EMERGING TECH CONFERENCE – Edge Intelligence

Volume 01, 2022, Page 267 – 270

Proceedings of Emerging Tech Conference:
Edge Intelligence 2022

End-point MCU implementation of Voice User Interface

Georgios Flamis

Senior Staff System Design Engineer, Renesas Europe, Alkminis str, Patras 26500, Greece

georgios.flamis.zc@renesas.com

Abstract

Rapid consumer adoption of connected smart speakers and voice assistants suggests that many people prefer to control products with speech rather than counter-intuitive button pushes. Consumers also seem to appreciate devices and home appliances that are smart without being connected to the cloud via internet. The Renesas Voice User Interface (VUI) solution, a third-generation reference platform built on the scalable Arm-based RA2, RA4, and RA6 series of MCUs, is especially targeted to applications in home automation, white goods, and small appliances. The solution gives access to a variety of developer tools and training resources that will take product designs to the next level, including design and training methods of the integrated machine learning model and full evaluation with on-line and off-line methods.

1. Introduction

Voice User Interface (VUI) is speech recognition technology that enables users to interact with a computer, smartphone or other daily life used device with voice commands as shown in Fig.1. The unique feature of VUI is the use of voice as primary mode of interaction, in contrast with traditional keyboard, mouse, display or touch screen.

Figure 1: Daily life used devices with VUI

The new Renesas easy-to-use hardware platform for VUI solutions is based on the high performance Renesas Advanced (RA) family of 32-bit microcontroller units (MCUs).

The RA Family delivers key advantages compared to competitive Arm Cortex-M MCUs by providing stronger embedded security, superior CoreMark® performance, and ultra-low power operation.

PSA Certified provides customers the confidence and assurance to quickly deploy secure IoT endpoint and edge devices, and smart factory equipment for Industry 4.0.

It is lined up into four product series. Each of these series has a unique feature set, making it ideal for various applications and market needs. The RA8 Series is the top-end product series, aiming for the highest integration and the highest performance. The RA8 Series supports operation at CPU speeds over 240 MHz with single or dual core, with the largest Flash and RAM integration to suit applications where performance really matters most. The RA6 Series offers the widest integration of communication interfaces, with integrated Ethernet and TFT display drivers. Memory densities range from 256KB Flash to 2MB Flash. The RA6 Series offers up to 240MHz performance running on the Cortex-M4 or Cortex-M33 core with TrustZone. The RA6 Series supports full security integration, making these devices widely desired for security applications. The RA4 Series balances the requirements for low power with the demand for connectivity. It offers up to 1MB Flash and a wide range of communication interfaces. The utilized core is the Cortex-M4 or Cortex-M33 with TrustZone and additional security IP integration. Memory densities range from 256KB Flash up to 1MB Flash. These devices provide a CPU frequency of up to 100MHz. On the lower end is the RA2 Series, where the low power requirements of an application matter most for these device definitions. To achieve the best performance, special power-down modes are provided, making these devices well suited for battery-powered applications. The RA2 Series provides memory densities of up to 256KB embedded Flash and a wide single voltage supply range of 1.6 to 5.5V. These devices use the Cortex-M23 core at up to 48MHz.

The Renesas Flexible Software Package (FSP) is an enhanced software package designed to provide easy-to-use, scalable, high-quality software for embedded system designs using Renesas RA Family Microcontrollers. It uses an open software ecosystem and provides flexibility in using bare-metal programming, included Azure RTOS or FreeRTOS, other preferred RTOS, legacy code, and third-party ecosystem solutions. The combination of the flexible open architecture of the FSP plus the wide choice of 3rd party solutions as part of the Arm ecosystem increases the range of choice for application

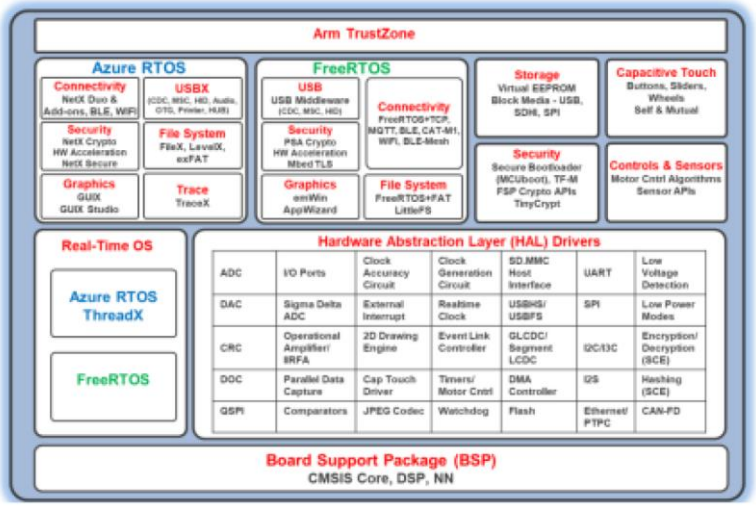

Figure 2: Flexible Software Package

development. This means that developers can choose the software model that best suits their needs while utilizing Renesas's excellent Arm-based silicon solutions as well as speed up the implementation time of complex areas like connectivity and security.

