

# О СОЗДАНИИ СИСТЕМЫ РАСПОЗНАВАНИЯ ГОЛОСОВЫХ КОМАНД НА БАЗЕ МИКРОКОНТРОЛЛЕРА С ИСПОЛЬЗОВАНИЕМ СЕРВИСОВ GOOGLE

Шумель В. В., Рудикова Л. В.

Факультет математики и информатики, кафедра программного обеспечения интеллектуальных и компьютерных систем, Гродненский государственный университет имени Янки Купалы

Гродно, Республика Беларусь

E-mail: shumel.vladimir@gmail.com, rudikowa@gmail.com

*В статье излагаются общие принципы создания системы распознавания и выполнения команд, поданных голосом, на базе микроконтроллера STM32F407 с использованием сервисов Google. Аргументируется актуальность подобного решения, описываются составляющие аппаратной части системы и приводится краткая методология разработки необходимого программного обеспечения. Приводятся выводы о целесообразности предлагаемого решения, его достоинства и недостатки.*

## ВВЕДЕНИЕ

Современные мобильные телефоны имеют различные возможности по распознаванию человеческой речи и выделения из нее конкретных команд для управления устройством, выполнения определенных запросов. Одним из самых известных и популярных решений подобного рода является Siri компании Apple, работающее на смартфонах iPhone и планшетных компьютерах iPad последних версий. С помощью него можно позвонить абоненту, произнеся его имя, сделать какие-либо запросы в сети или поменять настройки телефона. Главными недостатками этой системы является проприетарность, не дающая использовать ее вне устройств Apple и отсутствие поддержки русского языка (по состоянию на момент написания статьи).

Главный конкурент Apple, Google, начиная с версии 4.2 своей операционной системы Android, также предоставляет обладателям мобильных устройств на ее основе возможность распознавания голоса. Но более привлекательным является предоставляемый компанией сервис голосового поиска в сети Интернет, когда проговоренная в микрофон речь транслируется компьютером в текстовый запрос. Данный сервис не является проприетарным, и может быть использован на любом устройстве, подключенном к сети Интернет [1]. На основе этого факта и построена архитектура предлагаемой системы.

## I. АППАРАТНАЯ ЧАСТЬ СИСТЕМЫ

Аппаратная часть предлагаемой системы состоит из микроконтроллера STM32F407, цифрового микрофона, драйвера сети Ethernet и некоторой обвязки (имитация каналов нагрузки, фильтры питания и пр.) [2]. Для универсальности и общности описания предполагается, что драйвер сети Ethernet соединен сетевым кабелем с Интернет-модемом или другим устройством, способным обеспечить описываемой системе доступ к сети Интернет.

Цифровой микрофон используется для преобразования аналогового сигнала, представляющего собой колебания воздуха, создаваемого голосом, в цифровое дискретное представление. Драйвер сети Ethernet обеспечивает микроконтроллеру связь с модемом сети Интернет. Микроконтроллер STM32F407 является основой устройства, обеспечивающего выполнение всех основных и вспомогательных функций.

## II. АРХИТЕКТУРА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ СИСТЕМЫ

Программное обеспечение системы состоит из управляющей программы микроконтроллера и сервисной части Google. Микроконтроллер подготавливает и поправляет на сервер Google аудиофрагменты речи в цифровом виде, и принимает обратно распознанный текст.

Программу микроконтроллера, в свою очередь, можно разделить на две основные части – бизнес-логику и вспомогательные модули. Вспомогательными модулями являются: стек TCP/IP, обеспечивающий связь с сервером Google через драйвер Ethernet, подпрограмма получения данных с цифрового микрофона, подпрограммы управления исполнительными устройствами.

В бизнес-логику входят: система автоматической регулировки усиления цифрового микрофона и фильтрации шумов на его входе, система подготовки файла аудиофрагмента, и интерпретатор полученных с сервера Google ответов на запросы.

## III. ОБЩИЕ ПОДХОДЫ К РАЗРАБОТКЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ МИКРОКОНТРОЛЛЕРА

Микроконтроллер работает по следующему алгоритму. При достижении громкости сигнала на входе определенного уровня микроконтроллер начинает вести запись аудиоданных в буфер, сжимая их с использованием кодека Speex. Если

затем уровень сигнала в течение заданного длительного времени не поднимается до начального уровня, запись прекращается, иначе устанавливается соединение с сервером и ему передается запрос определенного вида с записанными данными. Через какое-то время сервер возвращает распознанную строку в текстовом виде.

