



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 101 за 2022 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

В.А. Карташев, В.В. Карташев

**Диалог человека и работа по
заданному сценарию**

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



Рекомендуемая форма библиографической ссылки: Карташев В.А., Карташев В.В. Диалог человека и работа по заданному сценарию // Препринты ИПМ им. М.В.Келдыша. 2022. № 101. 14 с. <https://doi.org/10.20948/prepr-2022-101>
<https://library.keldysh.ru/preprint.asp?id=2022-101>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

В.А. Карташев, В.В. Карташев

**Диалог человека и работа
по заданному сценарию**

Москва — 2022

Карташев В.А., Карташев В.В.

Диалог человека и робота по заданному сценарию

В настоящее время широкое распространение получила технология вербальных диалогов человека с роботом по свободному и заданному сценариям. Обычно в этом диалоге человек не наблюдает внешность робота, что создает дискомфортные условия для ведения диалога человеком. В описываемом аппаратно-программном комплексе человек обменивается фразами с роботом, наблюдая на экране дисплея его анимированное изображение. Эксперименты показали, что даже достаточно простые средства анимации позволяют сделать диалог более естественным по сравнению с демонстрацией статичного изображения робота.

Ключевые слова: человеко-машинное взаимодействие

Vladimir Alekseevich Kartashev, Vsevolod Vladimirovich Kartashev

Scripted communication between a human and a robot

Currently the technology providing the means both of arbitrary and scripted verbal communication between a person and a robot has become widespread. Usually in this communication, a person does not observe the appearance of the robot, which creates uncomfortable conditions for a human to conduct a dialogue. In the described hardware and software, a person exchanges phrases with a robot and watches its animated image on the display screen. Experiments have shown that even fairly simple animation tools make it possible to make a communication much more natural compared to the one comprising the demonstration of a static image of a robot.

Key words: human-robot interaction

Введение

В настоящее время достаточно широкое распространение получила технология вербальных диалогов человека с роботом. Это общение может осуществляться в виде интеллектуального диалога с непредсказуемым сценарием и по сценарию, который задан заранее.

В качестве примера диалога по непредсказуемому сценарию можно привести диалоги с роботом NAO [1]. Этот автономный программируемый человекоподобный робот, производится компанией Aldebaran Robotics с 2012 года. Робот поддерживает беседу на произвольные темы, используя информацию, которая содержится в интернете.

Значительное распространение получила технология диалога по заданному сценарию. Примерами таких диалогов являются передача данных о расходе воды в квартире в Москве [2] и интерактивный помощник Сбербанка [3].

В статье [4] рассматриваются голосовые сервисы для бизнеса. Несмотря на то, что такие возможности появились сравнительно недавно, их набор и количество фирм, работающих в этой области, уже достаточно велик. В статье рассматриваются возможности голосовых чат-ботов по управлению звонками, поддержке голосовых меню и баз данных совершенных разговоров с клиентами, быстрым уведомлениям, использованию умной телефонии (направление звонков по настроенному сценарию с целью уменьшить времени ожидания), обзвону роботом клиентов и проведению опросов.

Общение человека с роботами на вербальном уровне позволяет заменить человека-оператора в достаточно большом количестве сценариев.

Следует заметить, что произношение слов и построение фраз, которые использует робот, заметно отличаются от беседы между людьми. Это обстоятельство негативным образом влияет на оценку качества общения с роботом. Одним из способов улучшения впечатления от диалога является использование визуальной информации о говорящем.

Визуальная составляющая беседы является существенной и для общения между людьми [5], так как позволяет собеседникам отвлечься от особенностей тембра голоса, произношения и быстрее сосредоточиться на сути беседы и нюансах осуждаемой проблемы. В описываемой работе визуализация говорящего робота позволяет отвлечь внимание человека от недостатков вербальной составляющей беседы.

В настоящее время разработанные программные средства способны достичь достаточно высокого качества визуализации образа говорящего человека [6]. В проведенном исследовании делается попытка увеличить субъективную оценку диалога с помощью достаточно простых решений, которые позволяют уменьшить концентрацию внимания на ошибках речи робота за счет отвлечения его на восприятие графического образа.

В работе описывается технологии создания мимики и речи робота, подготовка и управление диалогом.

Предполагается, что диалог воспроизводится программным пакетом PowerPoint. Выбор пакета для презентации обусловлен широтой его применения и сделать возможным оживление презентаций за счет включения в них диалогов с роботом.

Выполненное исследование показало, что даже сравнительно простые способы визуализации мимики говорящего робота позволяют уменьшить концентрацию внимания на недостатках синтеза речи и усиливают впечатления от диалога. Психологические аспекты такого результата, по-видимому, связаны с тем, что человек не успевает одновременно внимательно следить и за точностью речи, и за динамическими изменениями наблюдаемого образа.

