A Stroke of Genius for Smart TVs: Add Voice to Your Remote Control

Introduction

In this age of the Internet of Things (IoT), connected devices are getting smarter and smarter. We have smart phones, smart homes, smart cars, smart appliances, and even smart TVs. But this last example begs the question: If my TV is so smart, why is my remote control so dumb?

Anyone who has tried to use a remote control with a smart TV for more than simply watching their favorite program has probably been frustrated with the experience. Even setting up its Internet connectivity can be daunting. And forget about trying to enter a URL address in the browser. Yes, some TVs allow you to use a keyboard or even a smart phone, but none of these connections are simple or convenient. It's even a bit embarrassing when friends want to watch the big game but you're fumbling with buttons.

Remotes these days bring back memories of when PCs had the "C:>" prompt. The jump to friendly, GUI-based operating systems was a giant leap ahead for most PC users. It's time for remotes to follow suit. The question is: how can this be accomplished?

History of Remote Controls and What Makes Them "Smart"

The first wireless TV remote controls can be traced back to the 1950s with the ultrasonic Zenith Space Command. These RF-based controls were replaced by infrared (IR) technology starting in the 1980s, and unbelievably, what we use today is largely the

same. While there have been some changes in technology, the majority of modern remote controls are still IR-based and the user experience is roughly the 1980's one.

To enhance the end user experience, some TV manufacturers are implementing more advanced features on remotes such as two-way RF communication, no line-of-sight restrictions, and even QWERTY keyboard interfaces. However, manufacturers aren't advancing remote control features to match the capabilities inside the TV.

What's next presents a new level of remote functionality and ease of use: voice control. When remote controls are truly capable of "hearing" a user's voice command and translating it to a TV command, the remote functionality and ease of use can unlock what's on TV.

Remote Control Voice Recognition Benefits

Adding voice recognition to a TV remote control changes the whole user experience. And if it works correctly, every change is a good one.

Without voice recognition, most current remotes present frustrating exercises in button pushing, transmit delays, lost progress, painful spelling exercises, and so on. It's even worse if the room is dark!

With a voice-enabled remote the interaction becomes very fast since the user simply activates the remote and speaks a command, which can fall completely outside the TV's menu structure.

For example, from watching a program, a user could press an activation button on the remote and say something like, "Record the program 'Big Bang Theory' tonight at 7 pm." That's it. In the old paradigm, the user had a long, arduous process to accomplish this goal. With voice, it's just a few steps: 1) activate remote, 2) speak command, and 3) confirm action.

How Does Voice Recognition Work

How does voice recognition in a handheld work? Good question. It's not as obvious as we might think. The processing power and data required to perform voice recognition is beyond the scope of most remote controls, TVs, and even smart phones. In fact, voice recognition in today's smart phones is actually accomplished through cloud computing.

Remember the old days of voice tagging when you recorded a voice command and then linked it to a task such as dialing a number from your contact list? In theory, you could say, "dial Ken," and if you were lucky, the cell phone would then "dial Ken;" however, more often it would announce, "dialing Ben," and you would throw the phone out the window.

Voice recognition has progressed considerably in recent years and leaders in voice recognition include companies like Nuance Communications, Microsoft, Google, Amazon, and many others. When we use Siri, Google or Alexa for voice control, these applications digitize our voice and send it over the Internet where it is processed for a response. The complexity of this exchange is illustrated in Figure 1.

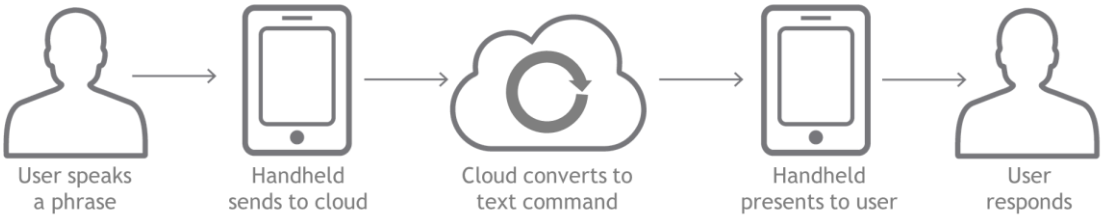

Figure 1: The Journey from User Voice to Cloud to User Action

In fact, with always-on features, simply saying "OK Google" from a Google web page or Android OS phone can trigger a search in which your voice command is digitized, processed in the cloud, and then converted to text for the search command. A key factor enabling voice commands in the TV market is the fact that smart TVs are already connected to the Internet and can leverage this considerable infrastructure.

