Smart speaker fundamentals: Weighing the many design trade-offs

Texas Instruments

Wenchau Albert Lo
System Engineer, Personal Electronics
Texas Instruments

Mike Gilbert
End Equipment Lead, Personal Electronics
Texas Instruments

There's no denying that voice-enabled speakers – better known as smart speakers – are a hot consumer product.

According to market research firm eMarketer, 35.6 million U.S. consumers used a voice-activated device at least once a month in 2017, with that number growing at a compound annual growth rate of nearly 50%.

Future market predictions are also optimistic. Juniper Research predicts that smart devices like the Amazon Echo, Google Home, Apple HomePod and Sonos One will be installed in a majority of U.S. households by the year 2022. They add that 70 million households will have at least one of these smart speakers in their home, with the total number of installed devices topping 175 million. This is especially impressive for a product category that didn't exist before November 2014.

But there's a lot more behind these simple-looking devices than microphones and speakers combined with an internet interface. Smart speakers incorporate many electronic functions implemented by dozens of sophisticated integrated circuits (ICs). Original equipment manufacturers (OEMs) entering the smart-speaker market with a differentiated product must make decisions about what to include; how to include it; and what trade-offs are acceptable in such a small, low-power device.

What does a smart speaker actually do, and how is it used in a home? In the simplest terms, it captures and digitizes the end user's voice command, passes results up to a web-based cloud service for interpretation, and then responds to end users by acting on the command or responding with results. Smart speakers can also search and play audio content from the web or from a *Bluetooth*®-connected device. As **Figure 1** illustrates, many smart speakers can now interact with other devices in the home, such as lights, door locks and climate-control systems.

Figure 1. *As a media player, smart speakers must be simple and elegant with quality sound. As a smart home hub, they must provide accurate voice recognition and connectivity to the entire suite of smart devices in the household.*

It's not just for this cycle that OEMs want to differentiate their product; instead, there's a battle to control the access and flow of information within the room, if not within the entire home, as the sole media and home automation hub.

