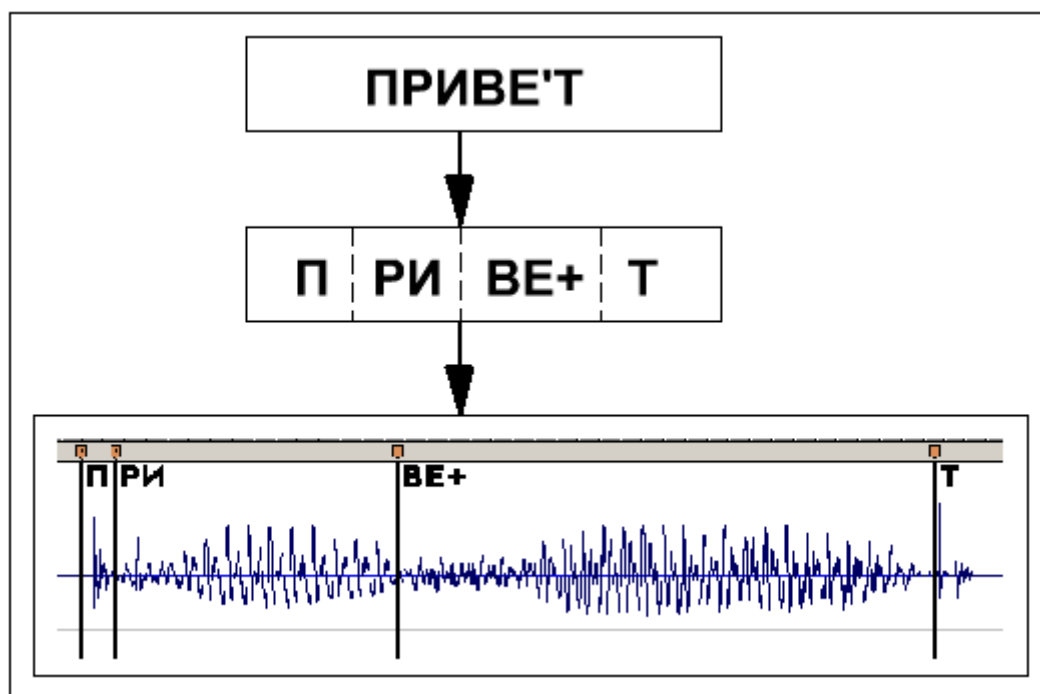


Модуль синтеза речи

Иногда требуется научить устройство говорить. Зачастую этот вопрос решается обычной записью в память нескольких фраз с целью дальнейшего их воспроизведения по каким-то событиям. Но более универсальным решением является использование синтеза речи. Здесь будет рассмотрена попытка создания простого встраиваемого синтезатора речи. Он разговаривает на русском языке и может быть интегрирован в устройства, имеющие свой микроконтроллер со свободным каналом UART. Возможные области применения – от игрушек до систем умного дома. Не стоит также забывать и об устройствах для людей с ограниченными возможностями.

Модуль синтеза является преобразователем произвольного текста в речь (TTS) и работает по принципу компилятивного фонемного синтезатора. Многим людям эти слова мало что скажут, поэтому рассмотрим его вкратце. Записывается речь диктора (слова, предложения). Затем в редакторе речь разделяется на элементы – фонемы (звуки «а», «бэ», «мэ» и т.д.). Из наиболее удачных фонем компонуется фонемная база. Теперь, чтобы синтезировать какое-то слово, нужно сопоставить его написанию последовательность фонем и проиграть их.



В данной конструкции используется упрощенная база от бесплатного компьютерного синтезатора Голос [1] на 685 фонемах. Чтобы база уместилась в чип памяти, я применил сжатие по алгоритму IMA ADPCM 4bit.