The Need for Voice in a Remote

Based on the fact that smart TVs have internet connections, you may ask, “Why do I even need a remote? Shouldn’t I be able to control my TV by just speaking to it since it’s now connected to the Internet?” The answer is “yes,” but there are several issues with that solution.

First off, for a TV to recognize voice directly without interaction from a remote control would require the TV to listen all of the time. Some TVs can do this today, and in fact do; however, the unanticipated consequence of this functionality has received negative press about privacy.

For the TV to constantly listen and decode user conversations for commands it must constantly send those conversations over the Internet. While this isn’t unusual, the feature did not apply adequate security and users’ conversations were wide open. Users are generally not aware of this, and if they were, they would either turn off the listening ability, or greatly curtail the content of their conversations in the rooms with their “listening” TVs.

Secondly, there are issues concerning the device’s ability to pick out commands from surrounding noise, or distinguishing voice commands from TV audio or background conversation.

By using a remote to initiate and stream voice commands users can greatly reduce these concerns since 1) the user proactively and knowingly engages with the TV remote control, and 2) the user is holding the remote control, which is designed to pick up sound from inches away and not from across a room.

Technology and Cost

The next question is, “With all these benefits, why are there not more voice remote controls?” Infrastructure, technology and cost are three key factors.

Infrastructure: Even if voice recognition is supported by hardware in the home, the back-end infrastructure to support it must be in place. This means the TV provider would need

to develop a voice recognition engine or pay for the service from a third-party. In the latter case, the user command would be translated to a text-based string that would need to be decoded by the TV into commands. The good news is this process is becoming more mainstream as operators try to differentiate themselves and improve the user experience.

Technology: As we all know, there are hurdles associated with getting voice recognition translated to text commands correctly, but these are quickly being overcome with the cloud computing process and dominant providers mentioned above. Given time and third-party intelligence, this hurdle is becoming smaller.

There is also the question of what wireless technology can get the voice data from the remote control to the TV or available internet connection without killing the battery life.

Typical voice recognition systems require 16-bit ADC resolution with 16 kSps, which results in 256 kbps of data. This means that unless the wireless technology has throughput of at least 256 kbps, some compression will be required.

Handheld IR rates are typically not sufficient for data bandwidth requirements; however, by using compression to accommodate the throughput requirements, wireless technologies like ZigBee® Remote Control have sufficient data rates and offer excellent battery life. I will talk more about this later.

Cost: It always comes down to cost: cost for the infrastructure, cost for the TV, and cost for the remote.

More About Remote Control Cost

Adding voice capability to a remote can double the bill-of-materials (BOM) cost of a standard RF remote control. A voice-enabled remote needs to support RF, add a

microphone and codec, and include supporting circuitry.

The following examples show block-diagram comparisons between IR, RF and RF+Voice. The IR-link capability always remains in each remote control, with RF or RF+Voice and associated BOM differences shown.

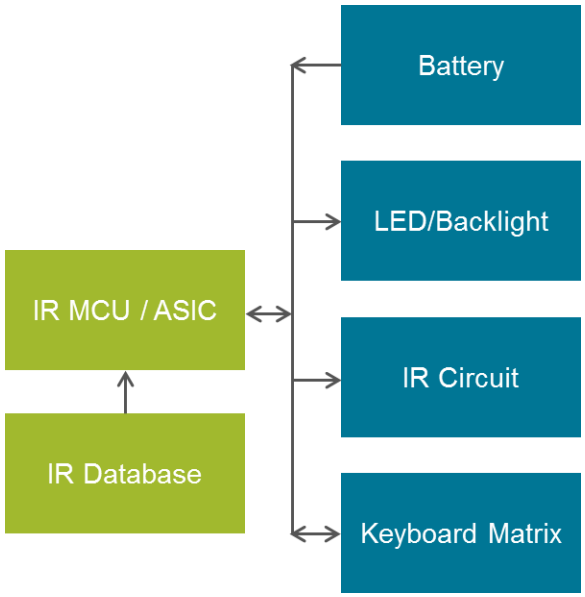

Figure 2: Example of IR Remote Control System

Figure 2 is a typical IR remote control block diagram. These are built with very low-cost MCUs or ASICs for IR control. In some cases, they will have additional nonvolatile memory that contains IR database codes needed for different devices such as TVs, DVD players, and so on. (Think “universal remotes.”)

