Dr. Paul Beckmann,
Aaron Grassian,
Matt Crowley

White Paper

Advancing Always-On Voice Command Systems with Ultra-Low Power Solutions

A Comprehensive Guide to Voice Command Systems in Portable and Battery-Powered Products

Contents

Executive Summary

More than ever, consumers are now seeking out always-on voice command products to enable a safe, and natural interaction between themselves, and the digital world. With COVID-19 spreading across the globe, the human behavior is shifting towards hand-free devices to avoid physical contact with public screens, buttons, and controllers.

As a result of this increased awareness of how public surfaces can transmit diseases, always-on devices are proliferating into smart cities, smart homes, and industrial applications.

However, until recent advances in power-efficient hardware and software were developed, portable and battery-powered products could not make use of voice command systems that are always listening. This white paper describes several breakthroughs in applications, techniques, and technologies that have made ultra-low-power voice command products possible.

SPOT™ by Ambiq

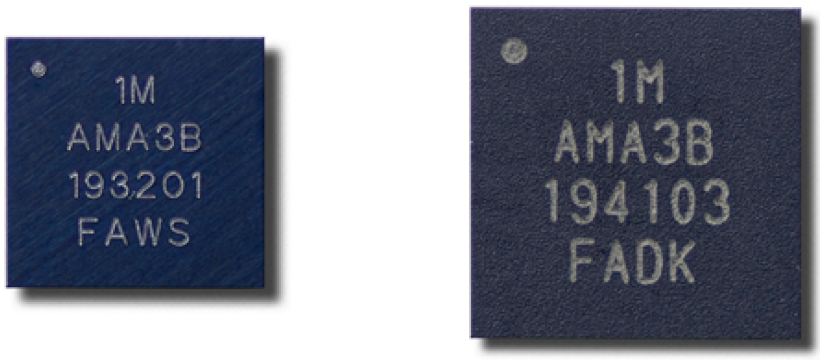
In battery-powered products, where the form factors and battery life are, primary concerns, the voice command system needs an ultra-low-power processor that can handle substantial audio processing tasks.

> In **battery-powered products**, the voice command system needs a processor that can handle substantial audio processing tasks.

Ambiq's Apollo microprocessor line is the preeminent example of ultra-low-power solutions for always-on voice command systems. Designed using Ambiq's proprietary SPOT platform (Subthreshold Power Optimized Technology), these microcontroller units (MCUs) and systems-on-a-chip (SoCs) can run on less than 1/10th of the current of a conventional audio processor.

Specifically, Apollo products (as shown in Figure 1), focus on ultra-low-power and always-aware voice command processing, which makes them ideally suited for ultra-low-power hearable, wearable, and other mobile applications.

Figure 1: Apollo3 Blue Products



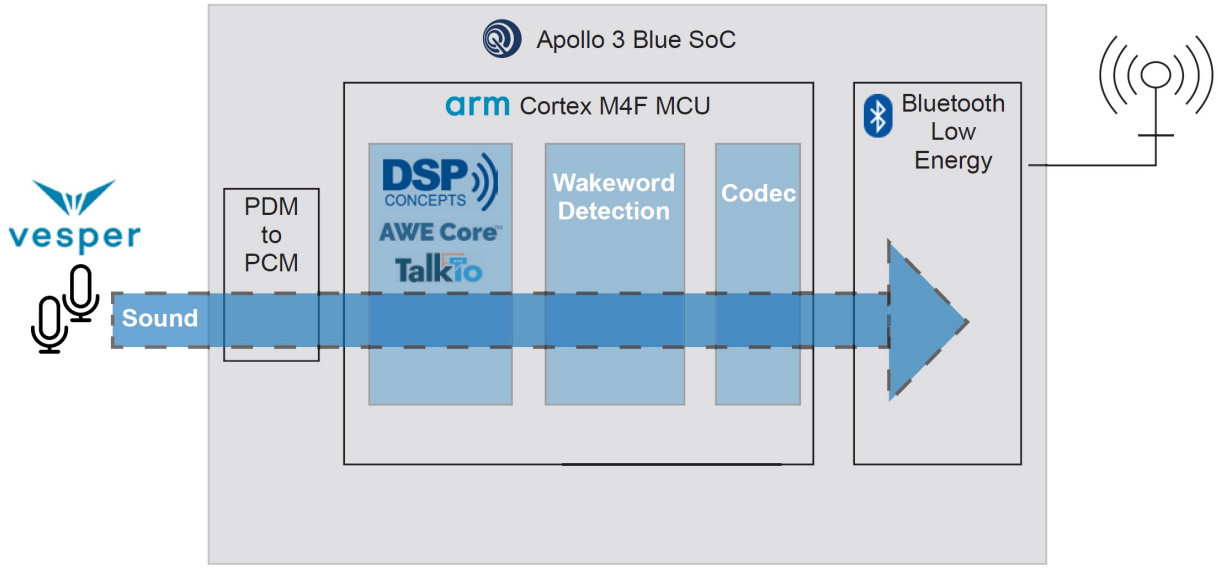
Adaptive ZPL™ by VESPER Technologies

Vesper's microphone delivers an unprecedented adaptive ZeroPower Listening™ engine that can actively monitor sound levels and activate the hibernating audio processor in response to a specific audio event. This powerful feature reduces the power consumption of the entire system by 10X. When combined, SPOT and ZPL technologies enable ultra-low-power always-on solutions for devices powered by small batteries.

Audio Weaver by DSP Concepts

DSP Concepts' Audio Weaver platform enables rapid prototyping of voice-enabled designs on the most advanced embedded processors using a simple drag-and-drop UI. Their TalkTo™ audio front-end algorithms enable reliable voice control in even the noisiest environments of everyday life, supporting applications in consumer electronics, automotive, as well as medical devices.

Figure 2: Ambiq Apollo3 Blue SoC Activated by voice from the Adaptive ZPL Microphones



Applications

While the concept of using always-on voice command is not new, it was not until recently that advancements enabled this feature in portable and battery-powered products. Remote controls, smartwatches, portable smart speakers, connected home devices, and wearable technology are just some of the top applications for always-on devices.

> Remote controls, smartwatches, connected home devices, and wearables are some of the **top applications** for always-on voice command.

Challenges

However, for battery-powered products to offer always-on voice command, manufacturers must overcome severe technological challenges. Because these products are always listening, power consumption must be dramatically lowered.

Typical tech consumers expect at least a full day's use from a single charge, and manufacturers are under constant pressure to improve battery life performance. Some products such as smartwatches are expected to last for weeks, and others like voice remote controls must last a full year at a minimum.

Manufacturers may need to compromise on form factors to accommodate oversized batteries while also accounting for users who have unreliable Internet connections.