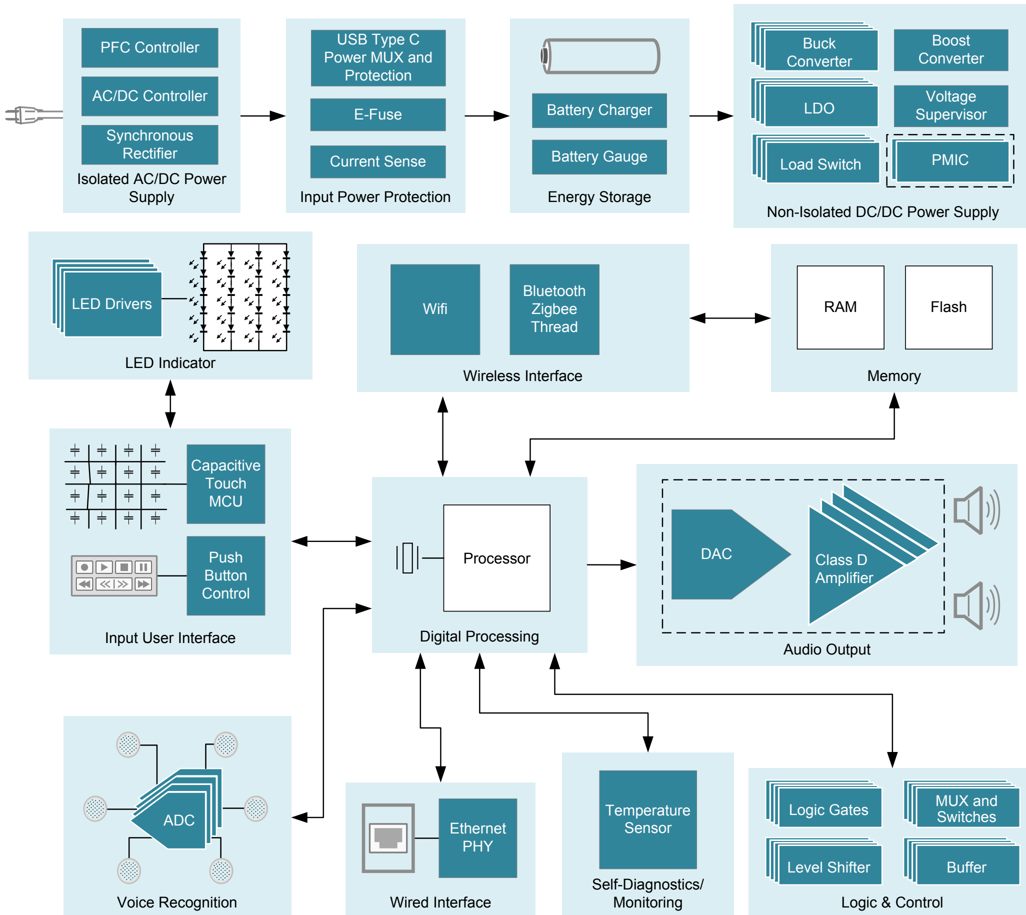

Figure 2. *TI's smart speaker system block diagram.*

Making the smart speaker real

Smart speakers require a considerable amount of circuitry to work – and to work well. It's a complex array and interconnection of analog, digital, mixed-signal and power-management subsystems, interfaces and more (**Figure 2**).

There are numerous design issues to address, including the number and type of microphones, audio output and speakers, power management, the user interface and wireless connectivity. For OEMs, the first question is whether to use a "black-box" chipset that includes a system-on-a-chip (SoC) for audio decoding and signal processing, a combination Wi-Fi® and *Bluetooth* radio with a microcontroller (MCU), and sometimes, a custom power-management IC (PMIC). This "canned" solution does not allow for much product differentiation, however. So let's look at some of the design areas and challenges.

Microphones

When choosing a microphone technology, the trade-offs that may not be all that clear. It's a choice between:

- A microelectromechanical systems (MEMS)-based "analog" microphone with an integrated preamplifier, paired with an external 24-bit audio analog-to-digital converter (ADC) that outputs formatted digital code to the SoC.
- A MEMS-based "digital" microphone with an integrated single-bit, first-order delta-sigma modulator ADC that outputs a pulse-duration modulation (PDM) digital bitstream that requires further filtering to create the formatted digital code. Either the SoC or a digital signal processor (DSP) specialized for voice recognition would have to handle this filtering. A stand-alone voice DSP offloads significant processing from the SoC, but also adds cost.

A digital microphone is more expensive than its analog counterpart, but the analog microphone requires an additional ADC in front of the SoC. Compared to an analog microphone with a separate ADC, a digital microphone also has a lower signal-to-noise ratio (SNR) and dynamic range, given both limitations on the transducer size to accommodate the ADC inside the microphone package and the performance limitations of the integrated ADC itself. Common digital microphones have an SNR on the order of 65-dB and dynamic range on the order of 104-dB, and since the ADC is integrated, there is no opportunity to enhance the SNR or dynamic range further with filtering and oversampling.

The analog microphone, on the other hand, combined with an external ADC, can experience an SNR or dynamic range (both are synonymous in an ADC) up to 120-dB. This external ADC is often a multi-channel precision audio, 24-bit ADC, employing third- or fourth-order delta-sigma modulators with high oversampling capability. They also integrate programmable, complex digital decimation filters; PGAs with configurable automatic gain control; and mini DSP for additional noise filtering and equalization. Given that a typical crowded room, or one with music playing, could easily have 60-dB of ambient sound levels, the lower dynamic range of the digital microphone may result in an inability to properly recognize voice commands unless they are significantly higher than the ambient sound (meaning that the end user would have to get closer to the microphones or the smart speaker would need to use more microphones).

Going from a 104-dB dynamic range to 120-dB has some amazing benefits that need to be put into perspective. A 6-dB improvement in dynamic range can double the voice-recognition range. At some point, increasing range is not practical or useful, but you have more dynamic range to work with. Adding another 14-dB of dynamic range will allow you to save cost by reducing the number of microphones required. In addition to being more costly, adding digital microphones may also be prohibitive by adding to the complexity of the layout by routing three signal traces (data and clock) for each microphone pair to the SoC, and by the number of PDM inputs available on the SoC itself. Add to that the fact that each trace can pick up and/or radiate noise, making electromagnetic interference a greater concern. Lastly, the clock lines that run to each digital microphone can introduce challenges with routing and with jitter. Today's analog microphones have differential outputs enabling common-mode rejection to the signal trace routing. The ADC also provides a bias supply for each microphone, reducing the complexity of the power tree for the array.

The combination of increasing the microphone range and sensitivity by using analog microphones with precision ADCs not only reduces cost and complexity, it can dramatically reduce frustrating command-recognition errors across a variety of noisy environments. As second-generation smart speakers start rolling out, this error rate will increasingly become an important market differentiator.