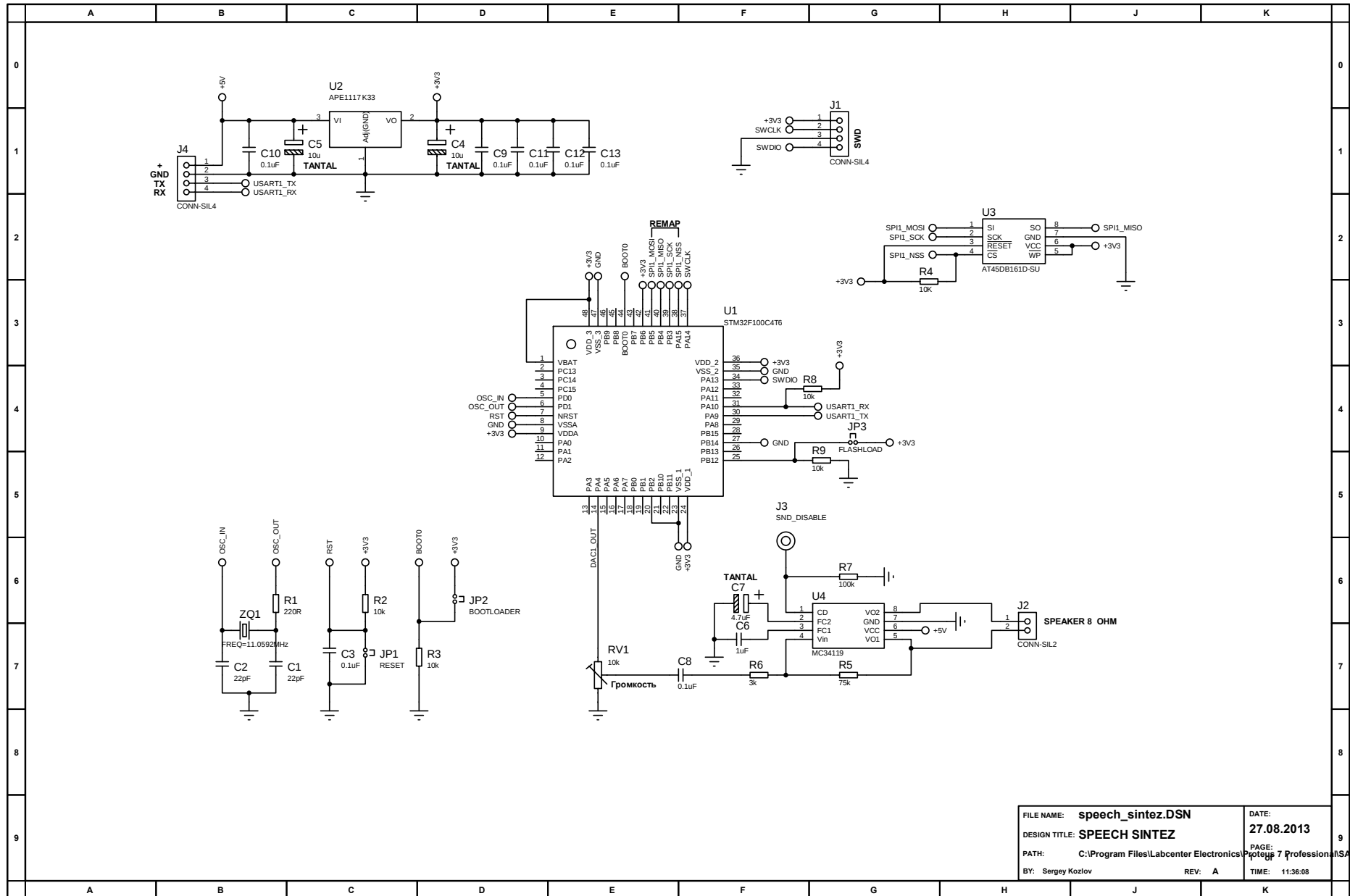
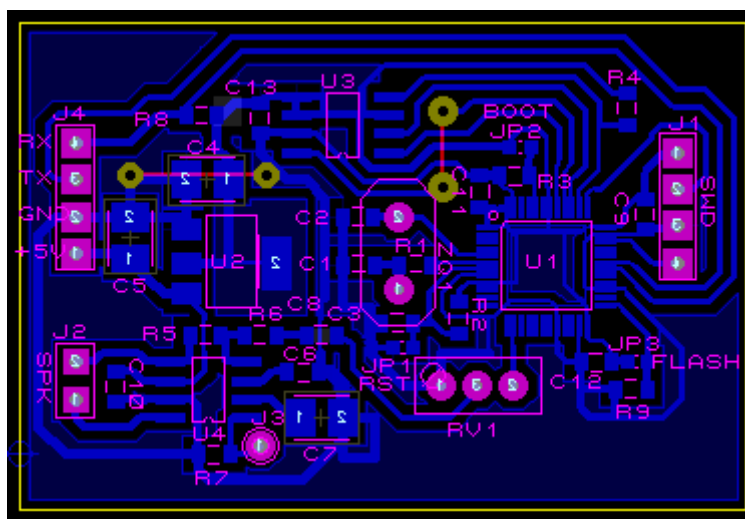