Hardware Considerations

The core hardware of a voice command system consists of one or more microphones (a microphone array)—operating in tandem—and a processor that receives audio signals from the microphones. Size, cost, reliability, and power are primary factors when selecting a microphone. For the audio processor, it must have enough computing power to process the signals from the microphone array, and also run all the algorithms needed for voice recognition.

> **Size, cost, reliability, and power** are primary factors when selecting a microphone, while the **computing power** is vital for audio processors.

Beyond the microphone array and audio processor, an always-on voice command product requires additional components, which can vary depending on the product, and the intended application.

However, nearly all voice command products need a wireless interface that can send, and receive data from cloud servers (by accessing the Internet) to offer more than basic capabilities. These products typically have audio feedback as well, which may confirm the user's command through alert tones or voice synthesis. A product's physical design can also have a massive impact on the performance of its voice recognition systems.

Software Algorithms

There are many different algorithms at work in always-on voice command products. The algorithms must listen for the wake word 24/7/365, isolate the user's voice from the surrounding noise when a voice is detected, and then produce a clean signal for the wake word detection engine to recognize the wake word reliably.

Voice recognition algorithms must listen 24/7/365 for the wake word and be able to **reliably recognize** the wake word.

Most voice command products incorporate some form of user feedback to confirm that the device is active, that it heard and understood the user's command correctly, and that it carried out the desired action. To better minimize acoustical differences, the microphones must be installed in the same configuration to the greatest extent possible.

In voice command products, the algorithm package needs to contain essential components, such as the sound detector, noise reduction and filtering, voice direction arrival detection, beam forming, acoustic echo canceling (AEC), adaptive interference canceller (AIC), wake-word detection, and local command set recognition. Most voice command products might also incorporate playback processing to accommodate user feedback, confirm the device is active, and check if the desired action for user' intent is properly carried out. Since these algorithms have complex functions, some of them require tuning for optimum voice recognition accuracy.

Introduction

Over the past decade, always-on voice command systems have proven their appeal and reliability in tens of millions of smart speakers. However, it was not until recently that always-on voice command systems could be effectively implemented into portable and battery-powered products. Thanks to the advancement of ultra-low-power solutions, today's manufacturers of consumer products can now offer a satisfying experience with always-on voice command systems on the go.

This paper will examine always-on voice command and its applications in portable and battery-powered products. It will also discuss the potential challenges that can arise when attempting to implement always-on voice command systems in portable products. Finally, this white paper will describe the hardware and software algorithm considerations that allow always-on voice command to be just as effective in portable, and mobile products as they are in AC wall outlet powered devices designed for home use.

What is an Always-On Voice System?

In typical voice command systems, the user must push a button (so-called Push-to-Talk) to "wake up" the system instead of using a wake word. However, in an always-on voice command system, the system is always on and constantly listening for a specific wake word, or trigger words such as "Alexa" or "OK Google." Current "smart speakers," such as the Amazon Echo, Apple HomePod, or Google Home, work this way.

> In an always-on voice command system, the system is always listening for a **specific wake word or trigger word**.

Today, always-on systems are mostly found in devices designed for home use. AC wall outlets power these systems due to the high power requirements. However, the development of ultra low-power solutions has made it possible for portable products to make use of always-on voice command without quickly draining their batteries.

What is the Lowest-Power Solution for Portable Products?

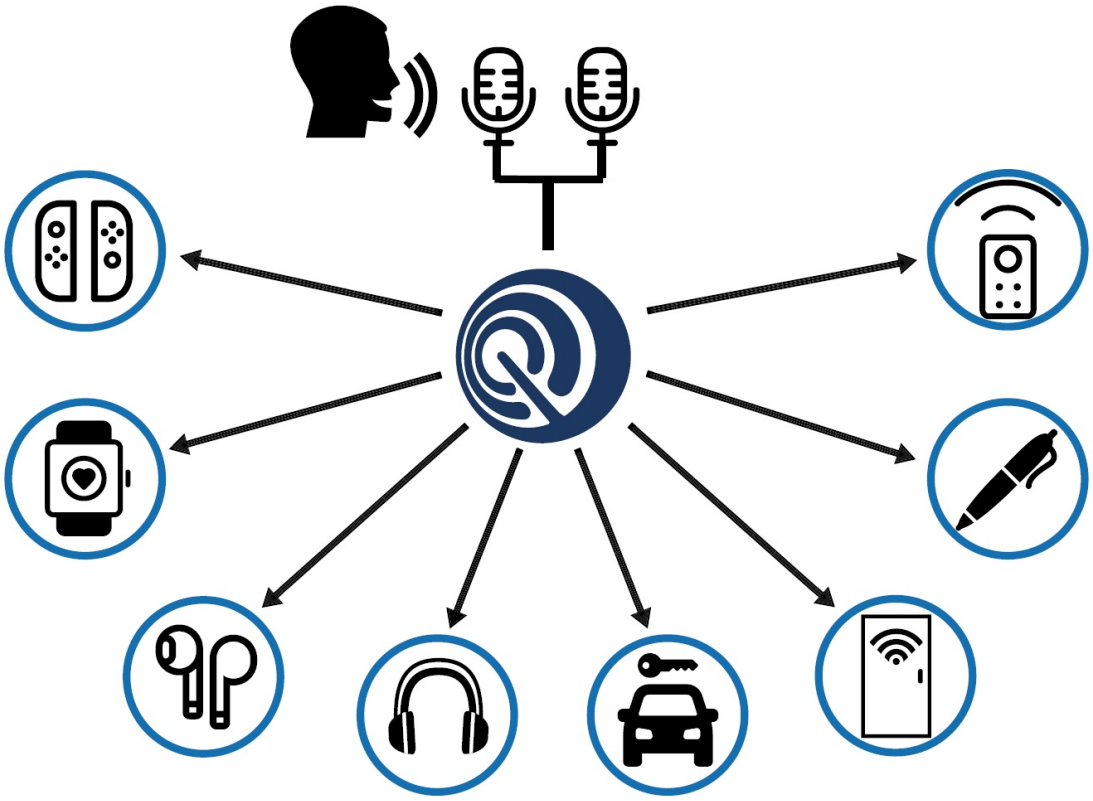
Powered by the SPOT (Subthreshold Power Optimized Technology) Platform, Ambiq has been able to help electronic manufacturers extend battery life and add new features in small form factor, battery-powered products, such as smartwatches, smartcards, and IoT sensors.

Ambiq's Voice-on-SPOT™, or VOS, is the lowest-power solution for always-on voice assistant integration and command recognition in hearables, wearables, and other portable battery-powered products. VOS enables extremely low power audio signal processing to deliver high-quality audio to voice assistant, and on-device Voice User Interfaces (VUI) with minimal battery

life impact. Companies can then deploy high-quality wake word and command phrase recognition everywhere in everyday life.