There's no need to reinvent the wheel when it comes to implementing multi-microphone designs and speech recognition. TI's PCM1864-based circular microphone board (CMB) reference design, shown in **Figure 3**, uses two 4-channel audio ADCs to interface with an array of up to eight analog microphones, and can extract clear user voice commands within noisy environments.

Speaker amplifiers and power

For the speaker amplifiers, there are trade-offs between output power (typically between 5 W and 25 W), power consumption, heat, size, speaker protection and sound fidelity.

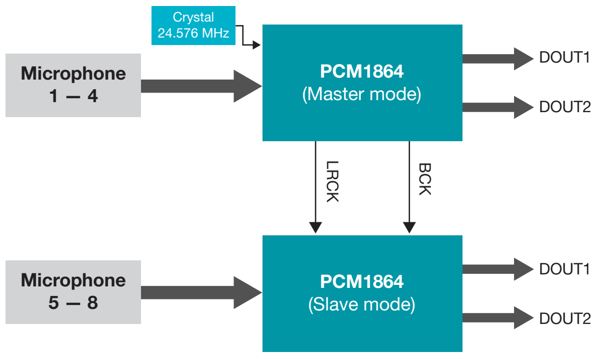

Figure 3. *Circular microphone board reference design.*

A simple speaker system with a single mid-range tweeter and woofer can produce good sound, while multiple speakers, combined with the latest audio-processing techniques, can offer a 360-degree audio experience.

You also have a choice between implementing a one-time room calibration to tune and optimally match the speaker's spectral characteristics, or taking a more complicated adaptive-tuning approach that compensates for movement within the sound area. [The TI PurePath™ Console] graphical development suite provides easy one-time tuning with impressive results.

On the power consumption and heat side, one approach to reducing ongoing power drain is to combine amplifier pulse-width modulation schemes with adaptive power supplies to reduce the power requirements of the speaker. This technique uses a variable (not fixed) switching frequency for the Class-D output, with the frequency change based on the audio content. In other words, more content results in more switching, and less content results in less switching.

To add efficiency, you can also dynamically adjust the amplifier's output power-supply voltage based on content. This technique is called envelope tracking. It tracks the audio content and boosts the voltage (output power) only when the music needs a boost in power, especially in bass-heavy parts (with lots of peaks in the signal content).

The [stereo evaluation module reference design of the digital input, class-D, IV sense audio amplifier] shown in **Figure 4** not only accepts digital inputs in multiple formats and delivers high-quality audio, but its Class-D topology includes additional features that minimize power consumption across a range of output levels without diminishing fidelity and performance.

Power management

As with most electronic systems, power management plays a significant role in system design. The ultimate goal is to provide power efficiently to dissipate less heat, allowing for a smaller and lower-cost system and, in the case of portable systems, extending battery run time. SoC and Wi-Fi chipsets are sometimes bundled with a dedicated PMIC, but you may still prefer the added board layout and supplier flexibility of a discrete implementation using individual DC/DC converters, low-dropout regulators and voltage supervisors to modify functions such as sequencing, change board layout, and reduce noise and/or cost.

You may also want to optimize the design beyond what a fixed, integrated solution offers, such as operating with lower quiescent current or using a higher switching frequency (such as 1.4 MHz up to 4 MHz) to achieve a smaller footprint, given the

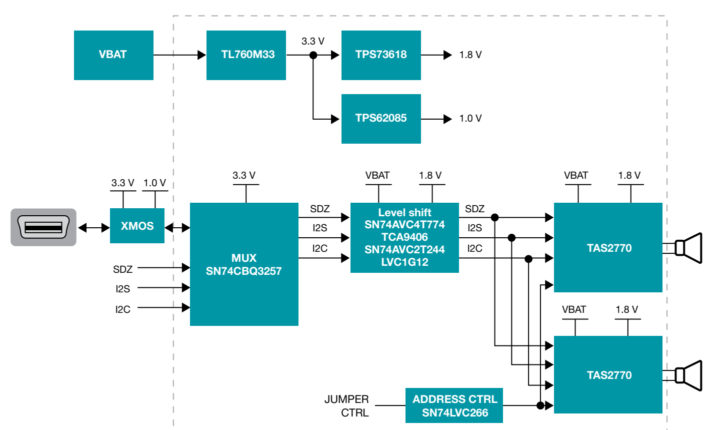

Figure 4. *Stereo evaluation module reference design.*

need for smaller inductors. Or you may want to use pulse skipping or eco mode to save power under light loads, while at the same time staying out of the audio band by not switching below 20 kHz (which may result in audible noise). Further, you may also desire system input voltage flexibility. These amplifiers require a 12-V to 24-V power supply that comes from an internal power supply or an external power adapter.

