EMERGING TECH CONFERENCE – Edge Intelligence

Volume 02, 2023, Page 89 – 92

Proceedings of Emerging Tech Conference:
Edge Intelligence 2023

Light Bulb control with DNN based voice user interface – the journey to design it

George Chardalias[1], Georgios Flamis[1], Vincent D'Agostino[2] and Suad Jusuf[1]
[1] Renesas Electronics
[2] D'Agostino Industries group
georgios.chardalias.ks@renesas.com, georgios.flamis.zc@renesas.com,
vindag@dagostinoindustries.com, suad.jusuf@renesas.com

Abstract

The purpose of a design journey is to show, in a demonstrative way, how to incorporate the DNN based voice user interface technology into a design. The base designs are by definition very simple to highlight the new technology and how it can be incorporated easily. Distinct lines are drawn between the application and the technology so the reader can more generalize how to incorporate it in their own design.

1 Introduction

Our goal with this Design Journey is to create a control for an LED-based light bulb based on the Renesas RA6/4/2 device family running Cyberon DSpotter for Voice User Interface (VoiceUI) technology. This technology is a non-connected speech recognition solution; it runs totally on the local MCU. While this limits its vocabulary and natural language support, it also allows for speech control of devices that are not connected to the cloud.

To accomplish this, we will design and build a circuit that can plug into any of the RA-Voice kit boards via the PMOD connector and write software to support the design. The software project accompanying this Design Journey is targeting the RA6 Voice Kit.

2 The Voice UI

One of the major differences between a connected, natural language solution, and a locally hosted speech recognition solution, is that a non-connected speech recognition system requires a stricter vocabulary set and syntax. This is because the interface is expecting certain words or phrases and cannot process natural language which can reorder words and still have the same meaning. This is the tradeoff for being able to run the speech recognition on an M-class device with a small memory footprint. Multiple language support (Cyberon supports 44 at the time of publication of this document) can be achieved by loading multiple language sets and switching between them at runtime.

For our lightbulb, we have created a system that, depending on the desired outcome, will take two or three commands to complete a task. In graphic form, it looks like the following drawing:

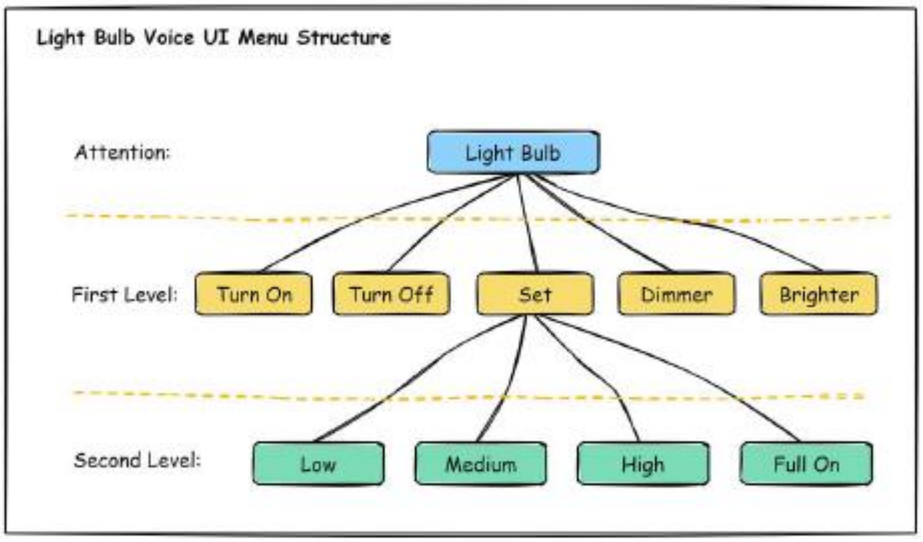

Figure 1 Voice Command options

As can be seen above, an attention word is first used to let the system know a command is coming. There are four commands (“Turn On,” “Turn Off,” “Brighter,” and “Dimmer”) that require no more information and complete the command. If the command word “Set” is used, then one of four preset levels is required next, to complete the command. “Dimmer” and “Brighter” traverse the presets one level at a time, and “on” and “off” move between extinguished and the last used preset level. Let’s look at what that means:

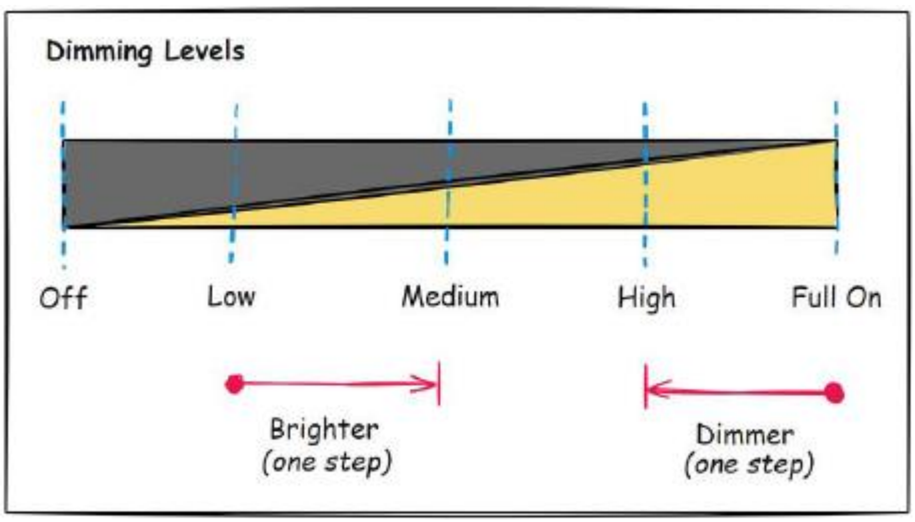

Figure 2: Light dimming levels

This defined command structure is easily created using the Cyberon DSpotter Modelling Tool and tested for recognition with their powerful command testing tools.

Unlike a “trained voice” solution, Cyberon has “pre-trained” the system for units of language sound, called phonemes. Each language has a set of phonemes. English, for example, has between 42 and 45 phonemes. Italian, by contrast, has 32-36. The DSpotter Modelling tool breaks down command text into these phonemes, which are pre-categorized. There is no need to take thousands of voice samples to train the system as you would with traditional speech recognition.

3 Hardware design

The hardware design for the bulb is relatively simple because the RA devices have timers with PWM outputs which can drive transistor gates directly. The use of transistors is driven by two things: firstly, the GPIO pins of any processor cannot handle the currents required to drive an LED at illumination levels, secondly, the forward voltage drop of white LEDs is very close to the 3.3V Vcc used by the processor. Typically, the LEDs are driven with a higher voltage. The processor pins cannot tolerate this higher voltage, so to isolate the I/O pins from both this voltage and current requirement, we can use small MOSFETs to drive the LEDs.

Since the PMOD connector has ground and 3.3V, we must create a higher voltage. For this, we are using a Renesas part on our plug-in board, the ISL9111A buck-boost controller, with the output set to 5.25VDC. The following illustration shows the overall design of the LED plug-in board.