Voice-on-SPOT (VOS) enables companies to deploy high-quality wake word and command phrase recognition everywhere.

Figure 3: Deploying High-Quality Wake Word and Command Phrase Recognition



Top Applications for Always-on Voice Command in Portable Products

Because the concept of using an always-on voice command in portable and battery-powered products is new, the applications for this technology are only starting to emerge.

Some possibilities include:

Remote Controls

Most current voice-command remote controls require the user to push a button to wake the system before speaking a command. Many require the user to hold the remote close to their mouth in order for the sound to be captured. An always-on system would allow the user to access the remote controller's functions when it is out of reach (or even misplaced). This new generation of remotes can bring "Netflix and chill" to a new level of comfort.

Automotive

In vehicles equipped with always-on voice command, the driver does not have to take their hands off the steering wheel, and fumble around for the wake button used in most voice command systems. It makes it safer for vehicle operation, and makes it easier to control the car's features, such as the GPS, environmental controls, entertainment systems, and remote operations like opening the rear hatch. In addition, other passengers are also able to access the voice command system.

Hearables

Always-on voice command allows users to start or stop playback of audio programs, select material, skip or repeat music tracks, answer phones, or access personal assistant features. All of these tasks are made easy with a paired headphone, or headset (via Bluetooth®) with voice commands to an Internet connected smartphone.

> **Always-on voice command** allows users to start, stop, or skip music tracks, answer phones, or access personal assistant features.

Because no button push is required, the user's hands remain free for other tasks, making a voice-command headphone ideal for sports or fitness, and office work, or as an in-ear personal assistant.

Smart Home Devices

By equipping an on-wall control panel with always-on voice command, the user can control home systems such as HVAC, lighting, and security from anywhere within the voice range. The user does not need to physically access the panel, or use a smartphone to call up the necessary control app, so there is no need for an expensive touchscreen. Furthermore, always-on voice command is usable for other smart home products, including auto-opening trashcans, or voice-controlled drapes and shades.

Options are now, or will soon be available on audio detectors for glass breaking or a baby crying, or that washing machine bearing that's wearing out. New functions and devices are easy to deploy, so there is no need to hire an electrician to run dedicated power lines.

Wearables

Always-on voice command systems can benefit many types of wearables. In a fitness tracker or collar-mounted device, always-on voice command allows the user to control the device while running, walking, or working out, without having to reach for the controls. Voice command is especially practical in smaller products, which may not have room for a sizable display and control buttons.

Voice command is especially practical in smaller products, which may not have room for a **sizable display** and **control buttons**.

A small wearable device can serve as a clip-on personal assistant, or as an interface with a smart speaker, or other devices that are out of voice range. Data can be communicated via Bluetooth to a paired phone, or through Wi-Fi to a local network. Emerging AR applications can use voice as a seamless user interface mode. When exercising, the user can leave their smartphone behind, but not their smartphone experience. Users can stream music to their earbuds, vocally start and stop their exercise timers, or stream music from their watch to their earbuds.

Challenges for Always-on Voice Command in Portable Products

Currently, some considerable challenges stand in the way of enabling voice command in many portable and battery-powered products. These challenges include:

High Power Consumption

An always-on voice command system must be active at all times to pick up user commands at any time. For smart speakers plugged into AC power, this is no problem. However, in battery-powered products, this is a more significant challenge, especially when battery run time is a primary concern for tech consumers.

Reducing power consumption is a challenge for engineers as well, since they must often minimize battery size to maintain a compact form factor. Specifically, the processor tasked with recognizing the wake word must be ultra-responsive, with at least one microphone that must always be active.

> In an always-on system, the **processor** is tasked with recognizing the wake word with at least one **microphone** that must always be active.

In larger systems, special-purpose components can isolate some of these functions, which allow most of the device's other components to be powered down when the device is idling. On the other hand, smaller portable products tend to rely on a system-on-a-chip (SoC), in which a single component performs almost all of the device's functions. In these products, there may be a few, or no inactive components that can be shut down.

Battery Life Expectations

Most tech consumers expect to get at least a full day's use (8 hours) from a product without recharging or replacing its batteries. Today, most active headphones and earphones can run for 18 to 20 hours, while inexpensive models can usually manage 10 hours. Some of the latest wearables, such as "true wireless" earphones and clip-on wireless speakers, have battery run times in the range of 5 hours. Regardless, manufacturers are under tremendous pressure from consumers and reviewers to improve battery life performance, especially in voice command products that serve as control interfaces.

Currently, consumers expect the batteries in remote controls to last at least 10 to 12 months. When new features like Always-on-Voice Command are added, this lifetime needs to be maintained even as many remotes migrate from AA, to AAA bat-

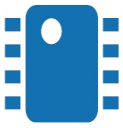
teries, which have less than half the milliamp-hours (mAh). On-wall control panels for smart home systems tend to run for about a year (or even two) on a set of AA or AAA batteries. It's impractical to use rechargeable batteries in an on-wall control panel, and unrealistic to expect consumers to change these batteries frequently.

Unreliable Internet Connection

Home products can rely on a nearly always-present Internet connection to offload most voice recognition processing to external servers. However, portable products often need to be paired to a smartphone, through Bluetooth Low Energy, to access the Internet, since cellular data connections are unreliable, or even impossible in many locations. Mission-critical applications, such as HVAC and security systems, need to function even during occasional Internet outages.

Because of unreliable Internet connections in portable applications, portable products using voice command must recognize, and process a small vocabulary of voice commands on their own, without help from external servers. Not only does this require more powerful processing, but it also limits the controllable functions of the voice command.

> Portable products using voice command must recognize, and process a small vocabulary of voice commands **on their own**.



Form Factor Compromises

The compact size of most portable products may require a fewer number of microphones used in an array. It may also force engineers to position microphones in a way that compromises their performance, and make the precise matching of response, and sensitivity of multiple microphones difficult or impossible.

> The form factor of wearables may force product designers to use **fewer microphones** and choose **smaller batteries**.

The form factor of wearables, and other compact portable products, also forces product designers to choose smaller batteries, which offer less power. For example, a typical AA alkaline battery might offer 3000 mAH of power. In contrast, a CR2032 lithium "coin cell" of the type used in many tiny tech products, offers only 220 mAH of power. Therefore, a product drawing 10 mA (or 10,000 µA) can run for 22 hours when powered by a CR2032.

Environmental Factors

Portable products are more exposed to challenging environments than home products. Wearable products must be at least sweatproof, which requires an ingress protection rating of IPx5, while portable products intended for the rugged outdoors, should be fully immersible (a rating of IPx7). Consumers expect their portable products will survive an occasional encounter with a rainstorm, or a trip through the washing machine.