Далее строка поступает на интерпретатор. Его задачей является парсинг строки на наличие ключевых слов, задающих определенные команды. Фактически, это самая главная часть программы микроконтроллера, от правильной реализации которой зависит корректная работа всей программы.

Одной из главных проблем, встающих на пути работы интерпретатора – выделение из общего потока человеческой речи только тех команд, которые адресованы именно ему. В идеальных условиях можно обеспечить качественную речь в микрофон без посторонних шумов, но больший интерес представляет использование системы, например, для управления «умным» домом. В этом случае система может ошибочно обработать фразу «Нельзя включать телевизор», адресованную дочери, которой следует выполнить домашнее задание, опознав команду «включить телевизор» и выполнив указанное действие. Чтобы исключить такие ситуации, можно ввести слово-идентификатор, редко используемое в повседневной речи, которое будет служить обращением к системе. Таким словом может быть, например, «Матрица», и, соответственно, команда должна звучать «Матрица, включи телевизор». Распознав ключевое слово в потоке речи, интерпретатор будет уведомлен о том, что следующие за ним команды, действительно, следует выполнять.

Другим методом решения данной проблемы может быть аудиозапрос подтверждения на выполнение заданного действия и ожидание одного из определенных ответов – «да» или «нет». Можно использовать также повтор команды, изложенный в изначальном виде – такие конструкции также редко используются в обычной человеческой речи.

Еще одной проблемой работы интерпретатора является человеческое нежелание уподобляться роботам и действовать по одному сценарию. Так, просьба одного человека другому включить освещение может звучать очень многими способами: «включи свет», «зажги свет», «включи освещение», «зажги освещение», или даже сленговое «дай света». Учет всевозможных комбинаций звучания одной и той же команды может стать довольно сложной, но интересной задачей.

Следует также отметить и необходимость фильтрации музыки или иных шумов от других электронных источников звука. Эту проблему можно решить, если микроконтроллер систе-

мы будет единственным источником звука в помещении. В этом случае аудиофрагмент с микрофона можно обработать до отправки на сервер, вычтя из него сигнал, излучаемый системой, и получив на выходе только посторонний звук (человеческую речь).

#### IV. ДОСТОИНСТВА И НЕДОСТАТКИ СИСТЕМЫ. ОБЛАСТЬ ПРИМЕНЕНИЯ

Главным преимуществом предлагаемой системы является ее относительная дешевизна и компактность, позволяющие встроить ее в состав многих устройств, например, в комплекс управления «умным» домом. Кроме того, система является достаточно гибкой с точки зрения конфигурируемости, позволяя сделать ее настройки под заданные требования.

Большим недостатком системы является зависимость от сервисов Google. Условия предоставления услуги не гарантированы и могут быть изменены компанией в любой момент или даже заблокированы. К недостаткам также можно отнести относительную сложность системы, для построения которой необходимо обладать навыками в разных областях программирования.

Перечисленные достоинства и недостатки определяют область применения описываемой системы – приложения, отказ от которых не повлияет на жизнедеятельность человека или предприятия. Это может быть, например, часть системы по управлению «умным» домом, «умными» автомобилями, где голосовыми командами можно управлять включением и работой музыкальных приборов, вентиляции, регулировки температуры и прочего. Общая система управления обязательно должна иметь возможность альтернативного контроля на случай отказа описываемой системы.

#### ЗАКЛЮЧЕНИЕ

В статье было рассмотрены общие подходы к построению системы голосового управления на базе микроконтроллера с использованием сервисов Google. Также описаны аппаратная часть системы, подходы к реализации архитектуры, общие подходы к разработке программного обеспечения. В завершении были даны примеры использования систем подобного рода.

1. Пульц, М. Доступ к функциям Google по распознаванию голоса из браузера Chrome 11 // Персональный блог Майка Пульца [Электронный ресурс]. – Режим доступа : <http://mikepultz.com/2011/03/accessing-google-speech-api-chrome-11/> – Дата доступа : 19.09.2013.
2. Описание отладочной платы STM32F4DISCOVERY // Официальный сайт компании ST Microelectronix [Электронный ресурс]. – Режим доступа : <http://www.st.com/web/catalog/tools/FM116/SC959/SS1532/PF252419> – Дата доступа : 17.09.2013.