Структура аппаратно-программного комплекса

Структура аппаратно-программного комплекса для воспроизведения речи робота показана на рис. 1. Комплекс включает в себя 2 компьютера с дисплеями и клавиатурами и модуль управления диалогом. Лицо, которое участвует в диалоге с роботом, в дальнейшем называется актером.

Диалог разделен на реплики. Репликой называется элемент диалога, в котором произносит текст одна из сторон.

Управляющий компьютер озвучивает реплики робота и используется для программирования модуля управления диалогом. Воспроизведение речи осуществляется программой PowerPoint (на рис. 1 она указана в виде сокращения PP).

В силу того, что слайд всегда содержит некоторый графический образ и/или текст, имеется возможность вывести на экран дисплея текст, который воспроизводит актер (эта функция управляющего компьютера названа Телесуфлер). Телесуфлер позволяет уменьшить требования к степени подготовки человека к диалогу.

Компьютер 1 визуализирует лицо робота в процессе диалога путем показа слайдов с мимикой робота программой PowerPoint.

Воспроизведение слайдов в нужном порядке осуществляется модулем управления путем эмуляции нажатия кнопки Page Down на клавиатуре управляющего компьютера и кнопок Page Up и Page Down на клавиатуре компьютера 1.

Слайды разбиты на группы по два в каждом. Первый слайд в группе содержит изображение с открытым ртом, второй – с закрытым ртом. При проигрывании реплики по очереди показываются первый и второй слайд пары. Для этого модуль управления эмулирует поочередное нажатие клавиш Page Up и Page Down на компьютере 1.

При последовательной демонстрации слайдов с открытым и закрытым ртом у зрителя возникает ощущение, что речь робота исходит из его рта. Использование слайдов с движением глаз позволяет усилить ощущения, что человек общается с живым существом.

Модуль управления диалогом включает микроконтроллерную платформу Arduino и кнопку «Пуск», которая переводит систему управления в режим воспроизведения реплики робота. Кнопку нажимает участник диалога или оператор в конце реплики, произнесенной актером.

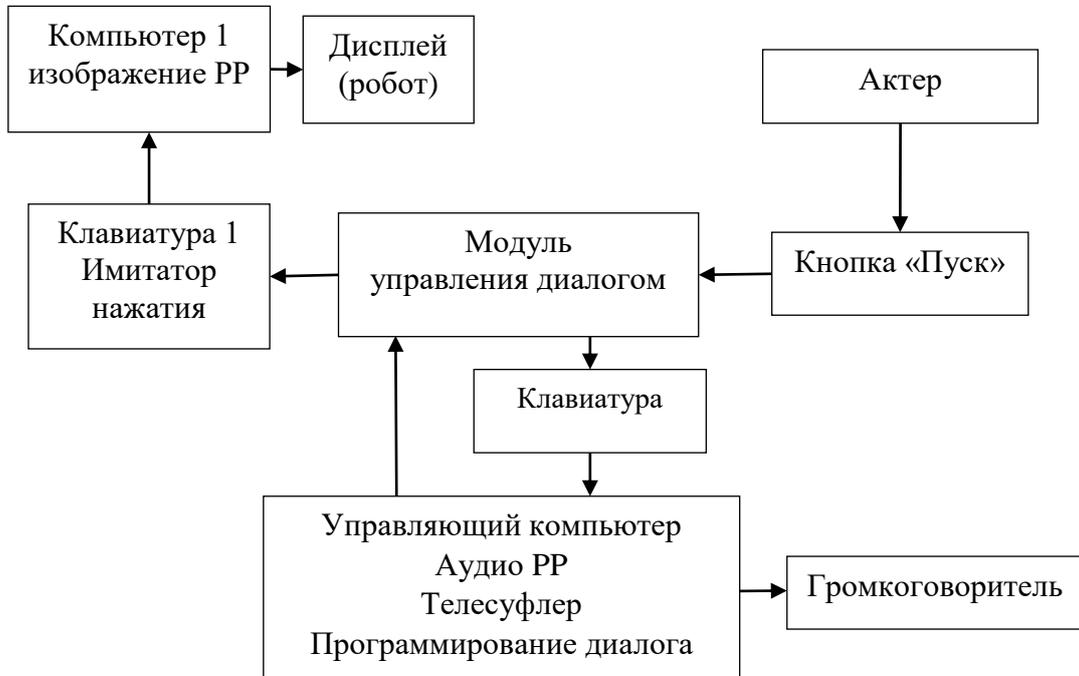


Рис. 1. Структура аппаратно-программного комплекса

Управление репликами робота

Речь робота воспроизводится управляющим компьютером. Визуализация его лица осуществляется компьютером 1. При воспроизведении реплики робота задействованы оба компьютера.

