International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 6th Apr 2018. Link :http://www.ijiemr.org/downloads.php?vol=Volume-7&issue=ISSUE-04

Title: **IOT BASED VOICE CONTROLLED HOME AUTOMATION USING NODE MCU**

Volume 07, Issue 04, Pages: 23–29.

Paper Authors

DR.K.B.S.D.SARMA, PROF. PRAVIN AKULA, PAVANI SUMA .Y, MANOJ RAVURI, JAYA SREE S

International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IOT BASED VOICE CONTROLLED HOME AUTOMATION USING NODE MCU

[1]DR. K.B.S.D.SARMA, [2]PROF. PRAVIN AKULA, [3]PAVANI SUMA .Y, [4]MANOJ RAVURI, [5]JAYA SREE S

Abstract— In this paper we give a concept of a device that provide voice interface (VI) for issuing commands and receiving messages. Proposed device is designed having Internet of Things (IoT) paradigm in mind and connected to the internet via IEEE 802.11. It uses Google speech to text (STT) and text to speech (TTS) web services for voice analysis and synthesis. Hardware prototype description is given in detail. Key aspects like price, performance, memory footprint and energy consumption are analyzed on a case study. We discuss potential applications when such a device is deployed as an interface of home automation system.

Keywords— Human voice, Speech processing, Internet of Things, Speech to Text, Text to Speech, embedded device, case study

I.INTRODUCTION

Embedded computers are indeed the foundation of modern technology. Every modern electrical appliance has some processing unit hence it is a computer based system. Computer technology is moving from embedded devices to pervasive or ubiquitous computing which means that we have more and more computers that are embedded everywhere and in everything. With the advancement of Internet in both coverage and speed such computers are often having connecting capabilities forming what is now popularly called Internet of Things (IoT).IoT paradigm assumes many devices connected over conventional internet network. These devices usually have restricted resources so moving part of the service implementation to a cloud infrastructure is a prominent solution. On the other hand, having to interface with many devices could be very cumbersome. For IoT deployment in everyday life, devices need to be designed ergonomically.

Authors of this paper see human voice as a potential interface for one or more devices in IoT ecosystem enabling issuing commands and receiving information via voice messages.In our previous work [1] we have demonstrated good potential of the Google STT service [2] as automatic speech recognition (ASR) for issuing voice commands. Two Android applications were developed and recognition performance is evaluated on a set of commands. Results are further improved by matching recognized words and phrases with possible commands by algorithms based on Longest Common Subsequence (LSC) and Levenshtein distance.Many authors use Android devices as an IoT device. Android has an important role for emerging IoT since it offers stable development platform. Usually it runs on a powerful hardware that is not affordable for all IoT applications. In this paper we are demonstrating how to provide a rich VI using a device with much smaller memory footprint and a fraction of processing power.

II.DEVICE PROTOTPE

In this paper we describe prototype of a VI device for IoT (VI4IoT). As a main processing chip VI4IoT uses 500MIPS xCORE multi-core microcontroller that has eight 32 bit logical cores [3]. It is connected to the Internet using TiWi-SL self-contained IEEE 802.11b/g Wi-Fi module based on Texas Instruments CC3000 chip. For recording and reproducing audio MikroElektronika SmartMP3 board with VS1053 CODEC chip is used. Among other formats it is capable of encoding ADPCM audio and decoding MP3 audio.Concept of a proposed device is shown in Fig 1. User issues commands using voice. Voice is recorded and immediately sent to STT cloud service using HTTP POST method. As a response it receives JSON object with multiple alternative transcript of analyzed voice sequence. Transcript alternatives are then matched against set of supported commands using previously developed algorithm [1].