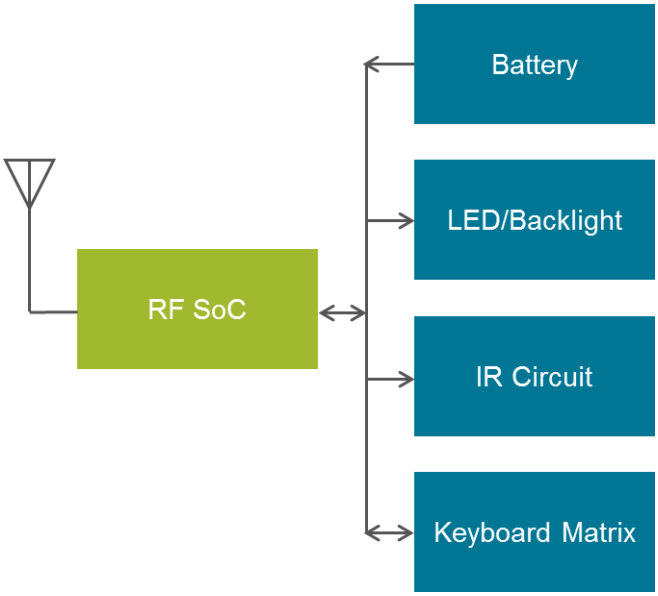

Figure 3: Example of RF Remote Control System

Figure 3 builds on the IR block diagram but replaces the microcontroller in the IR with an RF System on a Chip (SoC) and add an antenna.

While an RF SoC is typically more expensive than an IR MCU, the additional cost can be offset by the fact that the large IR database does not need to be stored, thus removing the nonvolatile memory cost. RF remote controls can download the required control codes from the TV or cable/satellite box over the two-way RF link. The TV and cable/satellite boxes have much more available memory to store codes, or can even pull data from the cloud. Pulling information from the cloud also allows for updated codes for newer devices that may not have been supported when the device was configured.

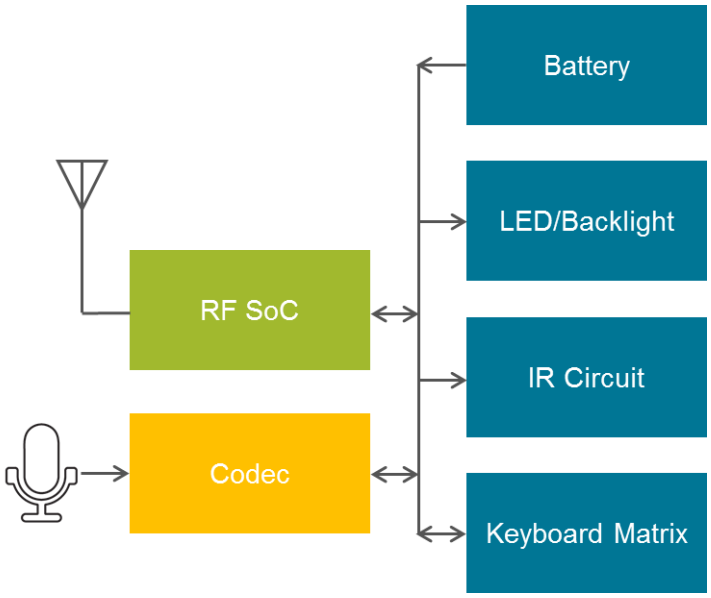

Figure 4: Example of a Voice-Operated Remote Control System

In Figure 4 we add voice capability to the RF remote control by inserting a hardware codec and microphone(s). These devices can significantly increase BOM cost. However, with the increased processing capabilities of today's wireless SoC chips, we can look at alternatives to hardware codecs. For example, the Silicon Labs EM341 ZigBee SoC is based on a Cortex M3 processor and has enough processing capability to handle not only the RF remote control requirements but also a soft codec.

Voice-Enabled Remote Control Example

Let's take a look at a full-featured remote control reference design that supports IR, RF and voice capabilities. In this case, we are digging into the Silicon Labs ZigBee Remote Control reference design (EM34X-VREVK). This ZigBee Remote Control device supports voice, IR with IR database, backlit keyboard, and an acceleration sensor for activating the backlight.

Figure 5: Silicon Labs ZigBee Remote Control Reference Design

For voice audio, we need to support 256 kbps throughput. ZigBee has a data rate of 256 kbps but actual throughput is typically 100 kbps or less for a point-to-point link. This means we need 4:1 compression on the audio before sending it over the air.

The reference design uses a hardware codec, mic, and voice capability. However, the RF SoC EM341 also supports a software codec that can provide significant cost savings with no feature reductions. The software codec is based on connecting the digital PDM (pulse

density modulation) microphone directly to the EM341's SPI and GPIO pins as shown in Figure 6.