However, the seals required to achieve these ratings may impair the function of microphones, and place limitations on the configuration of microphone arrays. Also, voice-recognition microphones in portable products, need to withstand the potential shock of falling from at least waist height, onto hard surfaces such as concrete.

Hardware Considerations for Voice Command in Portable Products

The core hardware for a voice command interface consists of a microphone array, and a processor that can receive, and interpret the audio signals from the microphones. Depending on the type of device, various other components may be needed, such as a wireless interface for Bluetooth Low Energy or Wi-Fi, plus speakers, amplifiers, LEDs, and displays to provide user feedback.

Microphone Array Design

Although it is possible to use a single microphone in a voice command product, most such products use a beamforming array of two to seven microphones to isolate the speaker from ambient noise better. The array allows the audio processor to focus the pickup pattern of the microphones on the user's voice, which improves the signal-to-noise ratio of the user's voice relative to the surrounding environmental noise. However, the demands placed by the form factors of portable and battery-powered products present many challenges not found in products designed to be plugged in.

Figure 4: Typical Microphone Pattern Used in Smart Speaker Designs

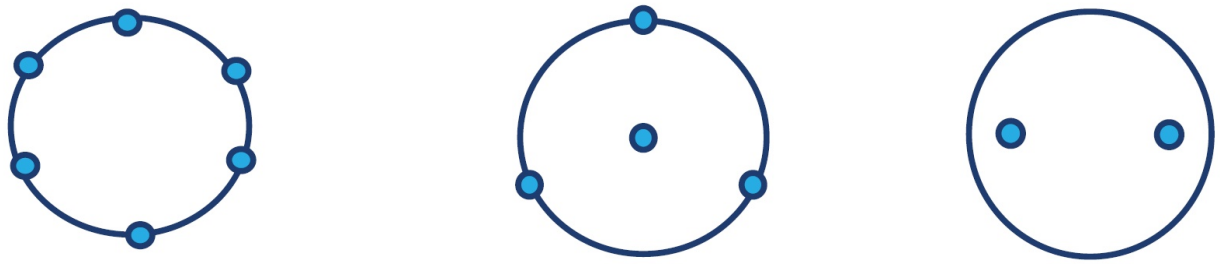


Figure 4 shows the typical mic pattern used in smart speaker designs: 6-mic (high end), 3+1 mic (standard) and 2-mic (low-cost) devices. Mic array diameter can range from 40-75mm. Portable smart speaker designs use 4-mic or 2-mic configurations with the mic array located on top of the product.

Figure 5: Typical Microphone Pattern Used in TV and Home Appliances

Figure 5 shows the typical mic pattern used in TV and home appliances: 2 or 4-mic linear array design. Mic spacing can be 25-60 mm for 2-mic and 25mm spacing between each microphone pair in 4-mic. Mics are facing up or front.

The DSP Concepts white paper, "Designing Optimized Microphone Beamformers," found that achieving the best possible signal-to-noise ratio is critical to the accuracy, and reliability of a voice command product.

The white paper also found that using microphones with tighter sensitivity tolerances could help performance by using microphones with ±1 dB tolerance, rather than the more typical ±3 dB. Since each microphone in an array may be in a different acoustical environment—due to the product's physical design—it is better to match the processor's microphone sensitivity rather than in the microphones themselves. It is also essential that the sensitivity of the microphone remains stable over the lifetime of the product.

Number of Microphones

DSP Concepts' research demonstrated that increasing the number of microphones improves voice user interface (UI) reliability. The more closely matched the sensitivity of the microphones are, the better the performance of the beamformer.

> The more closely matched the sensitivity of the microphones are, the better the **performance** of the beamformer.

The most practical way to achieve this is to balance the microphone sensitivity in the hardware after the microphones are installed. This way, the sensitivity adjustment compensates not only for the differing gain of the mics—typically specified to a precision of ±3 dB—but also for the acoustical effects of the enclosure on the microphones.

However, few portable products and almost no wearables have the space for such an array. True wireless earphones, for example, typically have room for only two mics in each earpiece with available microphone spacing of only 10 to 20 mm between the microphone pair. Also, the processing power required for such an array may be beyond the capabilities of the relatively small processors used in most portable devices. Therefore, software algorithms that perform beamforming and other voice UI optimization functions, must have the capability of being optimized for two or at most three microphones.

Microphone Selection

Because voice command products use multiple microphones, the primary factors in microphone selection for these products are usually size, cost, and quality. However, in portable and battery-powered products, a lower system power consumption is essential.

> The primary factors in microphone selection for voice command products are usually **size, cost, quality**, and **power consumption**.

Vesper MEMS piezoelectric microphones are already available on the market in a small footprint with Adaptive ZPL and a very competitive cost to address portable, and battery powered products. Another benefit of the Vesper piezoelectric MEMS microphone is that they are exceptionally stable, as the sensitivity does not shift during solder reflow, humidity, or temperature changes. All Vesper microphones are native IP57 with 200usec start-up time.

Real-World Products: Vesper VM3011 is the first MEMS digital microphone featuring the 2nd generation of Vesper ZPL (Adaptive), drawing only 10uA of power when in "Wake on Sound" mode. Considering that batteries in portable products typically dissipate about 50 µA of power even when fully powered off, the VM3011 has virtually zero effect on the battery life of a portable product. Most importantly, these ZPL microphones allow the rest of the system to hibernate in very-low-power modes, while the ZPL microphone monitors the environment and wakes up the processor when certain sounds occur.

Below is a study conducted by Vesper collecting (24 hrs) data from different households during weekday and weekend through Vesper dataloggers. Each datalogger had 4 of VM3011 with a different WoS threshold setting. The result as shown in Figure 6, shows that VM3011 allows Always-On with system hibernation between 81% and 92% of the time.