An internal AC/DC supply may provide the main power, but an external AC/DC wall adapter with 12- or 5-V outputs is more popular, depending on the speaker power required. This main power can be supplied through a Micro USB connector for low-power speakers or the newer streamlined USB Type-C™ for higher-power speakers, replacing the bulky traditional wall AC/DC adapter and barrel jack. Since these adapters can be different power levels, implementing USB Type-C would require some level of handshake from the speaker to the adapter, or employ input USB current-limit switches or battery chargers with integrated overcurrent and overvoltage protection.

For portable speakers, a technique called power-path management allows an external AC/DC wall adapter to charge the battery while powering the speaker "live" through an integrated regulator. If you need a higher speaker amplifier power rail (such as 12 V or 18 V), one option is to use a two-cell battery at 8 V, then boost it as needed for the speaker amplifier. The battery charger will need to boost the input to the higher battery voltage (if the adapter output is 5 V), and you'll need an additional boost converter for the speaker amplifier rail to achieve higher voltages during peak power conditions. In addition, the portable smart-speaker system must have a low standby-power rating and efficient step-down converters to provide a longer run time between recharge cycles when batteries are the sole power source.

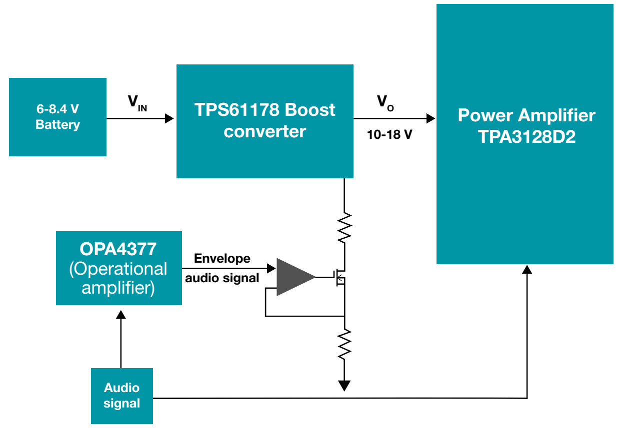

Figure 5. *Envelope-tracking power supply reference design.*

Since the speakers are the dominant power consumer, a power supply that is closely integrated with the needs of its amplifiers results in a power- and cost-efficient design. The envelope-tracking power supply reference design for audio power amplifiers shown in **Figure 5** is a good example of such a solution: It operates from a 5.4-V to 8.4-V input-voltage rail and delivers 2 × 20 W into an 8-Ω load (using a 7.2-V rail). In addition, it maintains high efficiency across the output-voltage range by changing the output voltage in accordance with the peak-to-peak envelope of the audio signal. Thus, it dynamically adapts the power amplifier's supply based on audio content, optimizing its power consumption.

User interface

You must decide what type of user interface to offer based on the desired end-user experience, since the human machine interface is a major factor in a smart speaker's market differentiation. The interface can range from lower-cost simple buttons and single-indicator LEDs, to an array of rotating LEDs, to a small LCD display, to an LCD display with touch inputs and haptic feedback.

LEDs are used at minimum to indicate status, and more recently, to enhance the end-user experience by generating moving colors in various patterns.

Although simpler systems may use single-color LEDs, most use red-green-blue (RGB) LEDs. If you chose a multicolor option, you will need to decide how many RGB LEDs to include, and whether the system processor, MCU or a newer multi-LED driver with integrated LED engines will control them; each choice brings cost, power and system-burden trade-offs. Using an integrated LED pattern engine offloads the processor as it manages pattern generation and drives an array of RGB LEDs even when the processor or MCU goes into low-power idle modes.

As shown in **Figure 6,** the [various LED ring lighting patterns reference design] illustrates how to design a multicolor RGB LED ring pattern subsystem using new multichannel RGB LED drivers with an integrated LED engine. The use of an ambient light sensor IC automatically controls the LED brightness.