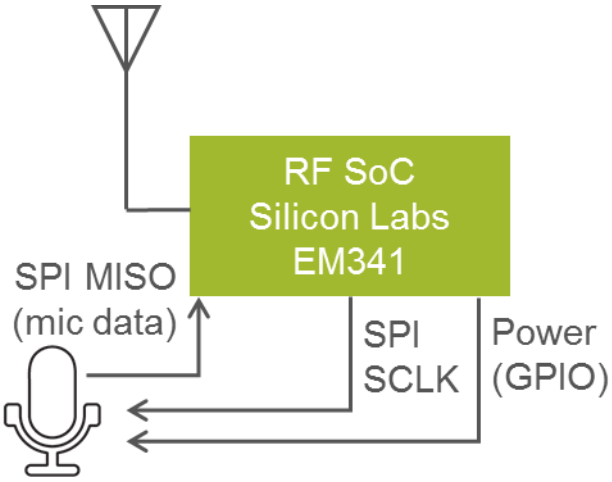

Figure 6: Connecting the PDM Microphone to the EM341 SoC

The EM341's Cortex M3 handles the PDM to PCM (pulse code modulation) filtering/decimation, equalization, and compression processes. The complete procedure from PDM output to ZigBee transmit is shown in Figure 7 and provided as a free library for the Silicon Labs ZigBee Remote Control application profile.

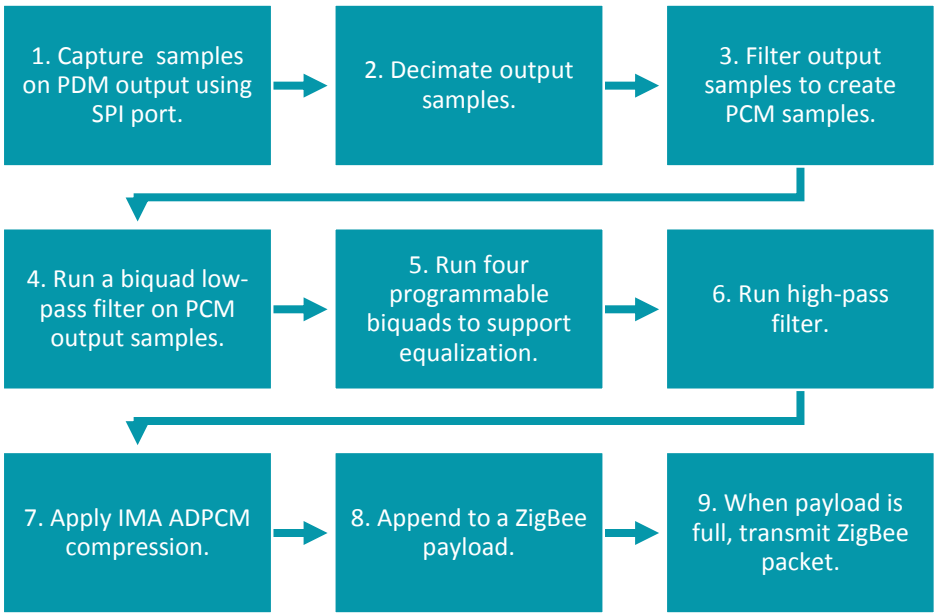

Figure 7: Process overview for PDM to ZigBee packet translation.

By using the software codec functionality, the performance meets the requirements for voice-to-text engines, but at a much lower cost. See Table 1 for estimated BOM savings.

	Software Codec	Hardware Codec	Units / Comments
Microphone Interface	SPI (SCLK and MISO only)	SPI 4-wire + I2C 3-wire	
GPIO needed	3	7	
Cost	$0.00	$0.50 - $1.50	Volume dependent
Audio Compression	IMA ADPCM	MS ADPCM	
Sensitivity @ 0dB digital gain	-26	-26	dBFS @ 94dB SPL
SNR (Required)	20	20	dB
SNR (Measured)	30.0	32.5	dB
EIN	64.0	61.5	dB SPL
THD	1.2	2.0	%
Frequency Response (-3dB)	160	160	Hz
	7.2	7.2	kHz
Maximum Acoustic Input	120	120	dB SPL
Minimal External BOM	· Digital Mic (INMP421) · 0402 Capacitor	· Digital Mic (INMP421) · Codec · PFET · +1.8V LDO · 0402 Schottky Diode · 0603 Capacitor · 0402 Capacitors · 0402 Resistor	

Table 1: Hardware/Software Codec Comparison

Summary

Voice control in a handheld TV remote changes the user experience positively. The process of adding voice control is challenging but has gotten easier in recent years with advances in voice recognition software.

The Silicon Labs voice-enabled remote control reference design has the complete remote in finished form factor with ZigBee-approved software libraries included. It also furnishes voice codec functionality with a soft codec, which saves money and manufacturing complexity without sacrificing performance.

More information on Silicon Labs ZigBee Remote Control solutions at http://bit.ly/1O4YdED