Maximizing Artificial Intelligence for Voice Activation and Control

Tips for Maximizing Lantronix SOMs for Battery Management, Password Security and More

By Victor Gonzalez
Senior Director of Engineering at Lantronix Inc.

The use of Artificial Intelligence (AI) is exploding in a variety of markets, including automotive, healthcare, security and consumer electronics. As a result, the demand for voice-enabled devices has increased significantly. In fact, Fortune Business Insights predicts that the speech and voice recognition market will grow at a CAGR of 19.8 percent from now until 2026 to reach $28.3 billion in annual revenues.

For engineers responsible for developing voice-enabled solutions for the AI market, it is paramount to choose the right components and to understand the different features that are available. To aid in the development of voice-activation products, Qualcomm® Technologies' Snapdragon 800 processers come equipped with Snapdragon Voice Activation (SVA) features. In collaboration with Qualcomm, Lantronix brings this advanced voice activation technology to OEMs with its Open-Q™ 800 System on Module (SOM) solutions, including the Open-Q™ 865XR.

Exceptional Power Management Capabilities

Lantronix's Open-Q SOMs support development of voice-activated devices with power management benefits, including:

- Ability for intelligent devices to conserve energy in low-power mode while waiting for a trigger word or phrase to awaken and begin processing audio input.
- A flexible technology that enables customized words or phrases.
- Powerful on-board CPUs that stay in sleep mode until a wake-up word is heard–particularly beneficial for battery-powered devices with constrained power availability.
- Audio Digital Processing System (aDSP), which allows for advantageous power conservation while the rest of the system is asleep. Until the wake word is recognized, the rest of the system will be in low-power mode.

Word-Recognition Sound Models

There are two common types of word-recognition sound models. Each model is dependent on how the audio modelling data samples were captured to differentiate between the voice of one or more users:

- ***User-Independent Voice Models***: The most commonly used model–it allows for any voice to wake the device. For example, smart home assistants allow the voice input of any user to be used as triggers for word recognition.

o **_User-Dependent Voice Models_** limit the operation/trigger/recognition to a single user's voice. By further training the modelled data input, it can match the characteristics of a single user's voice samples to make it user-dependent.

Setting up Customized Keywords

Utilizing Lantronix SOMs, engineers can set up a customized keyword User-Dependent Model (UDM) or User Independent Model (UIM) file comprising of several samples of the same keyword. Depending on the use case, either UIM or UDM can be desirable for product development. For example, UIM is typically mandatory in many devices.

At Lantronix, we have created a UIM based on samples provided by our customers in a proper database of spoken keywords. This involves many speakers and voice samples recorded according to specifications and lab environment. Proprietary tools are used to generate a UIM from the database of samples. From there, the model is integrated onto the device.

Alternatively, for demonstration purposes, there is a sample Voice Activation application, which allows the set-up of a limited set of wake words. This sample application is a good evaluation option, as it provides a few pre-built UIM keywords. It also allows the collection of at least five recordings of the keyword phrase from the primary end user in order to provide the phrases as a UDM to Qualcomm's SVA ListenEngine API. When the sound model is loaded and recognition is started, the ListenEngine configures the required hardware, and the software blocks begin detecting keywords.

Barge-In Considerations

Another consideration for SVA is the use case of barge-in. Barge-in describes a circumstance where the device may be playing audio but still needs to be able to listen and detect keywords from users. This feature is also needed for voice-command processing in cloud-based engines where the wake word and voice commands may run together.

When recognition is started for any Sound Model in non-barge-in mode, Microphone Activity Detection (MAD) hardware continuously listens for sounds near the microphone. In this case, the sounds are then matched against the sound model. If there is a match, ListenEngine notifies a callback associated with the app, waking up the CPUs from sleep mode.

By default, voice activation runs in the WCD audio codec module DSP (WDSP) (part of the Qualcomm Aqstic™ audio codec) and moves to the aDSP during an active playback. Once the audio playback activity is done, the detection moves back to the WDSP.
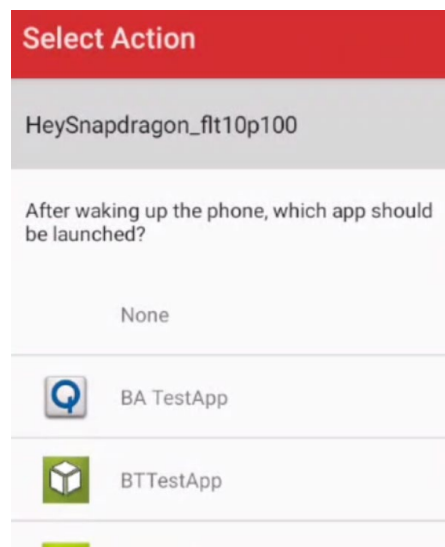
Sample Voice Activation Application
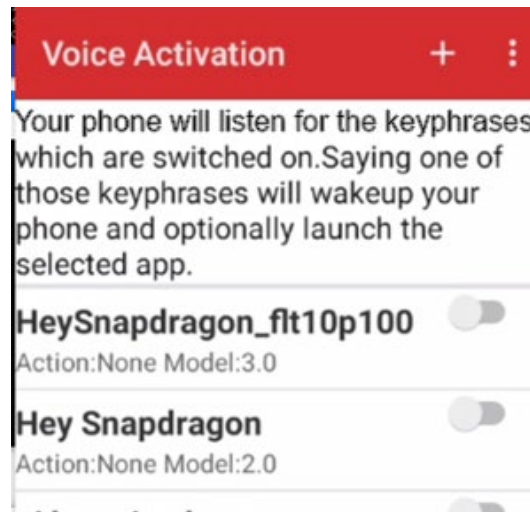
After using the application to select the keyword, you will see that SVA is up and running:


You can set up an application to be triggered by the wake-up word using the same tool:

To start the ListenEngine at the proper time, the MAD block needs to be set up. This can be accomplished via a software implementation on the aDSP, or by using the additional hardware vis-à-vis the WCD codec, depending on the product hardware design/configuration.

The following diagram shows some of the processing blocks involved for wake-word detection:
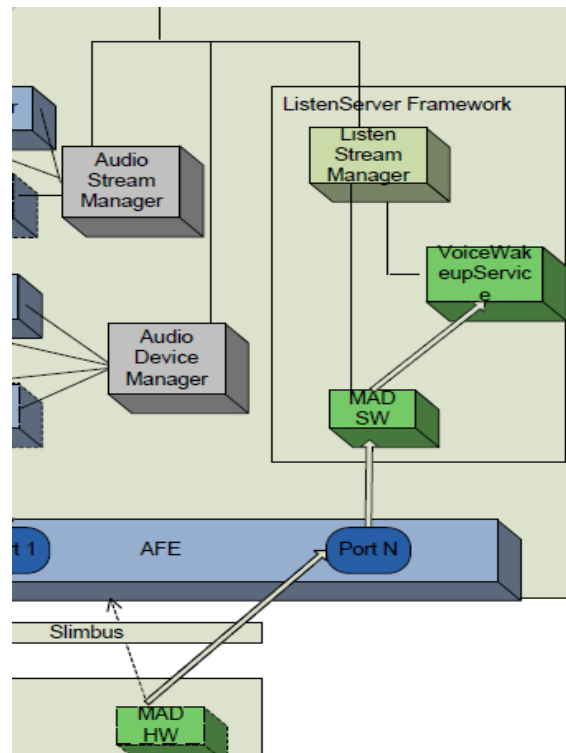
MAD hardware enables the system to run in a power-efficient mode by filtering out low-level noises. MAD hardware on the codec compares the energy level on the microphone over a 20-ms period 10 against a predefined threshold. If the energy level is higher than the threshold, it interrupts the Low Power Audio Subsystem (LPASS). MAD hardware is only capable of receiving input from one microphone. In this case of noise reduction, it would be the main microphone.
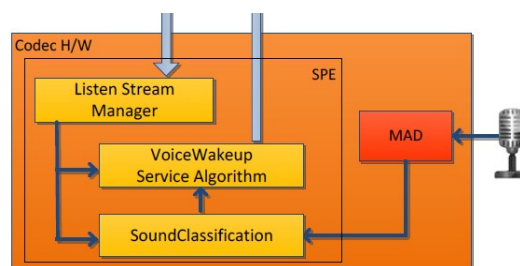
The above image shows SVA Engine (SPE) components. If SoundTrigger audio capture is enabled during startRecognition, the application is expected to read the contents of the audio buffer when detection occurs.

Guidelines for selecting a custom keyword that is suitable for identification under SPE include:

- Must contain at least 4 syllables.
- Must contain no more than 8 syllables.
- Must be at least 1 second long (even when spoken quickly).
- Should not be a phrase common in everyday conversation.

- Should maximize diversity in sound between each syllable.
- Should alternate vowel and consonant sounds for greater distinction.
- Must use the most significant distinct syllables at least 200 ms after the start of the phrase.
- Should meet the minimum confidence level for User Defined Keyword (UDK), which should be set to 69. However, if the missed detection rate is unacceptably high, drop the minimum confidence level to 65 or 61.

Lantronix is a great resource for developers and OEM engineers that are producing voice-enabled solutions for the AI market. We are able to help define and tune keyword identification and leverage the ability to define key words. To learn more, check out the Open-Q865XR and development kit, along with all of our embedded-compute solutions at www.lantronix.com.

Accelerate Time-to-Market with Symmetry Electronics

Global semiconductor distributor, Symmetry Electronics, has teamed up with Lantronix to offer their broad portfolio of embedded-compute solutions to design engineers. With hands-on training provided by Lantronix, Symmetry's technical staff serves as an extension of their sales team, offering expert-level consulting and support to customers. With an in-house Application Engineering team, Symmetry offers free comprehensive design services that help customers accelerate time-to-market and reduce costs. Browse the latest Lantronix products available at SymmetryElectronics.com.