Technical Article
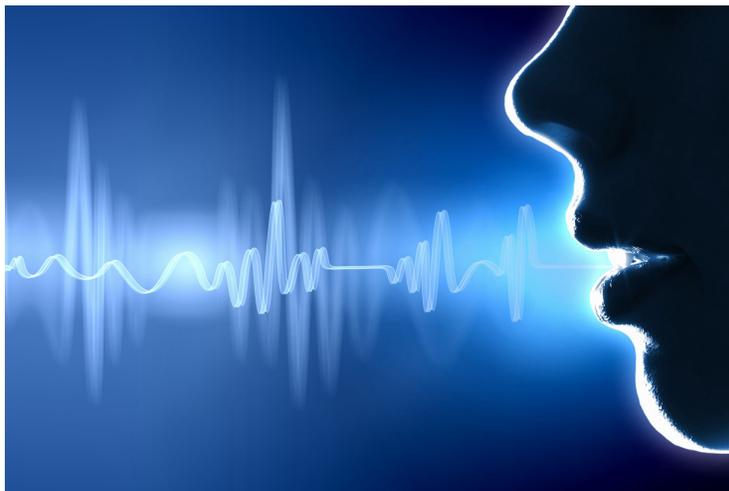# *When to Choose a DSP for Processing Voice Commands*

Debbie Greenstreet

In a twist of irony, the massive technological expansion of the telephone system – which led to the creation of the internet and all of its features and benefits – enables millions of simultaneous transactions without requiring a single human conversation. Yet in this efficiency-focused 21st century world, the plethora of voice-activated smart home devices confirms that live speech has become the de facto medium for smart home virtual assistants. Consumers can still pick up a remote and press the volume button; they can use their phones to place an online order and their hands to flip a light switch. But it seems clear that voice as a user interface is here to stay, and real-time signal processing is the key enabler of this modality.

Real-time signal processing consists of converting and processing analog signals by performing complex math operations. Digital signal processors (DSPs) are the most efficient way to process math in real time. While all processors can perform real-time signal processing math calculations, the DSP architecture by design ensures that these processes will happen faster, with less energy and less generated heat than more generic processor architectures.

Voice as the user interface – a new era in speech processing

 [Read the white paper here](#)

Voice capture and speech recognition devices and applications are not new. However, properly recognizing and capturing speech amid television noise or conversations requires far more processing than simply capturing speech from a single microphone in a device that's close by and in a relatively quiet environment.

Key design factors in a smart speaker-like solution include accurate voice discernment (given the user's distance from the speaker), the amount of ambient noise in the area and the need for two-way speaker conversation. For near-field processing, a relatively simple system of three or fewer microphones, wake word detection, fixed beamforming and signal-noise reduction may be all that's required. It is possible to execute such a configuration easily on a microcontroller (MCU) within a low-power, memory and cost footprint.

However, depending on the application, the consumer may expect to command the smart speaker from both the near-field and far-field, and to have the speaker accurately discern their speech over noises from sources such as TVs, smartphones, background conversations, wind and other ambient sounds. To be effective, this complex environment typically requires between four and eight microphones that in turn require adaptive beamforming

and adaptive spectral noise reduction (ASNR) algorithms, along with multisourcing selection functionality. This significantly increases the real-time signal processing complexity.

Applications such as video doorbells expand the processing complexity one step further, requiring process-intensive acoustic echo cancellation (AEC) to improve the user experience. AEC, along with beamforming and ASNR, tips the scales in terms of exceeding MCU efficiency, but DSPs can effectively process voice as a user interface engine.

DSPs continue to be the most efficient means of processing real-time audio commands, especially amid the ambient sounds and noises commonly present in our environment. Just as in a smart home interface where consumers prefer using their voice because it's the most efficient, DSPs are preferable for far-field, four-or-more microphone, or two-way-speaker smart home solutions where size, power, performance and cost are key metrics.

Additional resources:

- Check out this Audio Preprocessing System Reference Design for Voice-Based Applications
- Learn more about all of TI's audio and media processors.

IMPORTANT NOTICE AND DISCLAIMER