Алгоритм управления воспроизведением реплики робота показан на рис. 2. При произношении реплики слайды с изображением лица робота с открытым и закрытым ртом показываются через каждые 0,5 с.

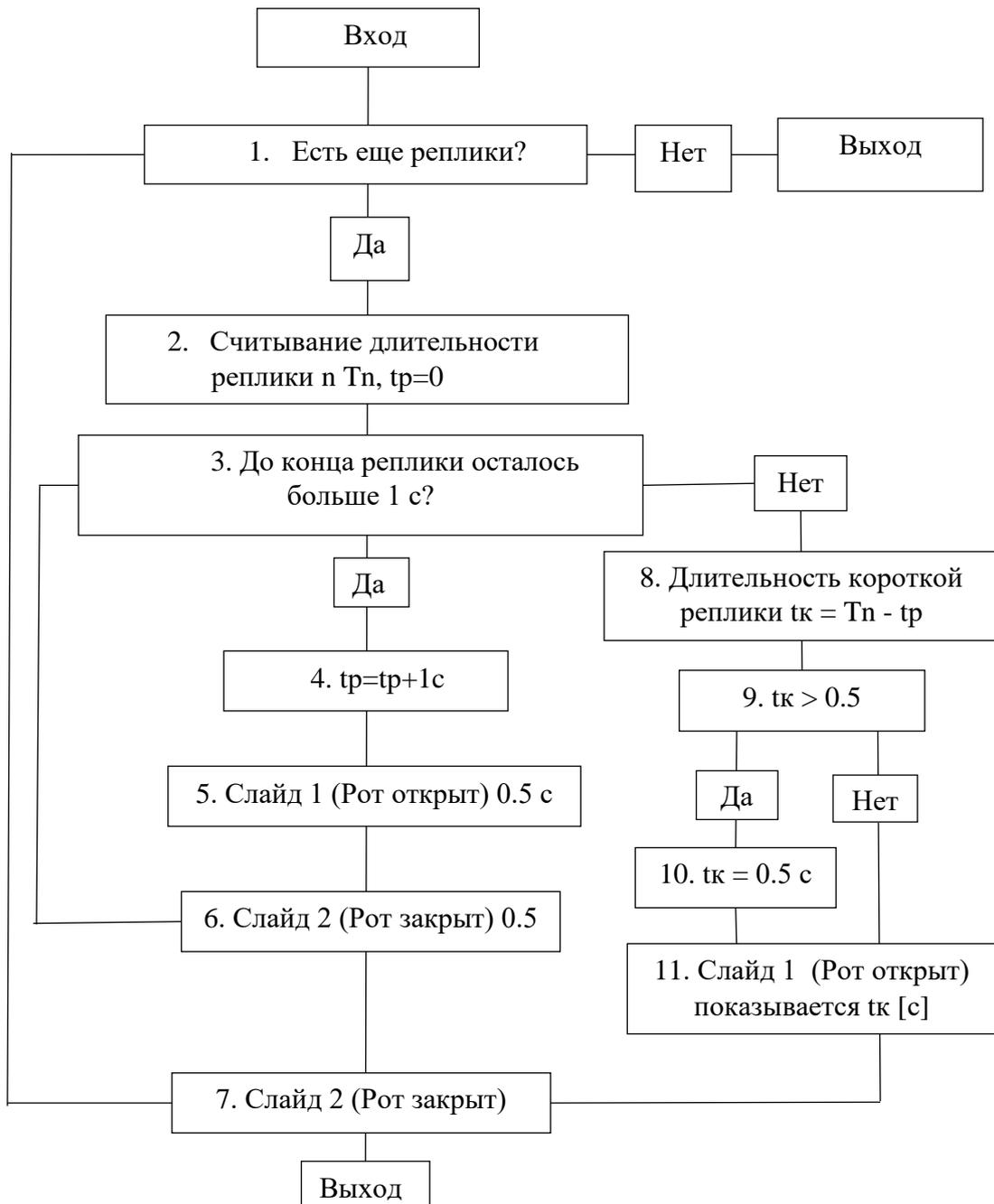


Рис. 2. Алгоритм управления воспроизведением

Частота смены состояния рта выбрана экспериментально. Она соответствует естественной частоте открывания рта человеком при разговоре, и поэтому визуализация речи с выбранной частотой смены кадров производит хорошее впечатление на зрителя в достаточно большом диапазоне эмоциональных состояний говорящего.

Реплика всегда начинается с изображения открытого рта и всегда заканчивается изображением закрытого рта. Два соседних слайда с открытым и закрытым ртом в дальнейшем назовем циклом визуализации. Цикл

визуализации назовем стандартным, если фазы открытого и закрытого рта делятся по 0.5 с.

Перед произнесением реплики с номером n в переменную T_n в блоке 2 алгоритма заносится ее длительность. Длительность реплик хранится в массиве, который заполняется при подготовке диалога.