2. Voice recognition engine

Based on Renesas ecosystem, Cyberon DSpotter is a local voice trigger and command recognition solution with robust noise reduction that consumes very low resources and provides high-accuracy performance. It supports multiple languages as well as many connectivity function and security capable depending on the selected MCU. The major features are listed below:

- Local voice recognition algorithm, no network connectivity needed
- Phoneme-based modeling
- Quick command customization: don't have to collect speech data in advance
- Optimization by model adaptation just with small amount speech data
- Global language support: 44+ languages worldwide
- Small footprint and cost effective (single DMIC +RA6E1 or RA4E1)
- DSMT Tool: wake-word and commands customization, performance tuning, testing, no-prior neural network knowledge needed
- Separation of recognition engine and command model, switching commands dynamically
- Low power and high efficiency RA MCU with strong security function

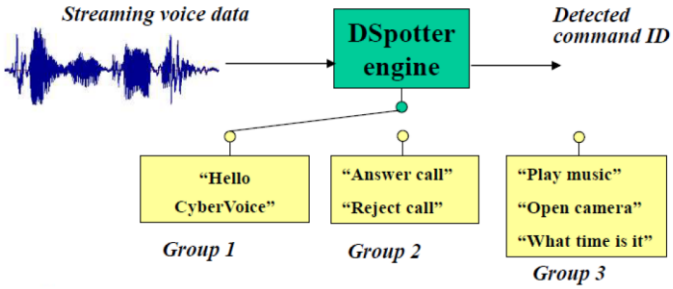

Figure 3: Flexible Software Package

3. System Architecture

Hit rate has been captured with voice commands mixed with different types of noise in levels suitable to create distinguished signal to noise ratios. The test bench of Fig. 4 is utilized and the results are summarized in Table 1.

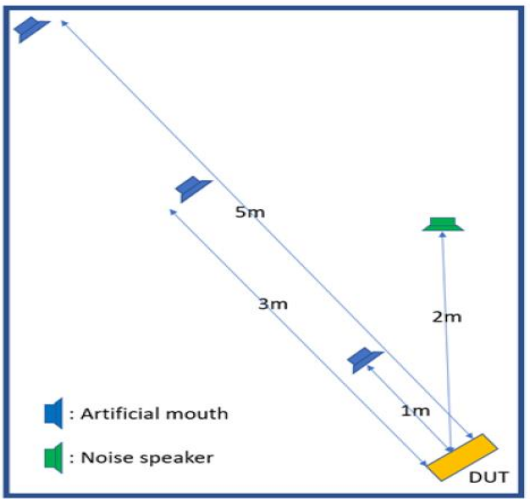

Figure 4: Evaluation test bench

SNR	Background noise	Distance	Hit-Rate	Alexa Requirements
(Clean)	None	1.5m	100.00%	90%
(Clean)	None	3m	100.00%	90%
10dB	Babble	1.5m	98.55%	80%
10dB	Babble	3m	98.84%	80%
10dB	Music	1.5m	98.26%	80%
10dB	Music	3m	98.55%	80%
10dB	TV	1.5m	98.84%	80%
10dB	TV	3m	98.55%	80%
5dB	Babble	1.5m	98.84%	80%
5dB	Babble	3m	96.24%	80%
5dB	Music	1.5m	98.84%	80%
5dB	Music	3m	97.08%	80%
5dB	TV	1.5m	93.37%	80%
5dB	TV	3m	90.72%	80%

Table 1: Results of Hit-Rate

4. Conclusions

The complete implementation of an end-point voice commands recognition system has been presented that is capable to execute on a simple MCU device. The reference design enables local voice recognition without a network connection and allows to quickly start building an enhanced VUI in minutes that recognizes voice commands to trigger the corresponding operation.

References

[1] RA Family of Arm® Cortex®-M based MCUs [LINK]
[2] Renesas Voice User Interface (VUI) TechnologySNR [LINK]