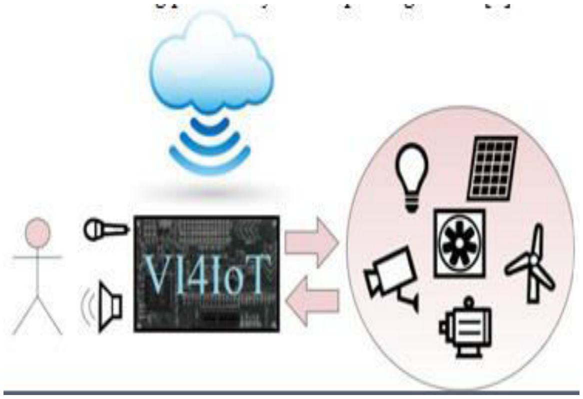
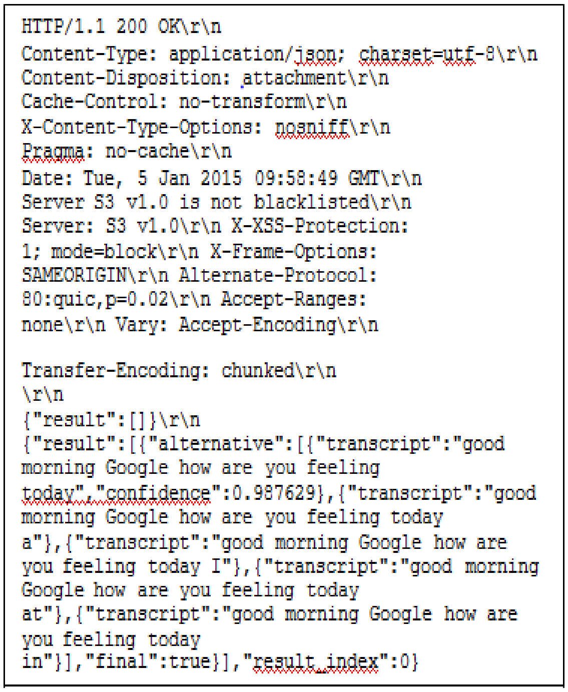


Fig. 1. Concept of a device that provides VI for issuing commands and receiving and reproducing voice messages.

When a command is recognized particular command handler routine is executed. We have implemented some representative commands having a home automation, information center or entertainment system in mind. Actuations of a command can results with two possible events. If actuator is directly connected to a VI device it can interface with actuator using binary state of general purpose input output pin (GPIO) or some intensity using pulse-width modulation (PWM) as we demonstrated in [4] or a SPI and I2C bus. In case controlled IoT device not connected to the VI4IoT device than the VI4IoT device sends a UDP datagram to a controlled device. As a third option, the VI4IoT device can communicate with web service like weather portal to get information about the weather forecast or similar. When information is received, we are forming sentences and use them for voice synthesis via Google TTS service. Synthesized voice is downloaded using HTTP GET request. Voice is coming in MP3 format and VI device decode it and reproduce with VS1053 CODEC chip.The microprocessor used in device prototype has a powerful architecture for concurrent task execution. Each logical core has a guaranteed processing time slice of up to 125 MIPS on a 500MHz using XMOS synchronization technology so it has a deterministic behavior. This is potentially good for implementation of various network protocols. But this device has only 64 Kbytes of memory on the chip. This is not enough for implementing voice analysis or voice synthesis on the device. In principal it is possible to use external memory to overcome this shortcoming. Since we are trying to make VI device low cost we have not used any external memory to explore what can be done in constrained resource target platform.

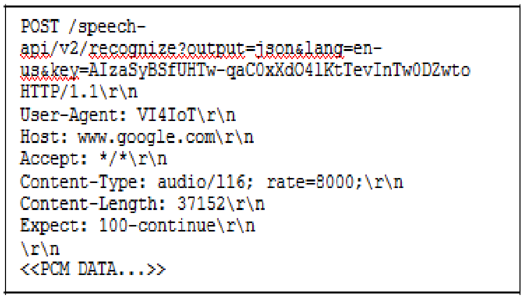
III.GOOGLE API

At fist Google released web service [2] that will help people search web and call businesses. Their goal was to make voice recognition and ubiquitously available around the globe. To achieve universal availability HTTP protocol is used. Search by voice service is integrated in Chrome web browser. Also HTML5 libraries are available to use both STT and TTS in any web site. Even though service is available on the internet, documentation is not. To port service on to VI4IoT device we have reverse engineered protocol using Wireshark software.STT service is available via a HTTP post method. In Fig. 2 we show example for sending raw 16 bit PCM data. HTTP post method header is preceding PCM data. Service response is in JSOM format like specified in post method as shown on Fig. 3