Если цикл визуализации до конца реплики укладывается полностью (блок 3), то к счетчику времени, прошедшего с начала реплики, добавляется 1 с (блок 4) и выполняется один шаг визуализации (блоки 5 и 6).

Если промежуток времени до конца реплики не позволяет выполнить стандартный цикл визуализации, то в блоке 9 проверяется условие, больше ли оставшийся промежуток времени 0.5 с. Если это так, то в течение 0.5 с показывается слайд с открытым (блок 11), затем показывается слайд с закрытым ртом (блок 7).

Если в блоке 9 установлено, что до конца реплики меньше, чем 0.5 с, то в течение оставшегося времени показывается слайд с открытым ртом (блок 11).

Реплика завершается показом слайда с закрытым ртом (блок 7).

Управление диалогом

Управление диалогом осуществляется модулем управления по алгоритму, показанному на рис. 3.

В начале работы в блоке 1 обнуляется счетчик реплик N_r . Затем актер может произнести свою реплику, если это предусмотрено диалогом (блок 2). После окончания реплики должна быть нажата кнопка «Пуск» (блок 3). Это может сделать актер или ведущий диалог.

Если значение N_r меньше, чем число реплик (блок 4), то значение счетчика увеличивается на 1 (блок 5) и управление передается блоку управления репликой робота (блок 6).

После того, как робот озвучил свою реплику, актер имеет возможность произнести свою реплику (блок 7). По ее окончании должна быть нажата кнопка «Пуск» (блок 8). Это действие приводит к передаче управления в блок 4 для проверки условия, все ли реплики произнесены.

Эксперименты показали, что необходимость нажатия актером или оператором диалога кнопки для озвучивания реплики роботом не вызывает существенных неудобств, в силу того, что задержка в нажатии кнопки не приводит к искажению предписанного хода диалога. Недостаток такого способа управления диалогом состоит в необходимости иметь помощника для ведения диалога, если руки актера должны быть свободны.