Figure 6: Vesper VM3011 Allows Always-On with System Hibernation

Datalogger – Weekday

User	User Profile	Datalogger Placement
User 1,2	Young professional who is the only one in the household	Living Room Coffee Table
User 3	Family of 3 with 2 adults and 1 child	Living Room Coffee Table
User 4	Family of 4 with 2 adults and 2 children	Living Room Coffee Table

	6dB	9.5dB	14dB	16.9dB
Normal Mode	13%	11%	7%	5%

Datalogger – Weekend

User	User Profile	Datalogger Placement
User 1	Family of 5 with 2 adults, 3 children and 2 pets (dog and cat)	Living Room Coffee Table
User 2	Family of 4 with 2 adults and 2 children	Living Room Coffee Table
User 3	Family of 4 with 2 adults, 2 children and 2 pets (dog and cat)	Living Room Coffee Table

	6dB	9.5dB	14dB	16.9dB
Normal Mode	35%	31%	21%	15%

(5 x Weekday + 2 x Weekend) / 7

	6dB	9.5dB	14dB	16.9dB
Normal mode	19%	17%	11%	8%
ZPL Mode	81%	83%	89%	92%

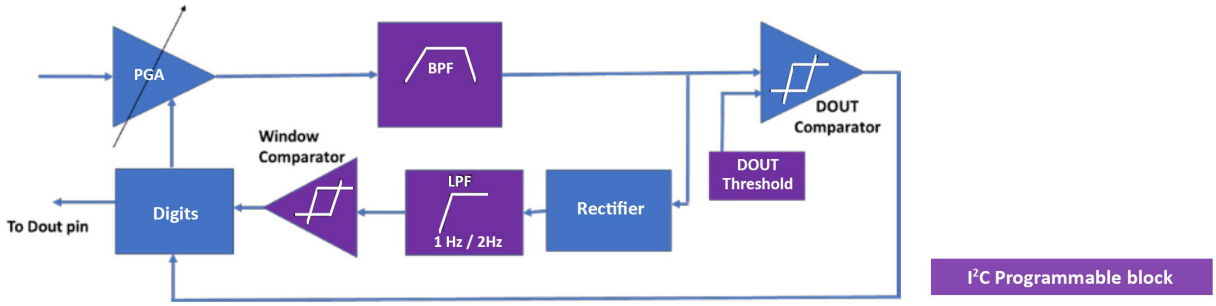
A piezoelectric microphone element in the VM3011, is monitored by a very low-power analog circuit, which monitors and tracks the ambient sound level to activate the system only when a sound is detected above the background noise. A WoS (Wake On Sound) threshold allows optimizing the VM3011 for the best performance in a variety of noise environments. A Passband-filter helps to narrow the selection of specific sounds of interest, between 200 Hz and 8 KHz, to pick up human voices better and reject environmental noises such as machinery rumbles, and wind noise. This circuit is known as "Adaptive ZPL."

The adaptive ZPL offers multiple programmable blocks via I^{2}C:

1. Pass-band-filter to narrow the selection of specific sounds.
2. WoS Threshold 2x-8x above RMS level.
3. Low-pass-filter to program the update rate of the RMS background noise.
4. Window-Comparator for fast vs. slow-moving background noise adaption.

Figure 7 is a simplification on the ZPL:

Figure 7: ZPL Multiple Programmable Blocks via I^{2}C



A single piezoelectric microphone can trigger the microphone array, audio processing circuitry, and Internet connection (if applicable) of a voice command product.

Audio Processor Considerations

In any voice-command product, the audio processor—whether a dedicated Digital Signal Processor (DSP) or a processing core within an SoC—must have the necessary computational capability to process the signals from all of the microphones in an array, and to run all of the algorithms necessary for voice recognition.

> The **audio processor** must be able to process the signals from all the microphones, and run all the voice-recognition algorithms.

The more advanced algorithms and the more microphones the chip can accommodate, the better the signal-to-noise ratio, and the more accurate the voice recognition is. In portable and battery-powered products, however, the processor must also consume as little power as possible to maintain adequate battery life in the product.

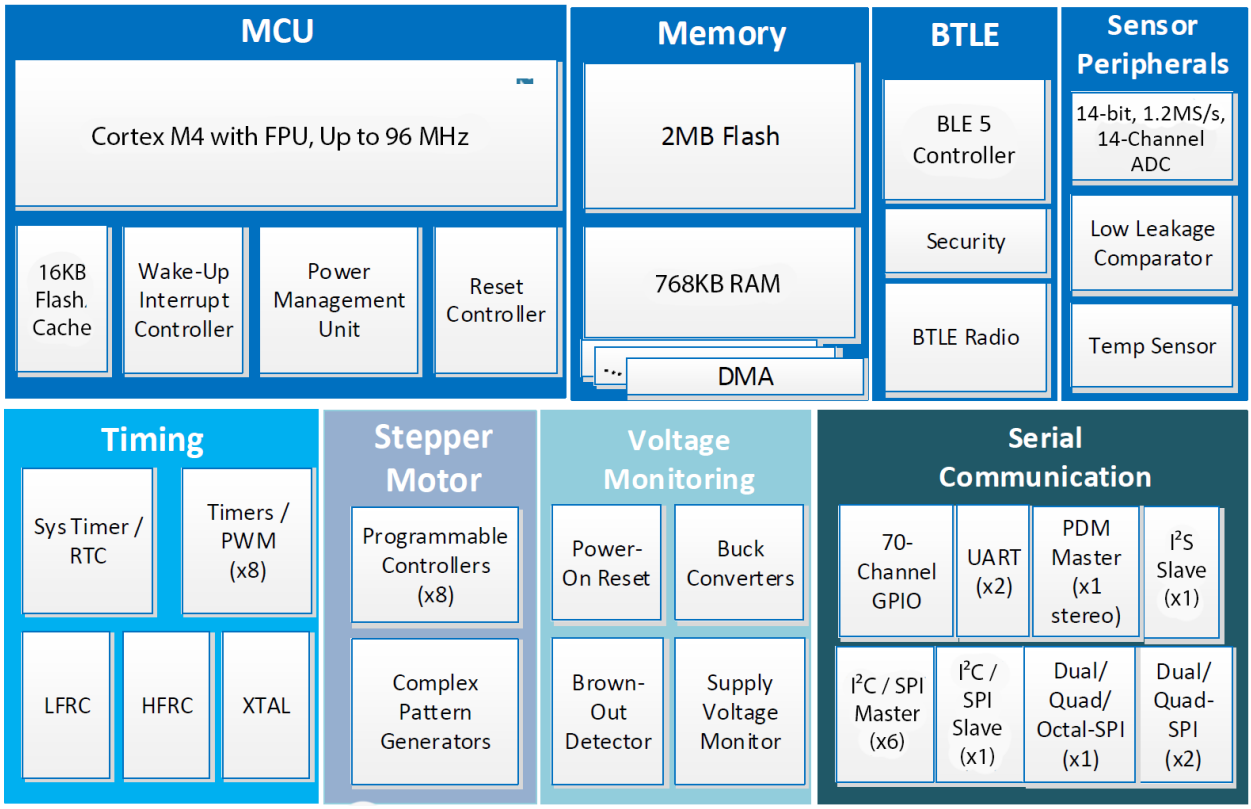
Real-World Products: One processor line, explicitly designed to handle substantial audio processing tasks in battery-powered products with small form factors, and battery power is Ambiq's Apollo line. These microcontroller units (MCUs) and systems-on-a-chip (SoCs) are designed using Ambiq's SPOT (Subthreshold Power Optimized Technology) platform, which allows them to run on less than 1/10th of the current of a typical audio processor.

