифра соответствует уровню, а список — это транскрибированные команды; полужирным выделены ошибки распознавания).

```
{
  "50": [
    "возврат в референтную точку",
    "запустить в Пензе",
    "следующий инструмент",
    "показать таблицу корректоров",
    "запустить программу",
    "следующий кадр",
    "остановить программу"
  ],
  "60": [
    "возврат референтную точку",
    "запустить шпиндель",
    "следующий инструмент",
    "показать таблицу корректоров",
    "запустить программу",
    "следующий кадр",
    "остановить программу"
  ],
  "65": [
    "возврат в референтную точку",
    "запустить шпиндель",
    "следующий инструмент",
    "показать таблицу корректора",
    "запустить программу",
    "следующий кадр",
  ]
}
```

```

    "остановить программу"
  ],
  "75": [
    "возврат в референтную точку",
    "запустить шпиндель",
    "следующий инструмент",
    "показать таблицу корректоров",
    "запустить программу",
    "следующий кадр",
    "остановить программу"
  ]
}

```

Качество распознавания можно признать приемлемым.

Данный эксперимент, однако, не доказывает, что голосовые модели из открытого доступа можно без доработок и учета специфики внедрять в производство. Эксперимент проведен с рядом допущений и упрощений (звук постоянный, широкополосный и т.д.).

Помимо шумового эксперимента на описанных выше принципах реализован конвейер обработки данных. Вначале произносится команда, и в ходе различных преобразований проходя по цепочке обработки, через компоненты, которые были описаны выше (ASR, NLU), она исполняется на станке.

Эксперимент реализован в три этапа:

- 1) локальная отладка, написание программных компонентов, сведение в цепочку;
- 2) отладка на лабораторном оборудовании в Университете ИТМО, с управляющей стойкой Smoothieboard [31];
- 3) отладка на предприятии компании БалтСистем на стойке NC-310 в сотрудничестве с инженерами компании.

Результатом эксперимента стала апробация выстроенного конвейера для управления станком, например, выход в референтную точку или перемещение по координатам. На рис. 4 представлен вывод на экран этапов работы программы, на рис. 5 — отображение команды на стойке станка.

```

Say :-
End of speech
переместить на икс двадцать игрек тридцать пять
ShallowClassifier predicted move {'X': '20.0', 'Y': '35.0'} for "переместить на икс двадцать игрек тридцать пять"
Please confirm execution "move" command
Say :-
End of speech
DefaultAdapter transformed move to G1 X20.0 Y35.0
Sent:
G1 X20.0 Y35.0

```

Рис. 4



Рис. 5

Хотя ученые исследовали тему управления станком с помощью голосового интерфейса [32—36], универсального стандарта для взаимодействия через этот канал не существует. Скорее всего, это связано с тем, что конечные пользователи не видят практической пользы сценариев реализации взаимодействия с использованием такого интерфейса. Для успешной реализации голосовой интерфейс должен быть не только способом закрыть/открыть дверь или переместить станок в нужные координаты [32—36], но и быть полноценным интерфейсом для управления станком с ЧПУ, если не полностью дублировать или расширять его функционал, то хотя бы его дополняющий. Практическая ценность заключается не в исполнении конкретной команды, что является выполнимой задачей при текущем развитии технологий, а в построении комплексных сценариев взаимодействия и приобретении практической потребительской ценности. Обычно такие исследования подтверждают гипотезу об использовании интерфейса для конкретного оборудования в качестве доказательства концепции [32—36].

Детали реализации выходят за рамки настоящей работы, стоит, однако, отметить, что выстроенный процесс позволяет выполнить не только G-коды, но и любые другие, которые разработчики стойки допускают как валидные: справочные, функциональные. Кроме того, существует способ сделать такой интерфейс универсальным для разного типа оборудования и разных производителей управляющих стоек, что открывает перспективы для встраивания интерфейса в качестве комплементарного для более удобного и безопасного взаимодействия со станком.

Заключение. Когда предприятие рассматривает возможность использования голосового интерфейса, необходимо:

— учитывать наличие любых других форм взаимодействия с оборудованием. Важно использовать графический интерфейс и голосовой ввод вместе, если есть такая возможность;

— учитывать уровень шума на рабочем месте при таком подходе. Шумное пространство накладывает свои ограничения, и чтобы реализовывать этот вариант, необходимо использовать речевые движки с высоким качеством распознавания даже в шумной среде;

— оценить, является ли голосовое управление более быстрым по сравнению с текущими подходами. Хотя текстовые сообщения могут быть не так удобны для оператора, нажатие комбинации кнопок может быть более эффективным, чем голосовой ввод;

— определить, упрощает ли этот интерфейс процесс управления. Подача команд без помощи рук может сделать управление более удобным и безопасным. Кроме того, вербальное общение – это естественный способ передачи данных и команд, противоположный графическому пользовательскому интерфейсу.