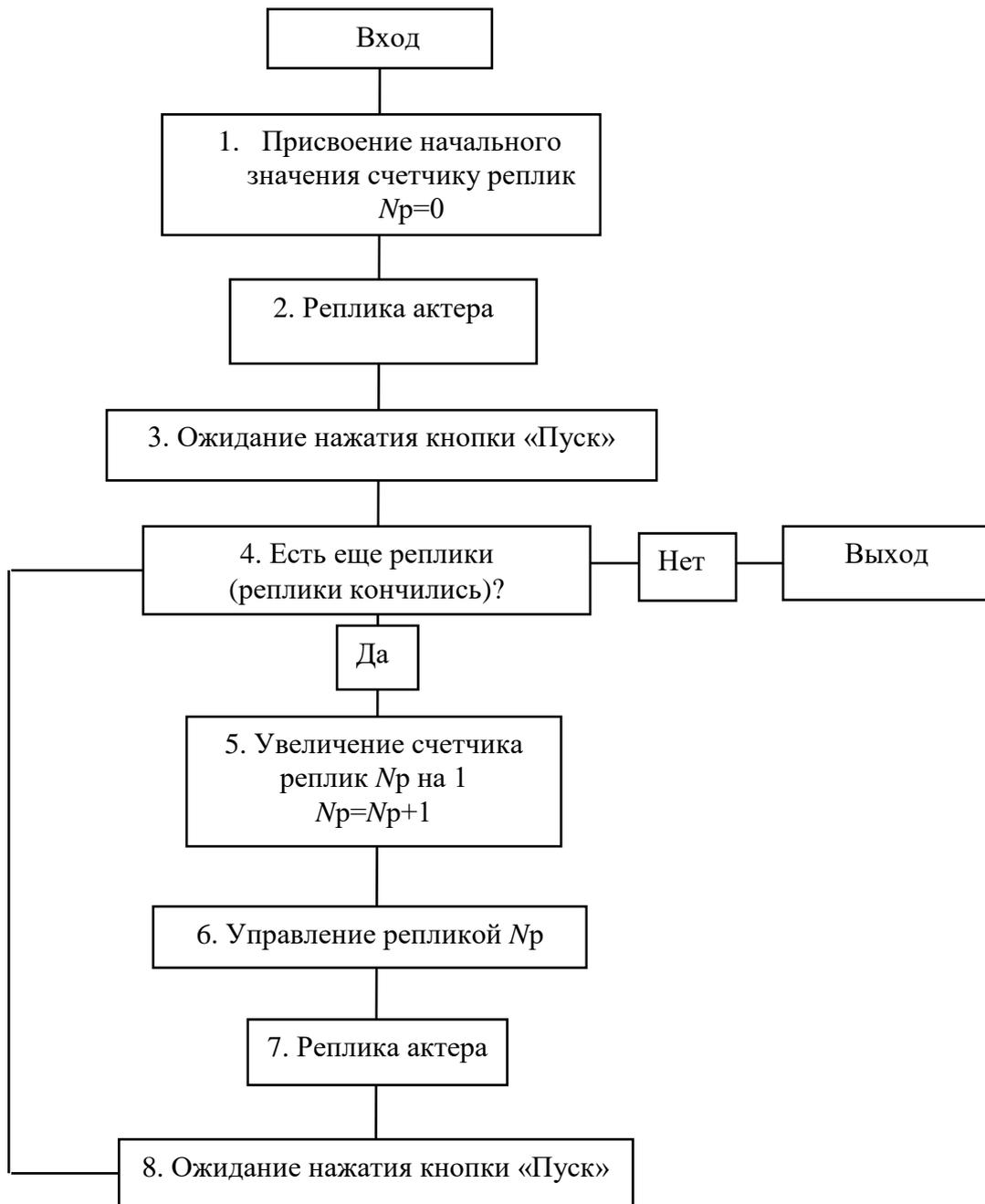


Рис. 3. Алгоритм модуля управления

С целью реализации возможности исполнения диалога без помощника был разработан способ автоматического определения момента окончания реплики человеком. Алгоритм определения окончания показан на рис. 4. Для его реализации требуется микрофон (блок 1), который подключен к аналоговому входу микроконтроллера Arduino. Предполагается также, что длительность реплики известна заранее.

Для произнесения реплики актером отводится заданный при подготовке диалога промежуток времени (блок 2).