Ambiq's SPOT-based Apollo2 is a 48 MHz Arm® Cortex® M4F-based MCU focused on sensor and voice processing that consumes only 10uA/MHz. Apollo2 Blue is available with a Bluetooth Low Energy channel for voice assistants.

Apollo3 Blue further lowers power to 6uA/MHz and increases frequency to 96 MHz. Its Bluetooth Low Energy radio is 5.0 compliant.

The Apollo2, Apollo2 Blue, Apollo3 Blue, and Apollo3 Blue Plus processors are capable of handling signals from multi-microphone arrays using DSP Concept's Voice UI algorithms, making them appropriate for ultra-low-power hearable, wearable, remote control, and other mobile applications.

Figure 8: Structure and Feature Diagram of the Ambiq Apollo3 Blue Plus



Ambiq will extend its ultra-low-power industry leadership with upcoming processors that will both dramatically lower uA/MHz, and increase Fmax. The new processors will achieve tighter beamforming patterns, better signal-to-noise ratios, and better voice recognition accuracy.

All of these processors have the compact size needed for products such as bands, smartwatches, and earbuds, and they measure from 2.5mm, to 5.3mm square depending on the package.

Additional Components

Beyond the microphone array and audio processor, a voice command product requires additional components. Specific component requirements depend on the application and form factor, but there are a few that almost every voice command product uses. As with the microphones and processors, these components must be chosen not only for their functions and performance, but also for their small size and low power consumption.

Wireless Interface

Voice command products need to send and receive data from external servers (by accessing the Internet) to offer additional capabilities. Smart speakers designed for home use connect through Wi-Fi to a LAN. With portable voice-command products, the connection occurs through Bluetooth or BLE (Bluetooth Low Energy) to a smartphone or tablet, which then connects to the Internet through a cellular data network or Wi-Fi.

> Voice command products need to **send and receive data from external servers** to offer more than the most basic capabilities.

User Feedback Components

Most voice command products incorporate some form of user feedback to: confirm that the device is active, heard and understood the user's command correctly, and that it carried out the desired action. These devices can be LEDs, such as the flashing lights atop the Amazon Echo and Google Home smart speakers. They can also be alphanumeric or graphical displays, which may be found on many remotes, and home automation wall panels.

These devices typically have audio feedback as well, which may confirm the user's command through alert tones or voice synthesis—yet another load placed on the processor. The unit must employ an amplifier and a speaker of some sort to reproduce the voice, and sometimes alert tones. Some products may even use multiple drivers with a beamforming algorithm to direct the response back at the listener.

Industrial Design Considerations

The physical design of a product can have a massive impact on the performance of its voice recognition systems. According to DSP Concepts' research, precise matching of microphone sensitivity is essential for reliable beamformer performance, and accurate voice recognition.

Manufacturing Consistency

If the microphones in an array are different distances from the edge of a product, for example, they will have a different frequency response and, therefore, differing sensitivity at different frequencies. To better minimize acoustical differences, the microphones should be installed in the same fashion to the greatest extent possible. Any seals around the microphones must also be consistent in design, materials, and installation.

Manufacturers should design voice command products carefully so that every microphone in the arrays is in a similar acoustical environment. Manufacturers can also ensure the best possible performance of voice command products by level-matching the microphones of each unit individually at the factory. This extra QC step ensures that any differences in microphone performance from minor manufacturing inconsistencies, do not affect the accuracy of voice recognition.

Manufacturers should design voice command products carefully so that every microphone in the arrays is in a **similar acoustical environment**.

The successful design of the Voice UI solution depends on a stable microphone array and playback signal chain, as well as optimized MIPS, and memory usage. DSP Concepts Audio Weaver software comes with a RTASC (Real-Time Audio System Check) module, to iteratively verify the hardware and software functionality throughout the design cycle. With easy-to-use debug scripts, RTASC enables a risk-free, and agile product development process enabling voice-enabled products a faster time to market.

Software Algorithms for Voice Command in Portable Products

There are many different algorithms at work in always-on voice command products, all of which must be tuned to suit the product's design and application. These algorithms must listen for the wake word 24/7/365, isolate the user's voice from the surrounding noise when a voice is detected, and then produce a clean signal for the wake-word detection engine to recognize the wake word reliably.

Basic Algorithm Structure

Here are the basic components of a voice command algorithm package, presented in order from the microphone end, to the final signal output:

Sound Detector

Typically, a Vesper Adaptive ZPL monitors the signal from a single microphone. When the signal level exceeds a certain threshold—such as when a user speaks the wake word—the microphone sends a signal to power up the rest of the system, it is critical in portable products because it allows the shutdown of other components to save power.

> When a microphone's signal level **exceeds a certain threshold**, a comparator sends a command to boot up the rest of the system.

This function must also occur quickly so that the system can receive the wake word. For example, with the Vesper microphone, the microphone wakes up within 200 µs, much less than the time it takes to utter the first letter in any keyword. Therefore, no audio buffer is needed.

Noise Reduction and Filtering

To implement voice detection in noisy environments, such as a typical household, Vesper has designed an innovative IP called "Adaptive ZPL." The user can program this ultra-low-power analog circuit through I^2C. Adaptive ZPL configuration can be changed on the fly, and it is easy to use, therefore integrating with any application processors with available PDM and I^2C interfaces.

The adaptive ZPL circuit filters out unwanted sounds or noise outside of the user-programmed audio band filter. Meanwhile, it latches a DOUT pin only when a sound is detected above the user-programmed WoS threshold. In the Adaptive ZPL, the WoS threshold continuously traces and follows the RMS background noise with a refresh rate of 0.5 sec/1 sec (also user programmable), which drastically reduces false rejects.

Direction of Arrival Detection

Because voice command products use multiple microphones, the primary factors in microphone selection for these products are usually size, cost, and quality. However, in portable and battery-powered products, lowering the system power consumption becomes essential.

> For a microphone array to focus on a user's voice, it must first **determine where the user** is relative to the product.

The microphone array must also include precedence logic that rejects reflections of the user's voice from nearby objects. It must also adjust its operating threshold to compensate for ambient noise level, so environmental noise does not create false directional cues. Determining the direction of arrival may not be necessary for products such as earphones, in which the physical position of the user's mouth relative to the microphone array is already known.

Beamforming

A microphone array can process the signals from multiple microphones so that the array becomes directional. It accepts sounds coming from the determined direction of arrival while rejecting sounds coming from other directions.