Быстрое развитие Индустрии 4.0, включая интернет вещей и облачные вычисления, вряд ли устранил человеческие рабочие места (по крайней мере, на ранних этапах). Однако спрос на удобные способы взаимодействия между людьми и машинами определенно будет расти, и именно голос обеспечит более естественный и простой способ взаимодействия людей и машин.

СПИСОК ЛИТЕРАТУРЫ

1. Alexa and Google Assistant have a problem: People aren't sticking with voice apps they try [Электронный ресурс]: <<https://www.vox.com/2017/1/23/14340966/voicelabs-report-alex-google-assistant-echo-apps-discovery-problem>> (01.05.2020).
2. Amazon Alexa leader: COVID-19 has sparked 'a huge increase in the use of voice in the home' [Электронный ресурс]: <<https://www.geekwire.com/2020/amazon-alexa-leader-covid-19-sparked-huge-increase-use-voice-home/>>. (01.05.2020).

3. *Afanasev M. Ya., Fedosov Yu. V., Andreev Yu. S., Krylova A. A., Shorokhov S. A., Zimenko K. V., Kolesnikov M. V.* A concept for integration of voice assistant and modular cyber-physical production system // 2019 IEEE 17th Intern. Conf. on Industrial Informatics (INDIN). Helsinki, Finland, 2019. P. 27—32. DOI: 10.1109/INDIN41052.2019.8972015.
4. *Rüßmann M., Lorenz M., Gerbert P., Waldner M.* Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries. Boston: The Boston Consulting Group, 2015.
5. *Borisov O. I., Gromov V. S., Kolyubin S. A., Pyrkin A. A., Bobtsov A. A., Salikhov V. I., Petranovsky I. V.* Human-free robotic automation of industrial operations // IECON 2016. 42nd Ann. Conf. of the IEEE Industrial Electronics Society. Florence, Italy, 2016. P. 6867—6872. DOI: 10.1109/IECON.2016.7793922.
6. *Borisov O. I., Gromov V. S., Kolyubin S. A., Pyrkin A. A., Dema N. Y., Salikhov V. I., Bobtsov A. A.* Case study on human-free water heaters production for Industry 4.0 // 2018 IEEE Industrial Cyber-Physical Systems (ICPS). St. Petersburg, Russia, 2018. P. 369—374. DOI: 10.1109/ICPHYS.2018.8387686.
7. *Nelles J., Kuz S., Mertens A., Schlick C. M.* Human-centered design of assistance systems for production planning and control: the role of the human in Industry 4.0 // 2016 IEEE Intern. Conf. on Industrial Technology (ICIT). Taipei, Taiwan, 2016. P. 2099—2104. DOI: 10.1109/ICIT.2016.7475093.
8. *Romero D., Bernus P., Noran O., Stahre J., Fast-Berglund Å.* The operator 4.0: human cyber-physical systems and adaptive automation towards human-automation symbiosis work systems // Advances in Production Management Systems. Initiatives for a Sustainable World. Iguassu Fall, Brazil, 2016. Vol. 488. P. 677—686. DOI: 10.1007/978-3-319-51133-7_80.
9. *Magrini E., Ferraguti F., Ronga A. J., Pini F., De Luca A., Leali F.* Human-robot coexistence and interaction in open industrial cells // Robotics and Computer-Integrated Manufacturing. 2020. Vol. 61. P. 2—19. DOI: 10.1016/j.rcim.2019.101846.
10. *Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F.* Digital twin-driven product design, manufacturing and service with Big Data // The Intern. J. of Advanced Manufacturing. 2018. Vol. 94. P. 3563—3576. DOI: 10.1007/s00170-017-0233-1.
11. *Damiani L., Demartini M., Guizzi G., Revetria R., Tonelli F.* Augmented and virtual reality applications in industrial systems: a qualitative review towards the Industry 4.0 era // IFAC-PapersOnLine. 2018. Vol. 51, № 11. P. 624—630. DOI: 10.1016/j.ifacol.2018.08.388.
12. *Lorenz M., Spranger M., Riedel T., Pürzel F., Wittstock V., Klimant P.* CAD to VR — a methodology for the automated conversion of kinematic CAD models to virtual reality // Procedia CIRP. 2016. Vol. 41. P. 358—363. DOI: 10.1016/j.procir.2015.12.115.
13. *Yan J., Meng Y., Lu L., Li L.* Industrial Big Data in an Industry 4.0 environment: challenges, schemes, and applications for predictive maintenance // IEEE Access. 2017. Vol. 5. P. 23484—23491. DOI: 10.1109/ACCESS.2017.2765544.
14. *Sittón I., Rodríguez S.* Pattern extraction for the design of predictive models in Industry 4.0. Trends in cyber-physical multi-agent systems // The PAAMS Collection — 15th Intern. Conf. 2017. P. 258—261. DOI: 10.1007/978-3-319-61578-3_31.
15. *Li Z., Wang Y., Wang K.-S.* Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario // Advances in Manufacturing. 2017. Vol. 5, N 4. P. 377—387. DOI: 10.1007/s40436-017-0203-8.
16. US Voice Assistant Users 2019 [Электронный ресурс]: <<https://www.emarketer.com/content/us-voice-assistant-users-2019>>. (01.05.2020).
17. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone [Электронный ресурс]: <<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>> (01.05.2020).
18. What is Alexa for Hospitality? [Электронный ресурс]: <<https://developer.amazon.com/en-US/alexa/alexa-for-hospitality>>. (01.05.2020).
19. *Brown T. B., Mann B., Ryder N., Subbiah M. et al.* Language models are few-shot learners. OpenAI. 2020 [Электронный ресурс]: <<https://arxiv.org/pdf/2005.14165.pdf>>.
20. *Kildal J., Tellaeché A., Fernández I., Maurtua I.* Potential users' key concerns and expectations for the adoption of cobots // Procedia CIRP. 2018. Vol. 72. P. 21—26. DOI: 10.1016/j.procir.2018.03.104.