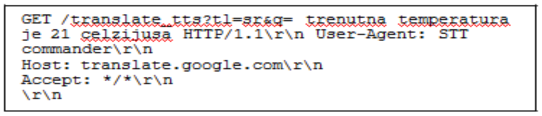
```
HTTP/1.1 200 OK\r\n
Content-Type: application/json; charset=utf-8\r\n
Content-Disposition: attachment\r\n
Cache-Control: no-transform\r\n
X-Content-Type-Options: nosniff\r\n
Pragma: no-cache\r\n
Date: Tue, 5 Jan 2015 09:58:49 GMT\r\n
Server S3 v1.0 is not blacklisted\r\n
Server: S3 v1.0\r\n X-XSS-Protection:
1; mode=block\r\n X-Frame-Options:
SAMEORIGIN\r\n Alternate-Protocol:
80:quic,p=0.02\r\n Accept-Ranges:
none\r\n Vary: Accept-Encoding\r\n

Transfer-Encoding: chunked\r\n
\r\n
{"result":[]}\r\n
{"result":[{"alternative":[{"transcript":"good
morning Google how are you feeling
today","confidence":0.987629},{"transcript":"good
morning Google how are you feeling today
a"},{"transcript":"good morning Google how are
you feeling today I"},{"transcript":"good morning
Google how are you feeling today
at"},{"transcript":"good morning Google how are
you feeling today
in"}],"final":true}],"result_index":0}
```

Fig. 3. Example of a received HTTP 200OK message that contains voice transcript in JSON format.



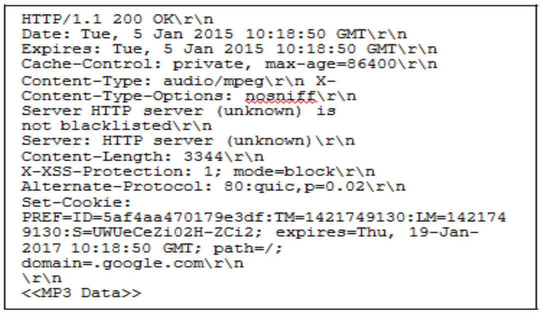
```
POST /speech-
api/v2/recognize?output=json&lang=en-
us&key=AIzaSyBSfUHTw-qaC0xXd041KtTevInTw0D2wto
HTTP/1.1\r\n
User-Agent: VI4IoT\r\n
Host: www.google.com\r\n
Accept: */*\r\n
Content-Type: audio/l16; rate=8000;\r\n
Content-Length: 37152\r\n
Expect: 100-continue\r\n
\r\n
<<PCM DATA...>>
```

Fig. 2. Example of a HTTP post method used to access Google STT service



```
GET /translate_tts?tl=sr&q= trenutna temperatura
je 21 celzijusa HTTP/1.1\r\n User-Agent: STT
commander\r\n
Host: translate.google.com\r\n
Accept: */*\r\n
\r\n
```

Fig. 4. Example of a HTTP get method used to access Google TTS service Firmware is divided in to several modules.



```
HTTP/1.1 200 OK\r\n
Date: Tue, 5 Jan 2015 10:18:50 GMT\r\n
Expires: Tue, 5 Jan 2015 10:18:50 GMT\r\n
Cache-Control: private, max-age=86400\r\n
Content-Type: audio/mpeg\r\n X-
Content-Type-Options: nosniff\r\n
Server HTTP server (unknown) is
not blacklisted\r\n
Server: HTTP server (unknown)\r\n
Content-Length: 3344\r\n
X-XSS-Protection: 1; mode=block\r\n
Alternate-Protocol: 80:quic,p=0.02\r\n
Set-Cookie:
PREF=ID=5af4aa470179e3df:TM=1421749130:LM=142174
9130:S=UWUeCeZi02H-ZCi2; expires=Thu, 19-Jan-
2017 10:18:50 GMT; path=/;
domain=.google.com\r\n
\r\n
<<MP3 Data>>
```

Fig. 5. Example of a HTTP get method used to access Google TTS service.

TTS service is available with HTTP get method. Desired language and phrase that is synthesized is passed as a query string with two name and value pairs. Example is shown on Fig 4. Figure 5. shows STT response as HTTP 200 OK message.

Actual synthesized voice is retrieved as a body of the message.