With some products, such as earphones and automotive audio systems, the direction of the user's voice relative to the microphone array is known, so the beamformer's direction may be permanently fixed. In devices such as smart speakers, remote controls, and home automation wall panels, the beamformer's desired direction of focus has to be determined, and the response of the array adjusts to focus in the direction of the user.

Acoustic Echo Canceling (AEC)

Acoustic echo canceling rejects the sounds (such as music or announcements) coming from the device itself so that the microphone array can pick up the user's voice more clearly. Because the original signal and the response of the device's internal speaker are known, the device knows to reject the signal that comes back through the microphone.

> AEC rejects the **sounds coming from the device itself** so that the microphone array can pick up the user's voice more clearly.

Selecting microphones with a high overload point, and minimizing the speaker distortion in the playback path, is crucial to achieving excellent AEC performance. This, in turn, results in a better music barge-in performance, particularly when playing low-frequency content at loud playback. DSP Concepts stereo AEC algorithms cancel out 35 dB of echo during music barge, which results in high wake word detection accuracies and improved user experience.

AEC is not necessary for products such as headphones and earphones because the sound coming from the product's speakers is confined, and typically not enough of it leaks out to affect the performance of the product's microphones.

Adaptive Interference Canceler (AIC)

Adaptive Interference Canceller rejects the interfering sounds, such as a TV playing in the living room, or microwave noise in the kitchen that are hard to cancel out with a traditional beamformer described above. Unlike other adaptive cancellation techniques, DSP Concepts' AIC algorithm does not require a reference signal to cancel out the interfering noises. Instead, it uses a combination of beamforming, adaptive signal processing, and machine learning to cancel out 30 dB of interference noise, while also preserving the desired speech signal.

AIC is necessary for products, such as remote controls, and smart speakers, that are typically operated in living room environments, where there are interfering noises and moderate to high reverb conditions. For example, when a user is operating a TV remote control, AIC will be able to cancel out the TV sound from the audio stream, to present the speech signal to the wake-word detection engine as if there was no interfering noise present.

Wake-Word Detection

Once the system detects sound and powers up, it must record the incoming audio and compare it to a stored digital file of the wake word (such as "Alexa" for the Amazon Echo). If the waveform of the incoming audio is sufficiently close to the stored file, the device becomes receptive to voice commands.

> If the waveform of the incoming audio is sufficiently close to the stored file, the device becomes **receptive to voice commands**.

In contrast to portable products, smart home speakers only need to recognize its wake word, as they offload other voice recognition tasks to an external, Internet-connected server. The wake-word detection model typically runs locally on the device, while some service providers such as Amazon, can enable additional wake word checks in the cloud.

Local Command Set Recognition

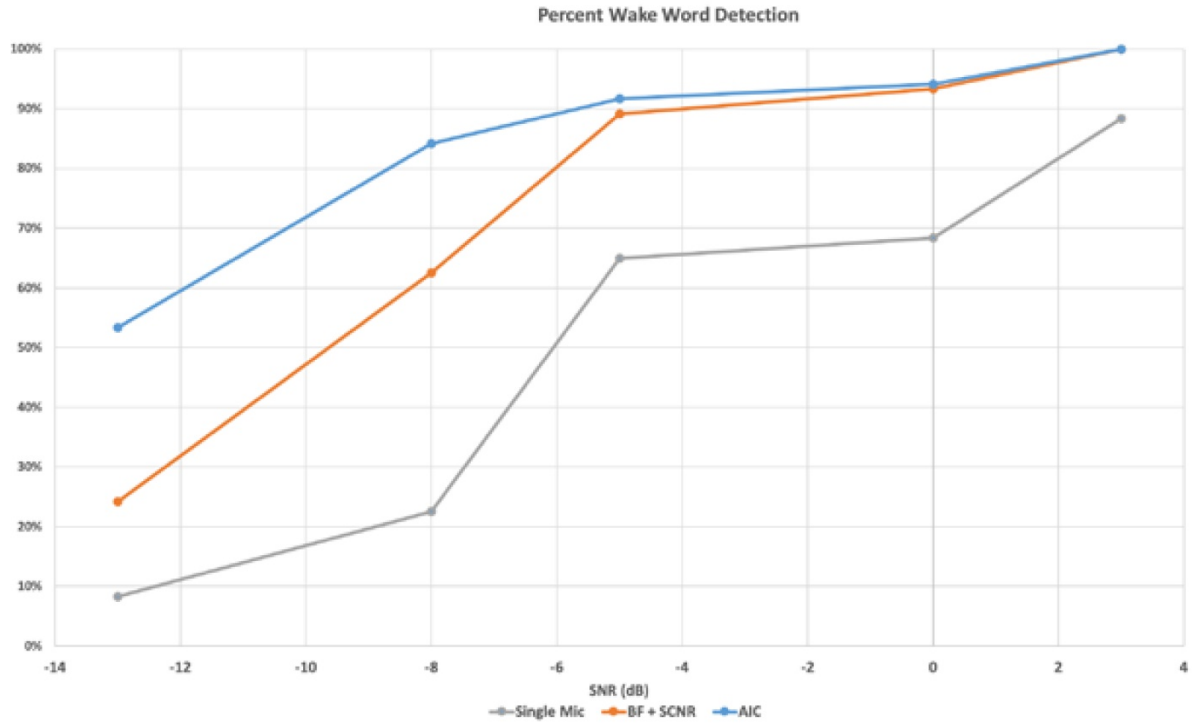
Because portable products cannot rely on an Internet connection as today's smart speakers do, they need to recognize a certain number of basic function commands on their own without the help of external servers.

> Portable products need to recognize a certain number of basic function commands **without the help of external servers**.

These commands are typically limited to basic functions such as play, pause, skip tracks, repeat, and answer calls. Recognition of these commands works in the same way as wake-word detection does. However, even though the command set is limited, the need for a local command set recognition increases the load on the processor compared with a smart home speaker. Portable devices, such as wearable headsets, can communicate over a Bluetooth or WiFi link with a mobile phone, which then performs the command processing.

Real-World Products: Figure 9 shows the percentage wake-word detection rate for a remote-control reference design using DSP Concepts' TalkTo™ algorithms running on Ambiq Apollo3 MCU. Percent wake-word detection is the measure of how many wake words the design accurately identifies. During tests, the user is located one meter away from the remote, and a music file is played on the TV two meters from the remote. The TV volume is varied from 62-78 dB sound levels, while the speech is played at 65 dB at the microphone array to achieve the SNR values shown in the Figure 9.

Figure 9: Performance Comparison of TalkTo Software Configurations for Remote Control



Results show that the AIC works best in the most challenging noise conditions where there is significantly low SNR. The 1-mic algorithm requires at least 3 dB SNR to achieve a wake word detection rate greater than 80%. The 2-mic beamformer together with Single channel noise reduction (SCNR) algorithm performs the same as AIC at 0 dB SNR.