Схема.

Принципиальная схема довольно проста. Всем делом заправляет микроконтроллер U1 STM32F100C4T6. Он работает на тактовой частоте 22,1МГц(11,059*2). Выбор в пользу STM32 был сделан из-за наличия у него встроенного ЦАП. Записи фонем хранятся в микросхеме памяти U3 AT45DB161 ёмкостью 2 мегабайта. На микросхеме U4 MC34119 реализован мостовой усилитель с выходной мощностью 0,25 Ватт. К его выходу(разъем J2) подключается динамик или наушники. Подстроечным резистором RV1 регулируется громкость звука. Вход J3 может использоваться для отключения усилителя мощности с целью уменьшения потребления схемы. Для этого туда надо подать напряжение >2В. Микроконтроллер и память питаются напряжением 3,3В. Оно поступает с выхода стабилизатора U2 APE1117-K33. На разъем J4 подаётся напряжение 4-7В от блока питания (контакты 1,2) , туда же выведены линии UART Tx и Rx(контакты 3,4), через которые речевой модуль связывается с внешним управляющим контроллером. Tx и Rx толерантны к напряжению 5В и без проблем согласуются с ТТЛ логикой. J1(SWD) служит для прошивки микроконтроллера. Перемычки JP2(BOOTLOADER), JP3(FLASHLOAD) используются в режиме прошивки микроконтроллера и загрузки дампа в память U3.

Плата и детали.

Печатная плата односторонняя, с двумя проволочными перемычками. Размеры 50*33мм.



Почти все детали применены в корпусах для поверхностного монтажа(SMD) за исключением двух компонентов: кварцевый резонатор ZQ1 и потенциометр RV1. Эти компоненты монтируются с противоположной стороны платы.

Кварц на 11,0592МГц в низкопрофильном корпусе HC-49/S(лодочка).

Потенциометр RV1 - многооборотный типа 3296W с номиналом от 5к.

Стабилизатор на 3,3В APE1117 можно заменить на AMS1117. У них много аналогов. Также подойдет 78L33 или KP1158EH3,3В с соответствующей доработкой печатной платы.

Микросхема памяти AT45DB161D-SU в 8-выводном корпусе SOIC, можно использовать AT45DB161D-TU(28-выводной TSOP). Здесь используется память с размером страницы в 528 байт(как идет с завода). Её можно переконфигурировать на 512 байт, но только один раз. И такая работать естественно не будет. Поэтому я рекомендую ставить новую микросхему.

Динамик должен быть сравнительно высокоомным($\geq 8 \text{ Ом}$). Желательно использовать головки от 0,5-1Вт т.к. слишком маленькие динамики плохо передают басы и страдает разборчивость речи.

Радиолюбители, ранее не работавшие с многоногими микроконтроллерами, могут испытать трудности с их пайкой. Для пайки таких микросхем я использую паяльник с жалом «микроволна»[2]. Главное – это перед пайкой обильно полить флюсом ноги микросхемы. После того как плата будет запаяна, её следует отмыть от остатков флюса. Иначе кварцевый генератор может работать со сбоями.

Прошивки.