21. *Gomes O., Pereira S.* On the economic consequences of automation and robotics // *J. of Economic and Administrative Sciences*. 2019. Vol. 36, N 2. P. 134—153. DOI: 10.1108/JEAS-04-2018-0049.
22. *Brettel M., Friederichsen N., Keller M.* How virtualization decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective // *Intern. J. of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*. 2014. Vol. 8, N 1. P. 36—37.
23. Write once, run anywhere? [Электронный ресурс]: <<https://www.computerweekly.com/feature/Write-once-run-anywhere>>. (01.05.2020).
24. Inside Sundar Pichai's Plan To Put AI Everywhere [Электронный ресурс]: <<https://www.forbes.com/sites/miguelhelft/2016/05/18/inside-sundar-pichais-plan-to-put-ai-everywhere/>>. (01.05.2020).
25. Connolly E. Principles of Bot Design. 2016 [Электронный ресурс]: <<https://www.intercom.com/blog/principles-bot-design/>>.
26. *Pearl C.* Designing Voice User Interfaces. California: O'Reilly Media, 2016.
27. Garmin reportedly paid hackers a multimillion dollar ransom to recover files after a cyberattack that left their services offline for several days last month [Электронный ресурс]: <<https://www.businessinsider.com/garmin-paid-multimillion-dollar-ransom-to-hackers-report-2020-8?op=1>> (01.05.2020).
28. Главный государственный санитарный врач Российской Федерации Постановление от 21 июня 2016 года N 81 Об утверждении СанПиН 2.2.4.3359-16 „Санитарно-эпидемиологические требования к физическим факторам на рабочих местах“ [Электронный ресурс]: <<http://docs.cntd.ru/document/420362948>>. (01.05.2020).
29. Industrial Sounds [Электронный ресурс]: <<http://soundbible.com/610-Industrial-Sounds.html>>. (01.05.2020).
30. Speech-to-Text [Электронный ресурс]: <<https://cloud.google.com/speech-to-text/>> (01.05.2020.)
31. Smoothieboards [Электронный ресурс]: <<https://smoothieware.org/smoothieboard>>. (01.05.2020).
32. *Norberto Pires J.* Robot-by-voice: experiments on commanding an industrial robot using the human voice // *Industrial Robot: An International Journal*. 2005. Vol. 32, N 6. P. 505—511.
33. *Rogowski A.* Robotized cell remote control using voice commands in natural language // 15th Intern. Conf. on Methods and Models in Automation and Robotics. 2010. P. 383—386.
34. *Rogowski A.* Industrially oriented voice control system // *Robotics and Computer-Integrated Manufacturing*. 2012. Vol. 28, N 3. P. 303—315.
35. *Rogowski A.* Web-based remote voice control of robotized cells // *Robotics and Computer-Integrated Manufacturing*. 2013. Vol. 29, N 4. P. 77—89.
36. *Jing L., Ting X., Nanyan S.* Research on the voice control and its audio signal processing in flexible manufacturing cell // *Sensors and Transducers*. 2013. Vol. 25. P. 91—97.