However, as the SNR worsens, AIC offers significantly better performance with a 10% improvement at -6 dB and 20% at -8 dB. In applications such as remote controls, where there is no way to use a reference signal to cancel out the TV noise, AIC algorithms robustness to non-stationary noises and reverberation is crucial to achieving reliable always listening operation.

Algorithm Tuning

The function of each of the above algorithms is complex. It must be adjusted to suit the application—especially in portable products, where the environment and use patterns are likely to be different from those of home products. Here are the algorithm functions that need tuning for optimum voice recognition accuracy:

Detection/Wake Threshold

The threshold levels for sound detection and wake-word detection must be set high enough to minimize false triggering of the device, but low enough that the user can address the device at an average speaking level. These wakeup thresholds also depend on the use case.

For example, a remote control that is 2-3 feet from the user should be set to a lower threshold, whereas a wearable device has to be set to a higher threshold to reduce false positives. In portable products, especially, it may be desirable for these levels to be adjusted dynamically so that the performance adjusts to compensate for varying levels of ambient sound. The function of the dynamic compensation will itself have to be tuned.

> The wake threshold must **minimize false device triggering** but still allow users to address the device at an average speaking level.

Noise Reduction/Canceling

Devices can be tuned to reject different types of noises depending on their application. For example, manufacturers know the spectrum of any given car's road and engine noise at different speeds, so the voice recognition system can be tuned to reject these sounds. The noise reduction or canceling algorithms can also function dynamically by adapting to the changing environment. However, this dynamic function also requires tuning.

> Devices with always-on voice command can be **tuned to reject different types of noises** depending on their application.

Beamformer Beamwidth

The tighter the beamwidth of the beamformer, the better it rejects environmental sounds, and reflections of the user's voice from other objects. However, setting the beamwidth too tight causes the unit to reject the user's voice if the user moves slightly.

The **tighter the beamwidth** of the beamformer, the better it rejects environmental sounds and reflections of the user's voice.

In products such as earphones and headphones, where the direction of arrival of the user's voice does not vary, it can set a tight beamwidth. However, in products such as remote controls and home automation panels, the beamwidth must be set wider to accommodate the movement of the user while the user is speaking.

Wake/Sleep Strategies

A key goal when minimizing power consumption is to put the device to sleep as often as possible, and to keep it asleep for as long as possible. However, this goal requires trade-offs. If the device is put to sleep too quickly after use, it may miss commands that follow the wake word, and require the user to speak the wake word again, which usually leads to frustrated users. However, if the device stays awake longer than necessary, it consumes more power than it needs to.

To **minimize power consumption**, the voice command device should sleep as often as possible, and be kept asleep as long as possible.

Reference Design

With the technologies described in the previous sections, it is possible to provide a complete always-on, voice-to-cloud solution for local voice commands and/or voice assistant integration in battery powered products.

Specifically, the Voice-on-SPOT reference design (as shown in Figure 10) combines the Vesper Microphones Wake-on-Sound capability, with the Ambiq Apollo MCU running DSP Concepts' flexible, high performance TalkTo algorithms to provide an industry-leading and ultra-low-power voice experience.

Figure 10: Voice-on SPOT Reference Hearable

A complete always-listening, voice-to-cloud solution for vocal voice commands and/or voice assistant integration in battery-powered products

Conclusion

Because of power consumption demands and form factor limitations, the capabilities of the audio processors used in most portable products are typically less than those of the processors used in home products.

However, ultra-low-power solutions such as Ambiq's Apollo processors and Vesper's ZeroPower Listening microphones, microcontroller units (MCUs), and systems-on-a-chip (SoCs) can run on less than 1/10th the power of a typical audio processor. By focusing on ultra low-power and always-on voice command processing, Apollo products can serve vehicles, remote controls, smart home devices, along with hearable and wearable technology.

The challenges of creating always-on voice command products that can run for many hours to many months on battery power—while achieving functionality similar to that of today's popular smart speakers—are considerable. But thanks to the products described in this paper, these challenges can now be overcome. The proper choice of components, combined with careful tuning to suit the application, can result in portable voice command products that deliver a satisfying and reliable experience for consumers.

Benefits of Ambiq's SPOT Platform

With Ambiq's SPOT (Subthreshold Power Optimized Technology) platform, tech designers and developers will have the ultra-low-power solution needed to create the voice command systems of the future. More specifically, tech designers and developers will have the tools to create an always-on, voice-to-cloud solution for local voice commands, and voice assistant integration in portable and battery-powered products.

Learn more about how Ambiq's SPOT Platform is powering tens of millions of commercial products or check out our expert resources to receive more support.

Authors

Paul Beckmann, PhD — Founder/CTO of DSP Concepts

Dr. Beckmann has extensive experience developing audio products and implementing numerically intensive algorithms. He spent 9 years at Bose Corporation where he was awarded the "Best of What's New" award from Popular Science for contributions made to the Videostage decoding algorithm. Paul was tasked by Dr. Bose to charter Bose Institute with industry courses on digital signal processing and holds a variety of patents in signal processing techniques. He received BS and MS degrees in Electrical Engineering from MIT in 1989 and a Ph.D. in Electrical Engineering in 1992, also from MIT. To find out more about DSP Concepts' powerful software algorithms, go to w.dspconcepts.com.

Aaron Grassian — Vice President of Sales at Ambiq Micro

Aaron is an international business development executive with experience in building, organizing, and managing global teams and channel partners for both public and private technology companies. Throughout his 15-year career, he has had repeated success in driving ultra-low-power solutions through multiple channels, targeting and winning initial designs, and quickly growing revenue through customer engagements and partnerships. He also has global customer experience in building and managing Asia sales for startup companies. Aaron holds a BS in Electrical Engineering from the University of Florida. For more information about Ambiq's industry-leading microprocessor technology, go to ambiqmicro.com.

Matt Crowley — CEO at Vesper

Matt is passionate about building great teams to bring disruptive technologies to market. Under his leadership, Vesper introduced its first product to market nearly five times faster than the industry average. Prior to joining Vesper, he pioneered the mass commercialization of piezoelectric MEMS devices at Sand 9, a fabless Micro-electro-mechanical system company. Matt has also advised Fortune 500 companies on operational and strategic issues. He received an interdisciplinary degree in Physics and the Philosophy of Science from Princeton University. Matt is also fluent in Japanese. To learn more about Versper's energy-efficient microphones, go to vespermems.com.

6500 River Place Boulevard, Building 7, Suite 200, Austin, TX 78730
www.ambiqmicro.com
sales_americas@ambiqmicro.com
+1 (512) 879-2850

A-MCUAP0-WPGA01EN v1.1
June 2020