Corresponding panel push-buttons may be inexpensive, but they are prone to mechanical failure and limited to a single function. They require that end users "push and hold" to effect action (up, down, scroll), an operation that is now dated and counterintuitive in the world of smartphones. In contrast, a capacitive touch-sensitive surface allows for more interaction and enhances the user interface. No physical force is required, and the same surface can detect end-user proximity and enable a backlight for easier use in the dark. A touch-sensitive surface can implement a more familiar interface by supporting "swipe" or "spin" instead of simple push, and offering it can help differentiate a smart speaker. A properly designed capacitive-touch controller works on variety of surfaces (plastic, glass or metal) and can be designed flush with the speaker case surface.

The [gesture-based capacitive touch interface reference design for speakers] shown in **Figure 7** provides an easy-to-use evaluation system for a multi-gesture capacitive-touch interface for smart speakers using TI's capacitive-touch MCU. The design allows for tap, swipe, slide and rotate gestures.

Wireless connectivity

Finally, there is the issue of literally getting outside the box. Without a connection to the internet, a smart speaker cannot function as intended. There are design decisions here regarding the best way to connect, given speed requirements and power constraints.

The most common form of smart speaker connects to the internet directly via Wi-Fi. Here, the bandwidth of IEEE 802.11n is more than sufficient, and it also allows for a multi-room wireless speaker mesh connection. However, Wi-Fi power amplifiers consume significant power and may limit the play time of battery-operated smart speakers. For this reason, Wi-Fi-enabled speakers are often plugged directly into wall power outlets or have AC adapters for continuous operation.