Сведения об авторах**Алексей Викторович Катриди**— аспирант; Университет ИТМО; факультет систем управления и робототехники; E-mail: katridi@yandex.ru**Юрий Валерьевич Федосов**— канд. техн. наук, доцент; Университет ИТМО; факультет систем управления и робототехники; инженер; E-mail: yf01@yandex.ruПоступила в редакцию
30.11.2020 г.**Ссылка для цитирования:** Катриди А. В., Федосов Ю. В. Особенности применения голосовых интерфейсов на производстве // *Изв. вузов. Приборостроение*. 2021. Т. 64, № 6. С. 485—497.

FEATURES OF THE USE OF VOICE INTERFACES IN INDUSTRIAL PRODUCTION

A. V. Katridi, Yu. V. Fedosov

ITMO University, 197101, St. Petersburg, Russia
E-mail: katridi@yandex.ru

Over the past decade, the quality of natural language processing has improved significantly due to enhancement in both software and hardware. Modern technologies demonstrate high results in all areas - from automatic speech recognition to text-to-speech and natural language understanding. This allowed a large number of companies to significantly increase the speed of service delivery, quality and usability of products in general. However, there are many other not yet used applications of voice interfaces, and one of them is the application of such technologies in industrial production within the framework of modern enterprises. There are a number of challenges that need to be addressed when using voice technology. The basic components of natural language processing are presented, the features of their application, as well as approaches to creating an effective and at the same time simple voice user interface. Some features of industrial design and production are considered, which should be taken into account when developing such systems.

Keywords: machine learning, natural language processing, Industry 4.0, industrial engineering, voice user interface, computer voice control

REFERENCES

1. *Alexa and Google Assistant have a problem: People aren't sticking with voice apps they try*, <https://www.vox.com/2017/1/23/14340966/voicelabs-report-alexa-google-assistant-echo-apps-discovery-problem>, 01.05.2020.
2. *Amazon Alexa leader: COVID-19 has sparked 'a huge increase in the use of voice in the home'*, <https://www.geekwire.com/2020/amazon-alexa-leader-covid-19-sparked-huge-increase-use-voice-home/>, 01.05.2020.
3. Afanasev M.Ya., Fedosov Yu.V., Andreev Yu.S., Krylova A.A., Shorokhov S.A., Zimenko K.V., Kolesnikov M.V. 2019 *IEEE 17th Intern. Conf. on Industrial Informatics (INDIN)*, Helsinki, Finland, 2019, pp. 27–32. DOI: 10.1109/INDIN41052.2019.8972015.
4. Rüßmann M., Lorenz M., Gerbert P., Waldner M. *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*, Boston, The Boston Consulting Group, 2015.
5. Borisov O.I., Gromov V.S., Kolyubin S.A., Pyrkin A.A., Bobtsov A.A., Salikhov V.I., Petranovsky I.V. *IECON 2016. 42nd Ann. Conf. of the IEEE Industrial Electronics Society*, Florence, Italy, 2016, pp. 6867–6872, DOI: 10.1109/IECON.2016.7793922.
6. Borisov O.I., Gromov V.S., Kolyubin S.A., Pyrkin A.A., Dema N.Y., Salikhov V.I., Bobtsov A.A. 2018 *IEEE Industrial Cyber-Physical Systems (ICPS)*, St. Petersburg, Russia, 2018, pp. 369–374, DOI: 10.1109/ICPHYS.2018.8387686.
7. Nelles J., Kuz S., Mertens A., Schlick C.M. 2016 *IEEE Intern. Conf. on Industrial Technology (ICIT)*, Taipei, Taiwan, 2016, pp. 2099–2104, DOI: 10.1109/ICIT.2016.7475093.
8. Romero D., Bernus P., Noran O., Stahre J., Fast-Berglund Å. *Advances in Production Management Systems. Initiatives for a Sustainable World*, Iguassu Fall, Brazil, 2016, vol. 488, pp. 677–686, DOI: 10.1007/978-3-319-51133-7_80.
9. Magrini E., Ferraguti F., Ronga A. J., Pini F., De Luca A., Leali F. *Robotics and Computer-Integrated Manufacturing*, 2020, vol. 61, pp. 2–19, DOI: 10.1016/j.rcim.2019.101846.
10. Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F. *The International Journal of Advanced Manufacturing*, 2018, vol. 94, pp. 3563–3576, DOI: 10.1007/s00170-017-0233-1.
11. Damiani L., Demartini M., Guizzi G., Revetria R., Tonelli F. *IFAC-PapersOnLine*, 2018, no. 11(51), pp. 624–630, DOI: 10.1016/j.ifacol.2018.08.388.
12. Lorenz M., Spranger M., Riedel T., Pürzel F., Wittstock V., Klimant P. *Procedia CIRP*, 2016, vol. 41, pp. 358–363, DOI: 10.1016/j.procir.2015.12.115.
13. Yan J., Meng Y., Lu L., Li L. *IEEE Access*, 2017, vol. 5, pp. 23484–23491, DOI: 10.1109/ACCESS.2017.2765544.
14. Sittón I., Rodríguez S. *The PAAMS Collection – 15th International Conference*, 2017, pp. 258–261, DOI: 10.1007/978-3-319-61578-3_31.
15. Li Z., Wang Y., Wang K.-S. *Advances in Manufacturing*, 2017, no. 4(5), pp. 377–387, DOI: 10.1007/s40436-017-0203-8.
16. *US Voice Assistant Users 2019*, <https://www.emarketer.com/content/us-voice-assistant-users-2019>, 01.05.2020.
17. *Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*, <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>, 01.05.2020.
18. *What is Alexa for Hospitality?* <https://developer.amazon.com/en-US/alexa/alexa-for-hospitality>, 01.05.2020.
19. Brown T.B., Mann B., Ryder N., Subbiah M. et al. *Language models are few-shot learners*. *OpenAI*, 2020, <https://arxiv.org/pdf/2005.14165.pdf>.
20. Kildal J., Tellaeche A., Fernández I., Maurtua I. *Procedia CIRP*, 2018, vol. 72, pp. 21–26, DOI:

- 10.1016/j.procir.2018.03.104.
21. Gomes O., Pereira S. *Journal of Economic and Administrative Sciences*, 2019, no. 2(36), pp. 134–153, DOI: 10.1108/JEAS-04-2018-0049.
22. Brettel M., Friederichsen N., Keller M. *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering*, 2014, no. 1(8), pp. 36–37.
23. *Write once, run anywhere?* <https://www.computerweekly.com/feature/Write-once-run-anywhere>, 01.05.2020.
24. *Inside Sundar Pichai's Plan To Put AI Everywhere*, <https://www.forbes.com/sites/miguelhelft/2016/05/18/inside-sundar-pichais-plan-to-put-ai-everywhere/>, 01.05.2020.
25. Connolly E. *Principles of Bot Design*, 2016, <https://www.intercom.com/blog/principles-bot-design/>.
26. Pearl C. *Designing Voice User Interfaces*, California, O'Reilly Media, 2016.
27. *Garmin reportedly paid hackers a multimillion dollar ransom to recover files after a cyberattack that left their services offline for several days last month*, <https://www.businessinsider.com/garmin-paid-multimillion-dollar-ransom-to-hackers-report-2020-8?op=1>, 01.05.2020.
28. <http://docs.cntd.ru/document/420362948>, 01.05.2020.
29. *Industrial Sounds*, <http://soundbible.com/610-Industrial-Sounds.html>, 01.05.2020.
30. *Speech-to-Text*, <https://cloud.google.com/speech-to-text/>, 01.05.2020.
31. *Smoothieboards*, <https://smoothieware.org/smoothieboard>, 01.05.2020.
32. Norberto Pires J. *Industrial Robot: An International Journal*, 2005, no. 6(32), pp. 505–511.
33. Rogowski A. *15th International Conference on Methods and Models in Automation and Robotics*, 2010, pp. 383–386.
34. Rogowski A. *Robotics and Computer-Integrated Manufacturing*, 2012, no. 3(28), pp. 303–315.
35. Rogowski A. *Robotics and Computer-Integrated Manufacturing*, 2013, no. 4(29), pp. 77–89.
36. Jing L., Ting X., Nanyan S. *Sensors and Transducers*, 2013, vol. 25, pp. 91–97.

Data on authors

- Aleksey V. Katridi** — Post-Graduate Student; ITMO University, Faculty of Control Systems and Robotics; E-mail: katridi@yandex.ru
- Yury V. Fedosov** — PhD, Associate Professor; ITMO University, Faculty of Control Systems and Robotics; Engineer; E-mail: yf01@yandex.ru

For citation: Katridi A. V., Fedosov Yu. V. Features of the use of voice interfaces in industrial production. *Journal of Instrument Engineering*. 2021. Vol. 64, N 6. P. 485—497 (in Russian).

DOI: 10.17586/0021-3454-2021-64-6-485-497