IV. FIRMWARE DESIGN

In a contrast with our previous work [1] where Android application is developed using Android API, implementation of firmware for VI4IoT is done in C language with some elements of the XMOS C language extension called XC [3]. XC is a C language extension that exposes XMOS architecture supported concurrency. It provides compile time race condition check and gives control of communication, timing and I/O port access. XC compiler is based on low-level virtual machine (LLVM) compiler so it can be used with C and C++ language.There is no operating system (OS) on xCore devices. Firmware developed as a monolithic program that executes in multiple threads. Each thread is mapped on a logical core. It is possible to have multiple threads executing on a single core. If executed on a same core, instructions from multiple threads are scheduled with a simple round-robin scheme. Memory access synchronization and thread suspension or pausing is supported in hardware. This provides very comfortable development environment for firmware development. Main interfacing between thread is via bidirectional links.

VI4IoT FSM loop	Command manager	
VAD	STT	TTS
SmartMP3 driver	SPI	TiWi-SL driver

Fig. 6. Firmware is divided in to several modules.

Firmware is divided in to several modules as shown on Fig6.Some of them are logical unit containing set of library functions and some of them are execution threads. Main thread is implemented as a finite state machine (FSM). It is executed in endless loop from when device is powered up. This practically means that VI4IoT is continuously listening for voice commands. FSM design is given on fig 7.

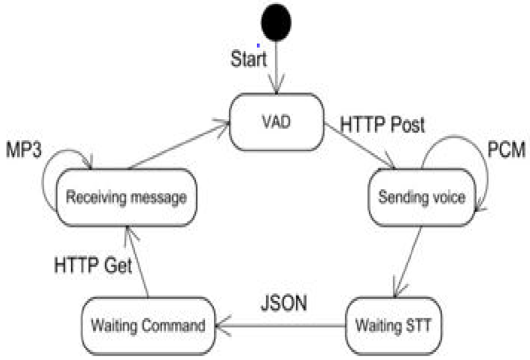


Fig. 7. Main thread is realized as FSM.

VI4IoD device starts in VAD state executing voice activity detection. We use algorithm for energy based foreground and background signal separation similar to [5] but less accurate hence less compute intensive. When voice activity is detected VI4IoT starts HTTP post method to Google STT web service. While voice activity is present device sends PCM data to service. When PCM sending is done device waits for a result of a voice transcript. It is received as a JSON object like shown on Fig. 3.Transcript is analyzed with already mentioned algorithm and command is executed. FSM waits for a command to be done and then generated HTTP get request to acquire voice message that will be reproduces to the user. After message playback is finished device is back in VAD state.

V.RELATED WORK

Many researchers are making contribution to this field finding new application for voice controls. In recent work [6] development of a wheelchair platform is presented. The WebKit open-source project is used as a cloud based voice recognition service. It would be interesting to see performance comparison with Google STT service. In this work only speech analysis is used. Authors in [7] use the same approach as ours using cloud infrastructure to enable speech recognition for portable maintenance aids with limited resources. In both papers speech synthesis is not mentioned. Soda at al. in his work [8] uses microphone array network to build a voice interface for smart home system. Also they explore usage of virtual agents [9] as web service to enrich user experience with the smart home system. His work is concentrated on Japanese language. Authors of [10] gave a description of a architecture to implement Estonian speech recognition for Android devices. Key advantage for use was that Google services offered support for Serbian language.Also, some commercial solutions are available today. Company HouseLogix offers two products that can be used as a VI for home control system in a smart house [11]. First one is a VoicePod Mobile application for smart-phones that is a natural language voice assistant. It offers flexible command phrasing and silent operation. Other solution is a proprietary VoicePod Tabletop Device that is designed to be used in quiet rooms. This solution is coupled with Control4 home automation control system [12]. Even more recently company Amazon introduced product called Echo [13]. It serves as a information service that can tell weather and news and also reproduce music content as a multimedia player.

VI. EVALUATION

Evaluation of a VI device is done simulating household environment with some home appliances that need to be turned off or turned on. Another use case was reading temperature from simulated thermometer and forming a message that contains a number of degrees in Celsius. Next use case was issuing a command with a query about current outside conditions and weather forecast.To analyze memory footprint and energy consumption we are looking in to amount of data that needs to be stored and processed on a device. Also same amount of data has to be transmitted and send to a local IEEE 802.11 access point. We are not compressing recorded voice since Google STT service only supports FLAC format for compressed audio. FLAC encoder is algorithmically complex for implementation on device that we propose. For this reason we are sending mono PCM data. As indicated in [14] this could save energy since processing dissipation can be larger than energy needed for transmission of uncompressed data. We have chosen 16000Hzsample rate and use 16 bit precision samples. For these parameters we got good results in previous work [1]. Having a voice command recognition implemented is also not feasible on proposed device due to memory constraints. Speech Synthesis on a memory constrained device is possible with some advanced techniques and some quality compromises. Still as indicated in [15] size of such implementation is around 100 Kbytes and that is over our budget. Not to mention that we would have space only for VI implementation on the device and no possibility to have any other feature.