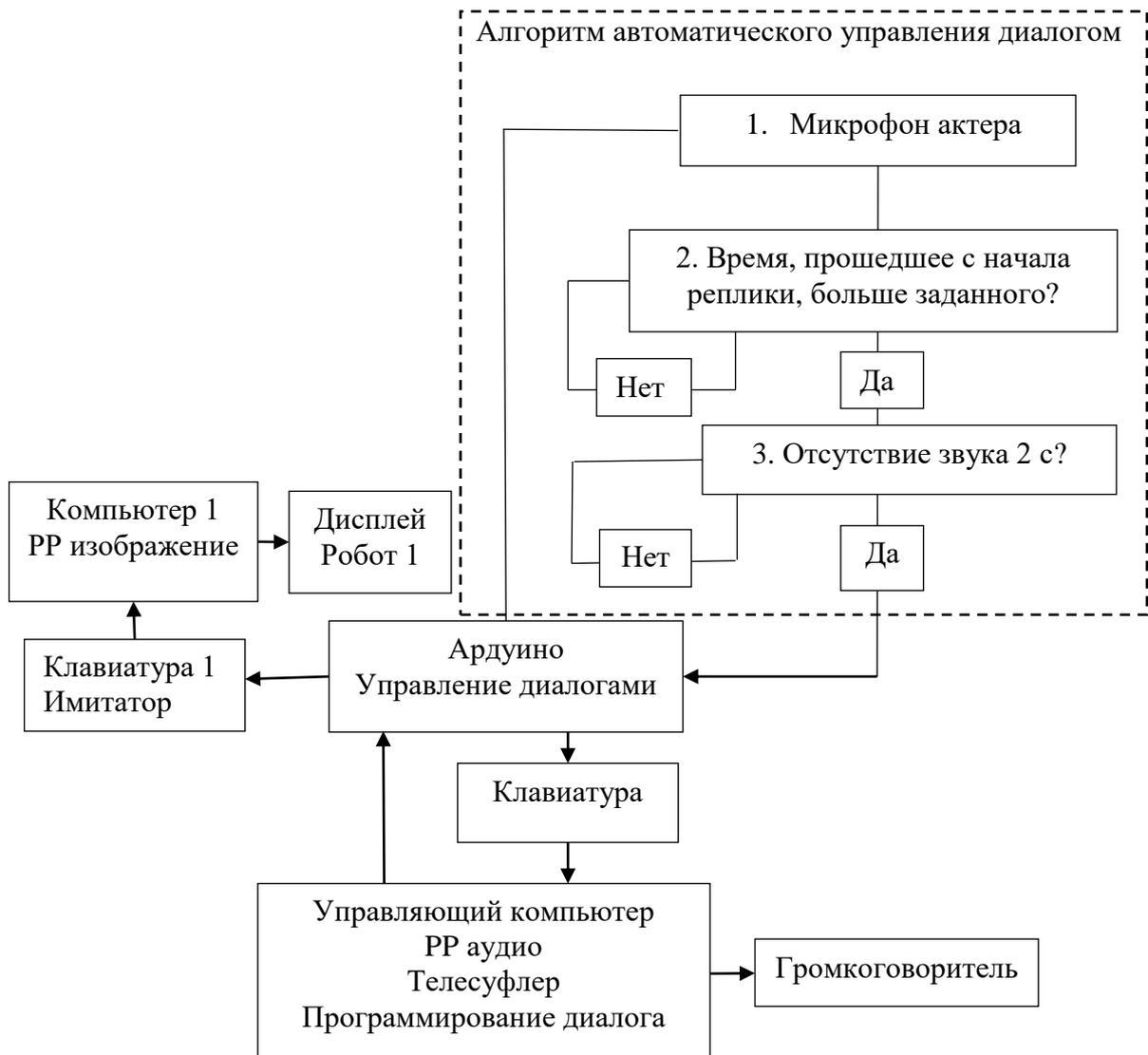


Рис. 4. Алгоритм определения момента окончания реплики

Затем (блок 3) система управления проверяет отсутствие звука от микрофона. Система управления принимает решение о завершении реплики, только если в течение 2 с сигнал от микрофона отсутствует (блок 3). Эксперименты показали, что промежуток времени такой величины не мешает восприятию диалога. При этом имеется возможность не ограничивать актера в длительности реплики и при необходимости произнести более длинную реплику.

Следует заметить, что рассмотренный алгоритм не свободен от недостатков. Основной из них состоит в том, что требуется заранее определить длительности реплик актера.

Попадание посторонних шумов в микрофон может привести к увеличению паузы в диалоге. В некоторых случаях это может оказаться полезным, например, если слушатели выражают свою реакцию аплодисментами или

смехом. В любом случае пауза между репликами не приводит к нарушению порядка воспроизведения диалога.

Подготовка диалога

Создание диалога состоит в подготовке аудио записей реплик робота, слайдов с изображениями лица робота с открытым и закрытым ртом, настройке PowerPoint и системы управления диалогом.

Удобным средством создания аудио реплики является озвучивание текста программным синтезатором речи. Такой способ позволяет значительно уменьшить трудоемкость работы над репликами и оперативно корректировать произносимый текст.

К сожалению, синтезаторы, имеющиеся в свободном доступе, преобразуют текст в речь с легко различимыми на слух отличиями от естественной речи. В них отсутствует возможность делать смысловые ударения, готовить фразы с восклицательной и вопросительной интонацией.

Отмеченные проблемы могут быть разрешены при воспроизведении текста реплики человеком и, при необходимости, дальнейшем редактировании записи аудио редактором.

Существенным недостатком такого способа подготовки диалога является большая трудоемкость отработки смысловой части реплик, в силу того, что даже небольшое изменение текста делает необходимым переозвучивание реплики.

Решение указанной проблемы до некоторой степени упрощается, если вести работу над текстом реплики с использованием синтезатора речи и только после того, как текст отработан, зачитать его вслух.

Текст реплик актера надо перевести речевым синтезатором в отрывки аудиозаписей и загрузить их в слайды презентации. При воспроизведении аудиозаписи реплик робота и актера будут чередоваться, что создает ощущение диалога между ними.

Аудиозаписи реплик проигрываются на управляющем компьютере программой PowerPoint. Условия озвучивания аудиозаписей задаются в настройках. В них надо указать, что реплика проигрывается сразу после начала показа слайда.

Создавать слайды с изображением лица робота удобно в любом доступном графическом векторном редакторе. В описываемом исследовании эта работа выполнялась с помощью программы Inkscape. Существенным достоинством такого редактора является бесплатность и достаточно большое количество доступных функций.

Для разных реплик можно использовать разные изображения лиц робота. Такое разнообразие существенно оживляет диалог.

Созданные изображения целесообразно сохранить в одном из растровых форматов jpg или png. Эксперименты показали, что использование этих

форматов обеспечивает приемлемую скорость визуализации изображений программой PowerPoint.

Настройка программы управления диалогом состоит в указании количества и длительностей реплик в специальном файле. При подготовке диалога этот файл заполняется в среде разработки Arduino IDE и записывается в постоянную память микроконтроллера.

Следует заметить, что разработанный алгоритм воспроизведения реплики обеспечивает достаточно качественное изменение мимики лица говорящего робота, если длина реплики задана с точностью 1 с. Это обстоятельство упрощает подготовку реплик и даже позволяет вносить в них небольшие изменения без перепрограммирования массива длительностей реплик.

Для демонстрации диалога достаточно запустить PowerPoint на компьютере 1 и PowerPoint на управляющем компьютере. Диалог может быть начат сразу после включения кнопки «Пуск» в модуле управления диалогом. Продвижение диалога осуществляется нажатием кнопки «Пуск» после очередной реплики актера.

Аппаратная реализация модуля управления диалогом

Структура модуля управления показана на рис. 5. В нем используется микроконтроллерная платформа Arduino nano (блок 1).