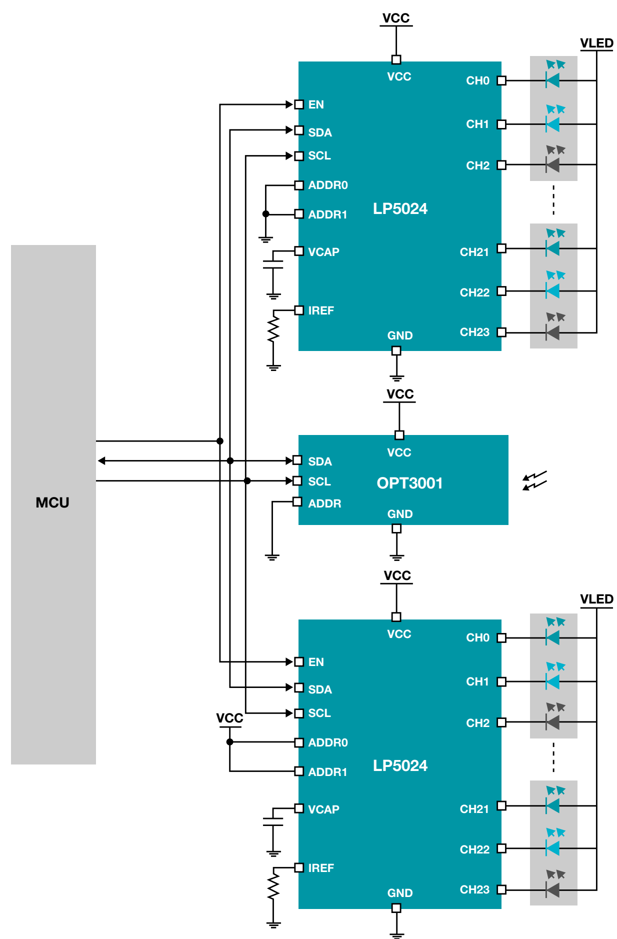

Figure 6. *Various LED ring lighting patterns reference design.*

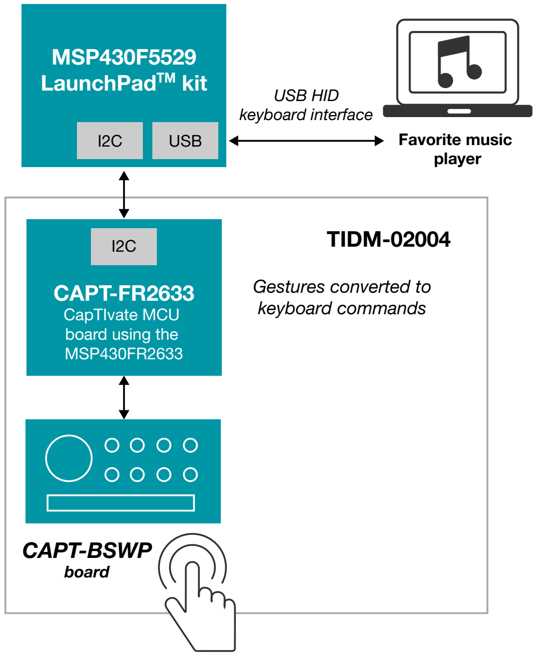

Figure 7. *Gesture-based capacitive-touch interface reference design.*

End users who want to use multiple smart-speaker units (for better room coverage or stereo sound quality) will need IEEE 802.11n/s support to implement a mesh network. In a mesh network, any one speaker can become the master (connected to the cloud) while the others act as slaves. If a speaker acting as master is powered off or loses the network, the mesh automatically assigns another speaker as master. The biggest challenge in a multi-speaker mesh network is synchronization. Wi-Fi controllers in a mesh network must have robust synchronization schemes to avoid significant user frustration.

Battery-powered portable speakers may offload Wi-Fi cloud connectivity to nearby mobile devices. For connectivity to mobile devices for indirect cloud connection and/or to listen to content stored on a mobile device, *Bluetooth* Classic (or *Bluetooth* Basic Rate) is required for continuous connection to stream the audio content, due to the bandwidth limitations and power schemes of *Bluetooth* low energy. When used in conjunction with *Bluetooth* Classic, *Bluetooth* low energy can control communication between devices.

Home automation is another function that currently exists in many homes as a separate entity – a stand-alone hub connected to the internet via Wi-Fi, as well as to specialized lights and thermostats via a wireless mesh network set up for home automation implemented by such standards as Zigbee®, Thread, Z-wave, etc. Smart speakers can legitimately lay claim to providing home automation via the internet as long as this additional stand-alone hub is also implemented.

However, to eliminate the need for end users to purchase this additional wireless hub, smart speakers can become the home automation hub by simply adding a multiband wireless MCU with an integrated RF power amplifier. The wireless MCU handles processing of the protocol stack and controls the radio, preventing the need to burden the existing SoC or Wi-Fi network processor while enabling communication through any of the popular long-range home-automation protocols, both in the 2.4-GHz and Sub-1 GHz bands. Because Wi-Fi and *Bluetooth* also use the 2.4-GHz band, you'll need to ensure co-existence through a combination of hardware and software built into the integrated wireless MCU.

Looking to the future

The smart speakers of the future will be more than stand-alone, audio-only units. As flat-panel TVs become increasingly thinner, their speakers need to be smaller, which negatively affects the TV's sound. As a result, soundbars (which enhance the sound of flat-panel TVs) are increasing in popularity. Adding voice recognition is the obvious next step of soundbar evolution.

To complete the picture, smart soundbars will incorporate a set-top box for wireless video streaming with only a single HDMI cable connected to the TV, which then acts as an extremely large display monitor. As flat-panel TVs become even thinner, the TV control circuitry and power supply may also be implemented into smart soundbars. Then, the smart speaker and smart soundbar will compete to be the hub for the overall home entertainment system. With the added connectivity for home automation, these devices will also compete to become the automation hub for smart homes.

Another added feature is smart speaker display. Adding display to the smart speaker is a natural extension of its functionality. Just as center console displays are proliferating in automobiles, consumers will demand the additional visual experience from a home informational/entertainment device as well. Additionally, the way in which content will be requested and displayed differs from that of a more personal, handheld smartphone or tablet experience. Since voice commands are the primary mode of requesting content and control, simplified search and control applications will be necessary to facilitate quick and accurate results. Further, displayed images can be simplified, with minimal need for touch interaction while also providing a large enough image suitable for distance viewing. This will allow consumers a more pleasurable experience when interacting with the smart speaker and will deliver crisp visual content.

With this added display functionality, smart speakers may then concede to the smart soundbar in the living room and focus on areas outside of the living room. Smart speakers may provide smaller personal displays, ranging from integrated LCD screens to larger ultra-short-throw high-definition projection by using TI DLP® technologies to create large displays on any surface. Smart devices located near high-traffic areas like the kitchen or family room will need to be aesthetically pleasing and nonintrusive. The addition of a tablet-size or larger flat-panel display generally may not always meet these criteria. Projection display technology gives end users a more interactive experience when asking the smart speaker for information (weather, recipe, traffic) and puts a face to the anonymous voice. In this way, the smart speaker's role and importance in the home continues to morph and grow, bringing with it opportunities for designers to start new trends and differentiate their designs.

For more information

- Explore **TI solutions and design resources for smart speakers**.