TABLE I. SUPPORTED VOICE COMMANDS SIZES IN BYTES

Command	Size (bytes)
Turn the light on	50546
Turn the light off	60342
Tell me current temperature	143544
Tell me current pressure	143978
Tell me current humidity	144924

To get a general picture about memory profile sizes of some supported commands are given in Table 1.

TABLE II. SIZES OF SYNTHESIZED MESSAGES IN MP3 FORMAT

Message	Average size (bytes)
Status message	3030
Current humidity	3448
Current pressure	3239
Current temperature	4702

Examples of synthesized messages are given in Table 2. All commands and messages are measured on sequences in Serbian language. Single audio channel MP3 data is encoded by TTS service using 22050Hz sample rate with 32kbps average date rate. Sizes are shown in Table 2. for a single message. When we consider that each message contains a different numeric value, storing all possible messages on the device will be hard or impossible. Our solution is flexible in such a way that only additional strings and string operation are needed to support a new feature. Still, good MIPS estimation for new features is not possible without knowledge of what they are. Main idea of our solution is that MIPS intensive voice recognition and synthesis is done in cloud.

VII. CONCLUSION

We have successfully demonstrated a proof of concept of a VI device for IoT. The proposed scenario shows a good potential to deploy such device to a household. It can serve as a central control unit or information center for home automation system. Since we use a powerful cloud services limitation to a number of controls that could be recognized or phrases and messages that VI device can reproduce is amount of memory for implementing command handles and small amount of memory to add new string constants for comparison. Localization or multilingual support of the cloud services also comes handy allowing people speak and listen in their own native language.

REFERENCES

[1] M. Stefanovic, N. Cetic, M. Kovacevic, J. Kovacevic and M. Jankovic, "Voice control system with advanced recognition," in Telecommunications Forum (TELFOR), 2012 20th, Belgrade, 2012.

[2] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar and B. Strope, "“Your Word is my Command”:Google Search by Voice: A Case Study," in Advances in Speech Recognition, Springer, 2010, pp. 61-90.

[3] D. May, "The XMOS Architecture and XS1 Chips," Micro, IEEE, vol. 32, no. 6, pp. 28-37, 2012.

[4] N. Vrga, N. Cetic, J. Kovacevic and D. Bardek, "HTTP server for control of home appliances based on XMOS platform," in 21st Telecommunications forum TELFOR, Belgrade, 2013.

[5] S. Goetze, J. Schroder, S. Gerlach, D. Hollosi and J.-E. Appell, "Acoustic Monitoring and Localization for Social Care," Journal of Computing Science and Engineering, vol. 6, no. 1, pp. 40-50, 2012.

[6] S. Andrej, K. Andrej, K. Davorin and S. Radovan, "Prototype of Speech Controlled Cloud Based Wheelchair Platform for Disabled Persons," in 3rd Mediterranean Conference on Embedded Computin, Budva, Montenegro, 2014.

[7] H. Wei and S. Xincun, "VUI system of the portable maintenance aids based on cloud computing," in Computational Problem-Solving (ICCP), 2012 International Conference on, Leshan, China, 2012.

[8] S. Soda, M. Nakamura, S. Matsumoto, S. Izumi, H. Kawaguchi and M. Yoshimoto, "Handsfree voice interface for home network service using a microphone array network," in Third International Conference on Networking and Computing, 2012.

[9] S. Soda, M. Nakamura, S. Matsumoto, S. Izumi, H. Kawaguchi and M. Yoshimoto, "Implementing virtual agent as an interface for smart home voice control," in 19th Asia-Pacific Software Engineering Conference, 2012.

[10] A. Tanel and K. Kaljurand, "Open and extendable speech recognition application architecture for mobile environments," in SLTU, 2012.