Модуль содержит кнопку «Пуск» (блок 2) для запуска реплик робота. Параллельно кнопке можно подключить кнопку для дистанционного управления диалогом, которую использует актер. Эту возможность можно осуществить с помощью кнопки с проводным удлинителем или с помощью беспроводной передачи сигнала.

Для индикации режима исполнения диалога используются два светодиода. Первый светодиод (блок 3) загорается, когда реплику произносит робот, второй (блок 4) - если реплика произносится актером.

Ко входу модуля управления можно подключить микрофон с микрофонным усилителем (блок 5). Микрофон позволяет вести диалог без нажатия кнопки «Пуск» после того, как актер произнес свою реплику.

Реле (блоки 6-7) используются для имитации нажатия клавиш Page Down и Page Up на клавиатурах управляющего компьютера и компьютера 1, предназначенного для визуализации лица робота.

Программирование микроконтроллера осуществляется в среде программирования Arduino IDE. Загрузка рабочей программы из управляющего компьютера выполняется по кабелю USB, который подключается к разъему (блок 13). По этому кабелю также подается питание в модуль управления. Модуль готов к работе сразу после подачи питания.

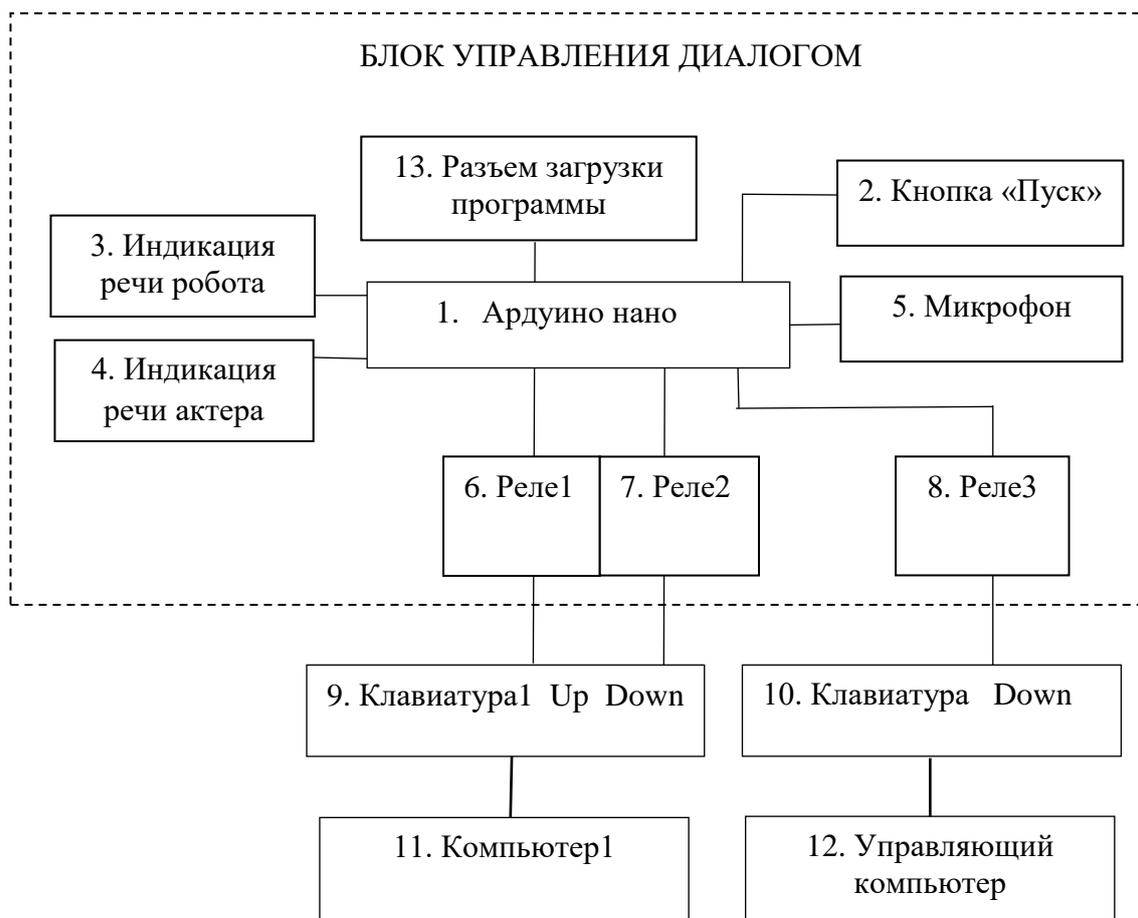


Рис. 5. Структура модуля управления диалогом

Результаты и выводы

В работе описан аппаратно-программный комплекс, который позволяет визуализировать лицо робота в диалогах, происходящих по заданному сценарию, с использованием программного пакета PowerPoint. Возможность введения в презентацию диалога позволяет оживить восприятие представляемого материала слушателями.