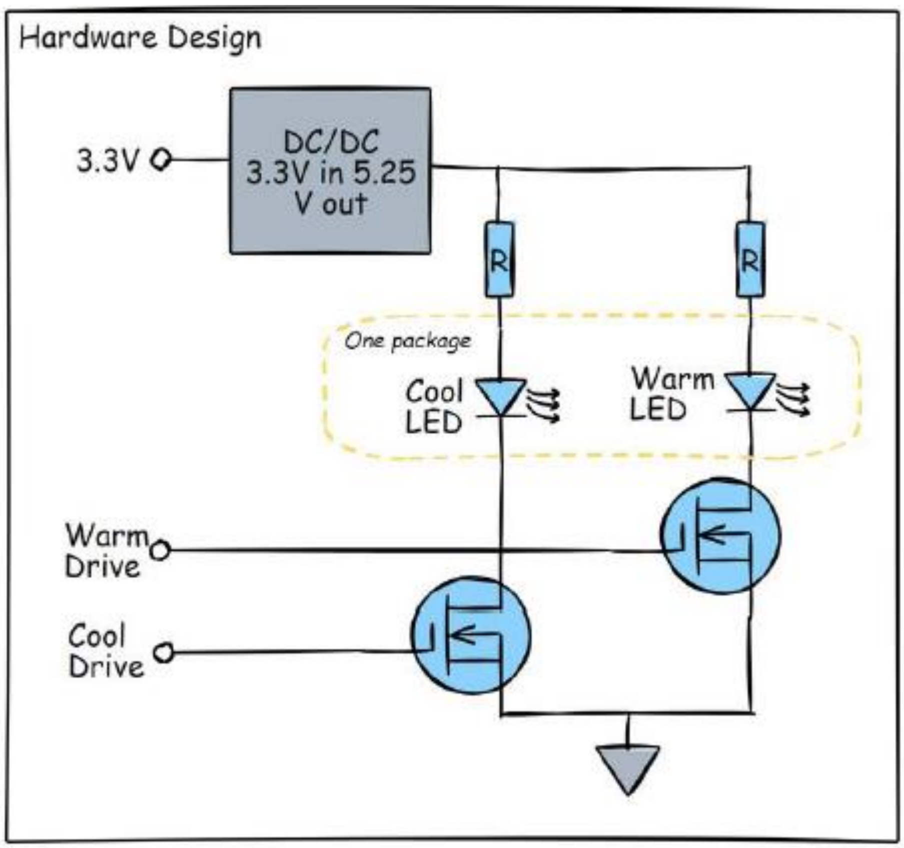

Figure 3: Drawing of the PWM driver

We will drive the two transistor gates with the output from a PWM running with a carrier frequency of 4kHz. This is well above the optical detection of the eye – even in the periphery where we are more sensitive, and well below where the switching transitions occupy a high percentage of overall time (this, in the end, is what makes transistors run hot – not what percentage of time they are on, but the percentage of time they spend in the linear region during switching).

The hardware for the processor and its peripherals is the one supplied with the voice kit. It has a processor, two digital and two analog mics, and a PMOD connector. Our LED board plugs into the PMOD connector. The other connection is the USB-micro to the PC for debug and serial monitoring (the JLINK-OB debugger creates a VCP along with the debug connection).

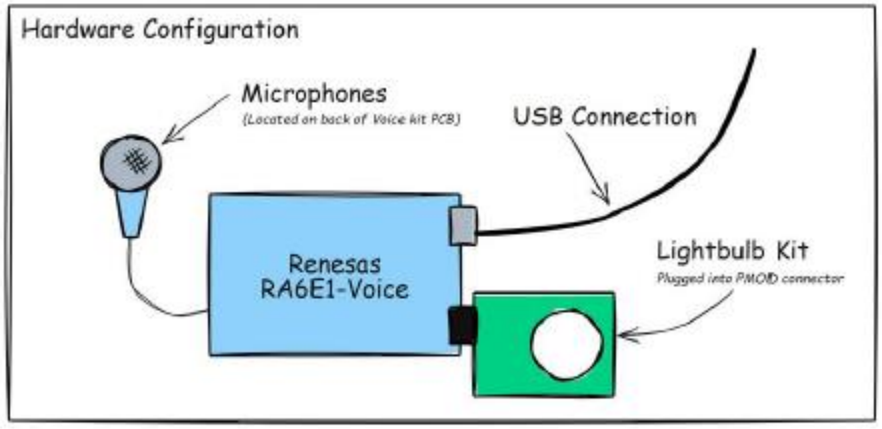

Figure 4: Hardware Configuration

4 Conclusion

Over the previous sections, it has been shown that incorporating the Renesas/Cyberon VoiceUI into products is fast, easy, straightforward, and very maintainable. In this design, the use of a linked hardware solution makes it easy to do two time-dependent tasks together without the overhead of an RTOS to manage the processes.

While this solution generates data for the outside world (the PWMs) the same opportunity is there for data acquisition with digital and analog peripherals.

Renesas has always had a rich set of high-performance peripherals: both digital and mixed mode. Together with the Event Link Controller (ELC) in the RA family, these peripherals can be configured to do many functions independently.

5 References

[1] VOICE-RA6E1 Engineering Manual, Renesas Electronics, 2022

[2] RA6E1 Group Datasheet, Renesas Electronics, 2022

[3] Cyberon Speech Recognition Technology, Renesas Electronics, 2023

[4] Endpoint Embedded Voice User Interface Solution, Renesas Electronics, 2022

[5] Designing Industrial IoT Applications with TinyML, Mouser Electronics, 2023