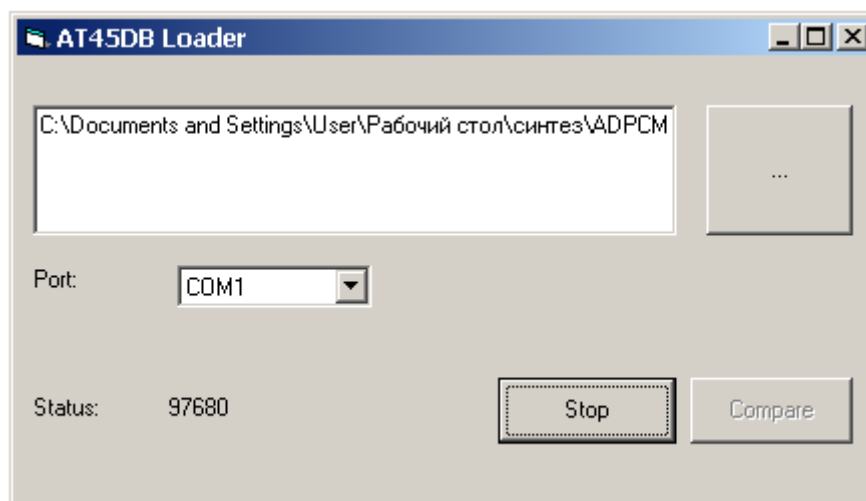
Микроконтроллер STM32 прошивается через интерфейс SWD(выведен на J1) или UART(Tx и Rx на J4).

Назначение контактов на разъёме SWD соответствует контактам разъёма отладчика ST-LINK на оценочной плате STM32VLDISCOVERY. Обладатели ST-LINK(и иных LINK-ов) наверняка прекрасно знают как им пользоваться, поэтому рассматривать прошивку через SWD не будем.

Когда отладчика под рукой нет, можно запрограммировать чистый чип через UART используя встроенный загрузчик(bootloader). Для его активации запаивается перемычка JP2(BOOTLOADER), к линиям Rx/Tx подключается обычный преобразователь уровней COM-порта в TTL(например на MAX232) или USB-TTL(например FT232). Далее подается питание и запускается программа Flash Loader Demonstrator. Файл прошивки – **sintez.hex**. Весьма подробно и с картинками этот процесс расписан в статье про загрузчик STM32[3]. Также дополнительную информацию по этому методу вы найдете в архиве с прошивками. После успешного прожига программы в контроллер перемычку JP2 следует убрать.

Надо отметить, что в отличие от контроллеров AVR, на этапе программирования STM32 нельзя получить мёртвый кирпич. В любом случае он будет виден как отладчиком, так и загрузчиком.

Теперь надо залить фоновые данные в чип памяти AT45DB161. Можно это сделать специальным программатором, который опять есть не у всех, поэтому я написал свою утилиту загрузки через COM-порт. Запаиваем перемычку JP3(FLASHLOAD), даём питание, запускаем программу AT45DB Loader, затем указываем файл **adpcm.bin**, выбираем номер порта и жмём кнопку Load.