Доступные синтезаторы речи для озвучивания текста реплик обладают рядом недостатков. Ошибки произношения, отсутствие смысловых ударений оставляют неприятные ощущения и затрудняют осмысление услышанной человеком реплики в силу того, что внимание концентрируется на ошибках.

Исследование показало, что даже сравнительно простые способы визуализации мимики говорящего робота позволяют уменьшить отрицательные ощущения от ошибок синтезированной речи. Психологические аспекты такого результата, по-видимому, связаны с тем, что человек не успевает одновременно внимательно следить и за точностью речи и изменениями мимики говорящего. В результате этих психологических особенностей восприятия речи человеком внимание слушателя концентрируется на общем смысле услышанного. На рис.

6 показаны выражения лица и ротовая артикуляция во время диалога по сценарию.

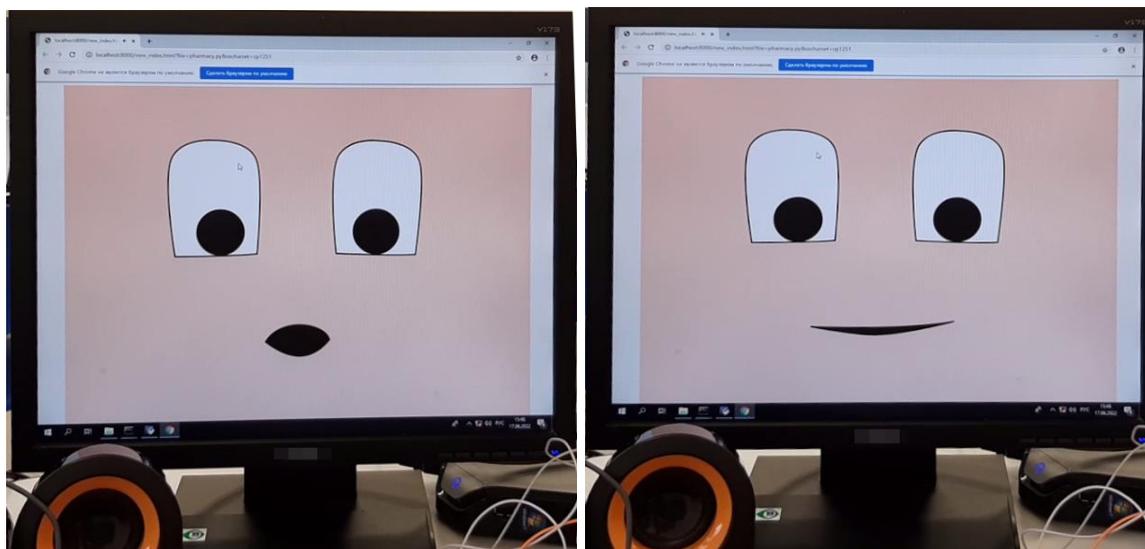


Рис. 6. Артикуляции во время диалога

Библиографический список

1. Beiter, M. An introduction to robotics with NAO / M. Beiter. — Текст : электронный // ALDEBARAN Robotics : [сайт]. — URL: www.aldebaranrobotics.com (дата обращения: 08.12.2012).
2. Передать показания счетчиков воды можно с помощью чат-бота.. — Текст : электронный — URL: <https://www.mos.ru/news/item/95202073> (дата обращения: 27.08.2021).
3. Голосовой помощник 900. — Текст : электронный // Сбербанк : [сайт]. — URL: https://www.sberbank.ru/ru/person/dist_services/golosovoy-pomoshchnik (дата обращения: 08.12.2022).
4. Дашиева, Я. Голосовой бот: примеры автообзвона + как выбрать + 10 программ — Текст : электронный // Агрегатор сервисов in-scale : [сайт]. — URL: <https://in-scale.ru/blog/golosovye-boty/> (дата обращения: 08.12.2022).
5. Лицом к лицу. О непосредственном общении между людьми. — Текст : электронный — URL: <https://dzen.ru/media/id/5eb3e6590ab5b766d085bab2/licom-k-licu-o-neposredstvennom-obsceniimejdu-liudmi-5ebbf9996f5f65f3407196> (дата обращения: 13.05.2020).
6. В Китае ведущим новостей стал робот с искусственным интеллектом — Текст : электронный — URL: <https://dzen.ru/media/kubmedia/v-kitaaveduscim-novostei-stal-robot-s-iskusstvennym-intellektom-video5be583590bb03600a95d400d> (дата обращения: 09.11.2018).

Оглавление

Введение	3
Структура аппаратно-программного комплекса.....	4
Управление репликами робота	5
Управление диалогом	7
Подготовка диалога.....	10
Аппаратная реализация модуля управления диалогом	11
Результаты и выводы	12
Библиографический список.....	14