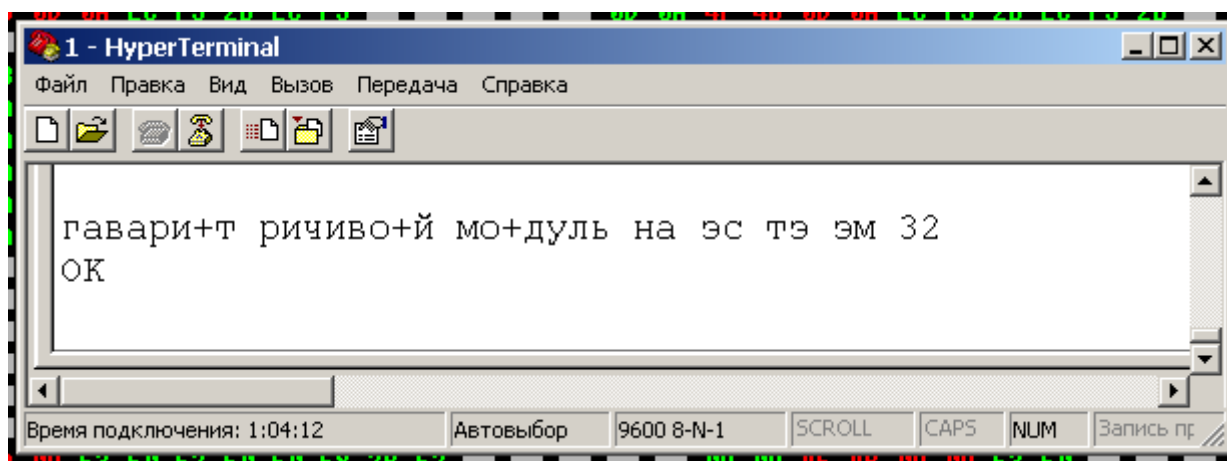


В поле Status появится надпись Chip Erase, говорящая о том, что идет процесс стирания Flash памяти. Секунд через 15 начнётся запись и побегут цифры.

После прошивки памяти надо убрать перемычку JP3. Теперь синтезатор наконец-то готов к работе.

Использование.

Как было ранее сказано, модуль синтеза не является полностью автономным устройством. Он получает тексты и команды от другого управляющего контроллера или компьютера. Проверить его работоспособность и изучить команды можно с помощью терминальной программы (например, HyperTerminal из ОС Windows). Параметры COM порта: 9600, 8N1. В свойствах установить передачу последовательности CR+LF в конце строки и отображение введённых символов на экране (режим эхо).



Текст желательно вводить так как мы его слышим (или как пишет малограмотный человек:)). Знаком плюс(+) помечаются ударные гласные. В простых словах с одной гласной буквой ударения расставляются автоматически. По нажатию на клавишу Enter запускается воспроизведение. Размер вводимой строки – до 250 символов. Если в строке больше чем 250 символов, то воспроизведение начинается автоматически по заполнению буфера без нажатия на Enter.

По окончании работы синтезатор возвращает **ОК** и готов к приёму нового текста.

Числа корректно проговариваются в диапазоне 0...9999. Большие числа разбиваются на части. Например, число 123456 будет произнесено как 1234 – 56.

Знаки точка(.) и запятая(,) добавляют небольшую паузу между словами.

Список поддерживаемых команд.

Команда	Описание	Ответ
STATUS	Запрос статуса синтезатора	READY – готов к приёму текста PLAY – идет синтез речи
STOP	Остановить синтез	OK – синтез остановлен и синтезатор готов к приёму текста.
SPEEDx x=1...5	Настроить скорость синтеза. Выше номер – больше скорость и выше тон голоса. По умолчанию используется SPEED2.	OK – новая скорость установлена
BEEP	Издаёт короткий звук «бип».	OK – воспроизведение звука завершено.

Команды должны быть записаны в верхнем регистре (т.е. большими буквами). Формат у ответа такой же как у команды, например: OK<CR><LF>. Где символ <CR> = 0x0D, <LF> = 0x0A.

Недостатки и возможности для улучшения.

Первый и самый главный недостаток – несовершенство речевого движка. Синтезируемый голос монотонный и очень похож на речь роботов из фантастических фильмов. Есть много «спотыкачек» и некоторые слова на слух воспринимаются не сразу. Улучшить качество можно за счет увеличения размеров фонемной базы и её более точной нарезки, что также влечёт за собой и расширение набора правил синтеза. Второй недостаток – это необходимость вводить слова с ударением. Добавление словаря на 100-200 тысяч слов сможет покрыть большую часть тем в русском языке и решить данную проблему. Но всё это потребует значительного увеличения объёма Flash памяти.

Хочу выразить благодарность автору синтезатора Голос за помощь по фонемной базе.

Ссылки.

1. Домашняя страница синтезатора Голос

<http://beriinfo.narod.ru>

2. Пайка планарных микросхем самодельной микроволной

<http://easyelectronics.ru/pajka-planarnyx-mikrosxem-samodelnoj-mikrovolnoj.html>

3. Загрузчик STM32

<http://easystm32.ru/useful-things/42-bootloader-stm32>

Файлы.

Схемы, платы, прошивки [firmware.zip]

Исходники фонемной базы [phonem_base.zip]

Образец записи синтезированного голоса [speech_test.mp3]