Synaptics®

AudioSmart Voice and Speech Processing

Contents

Synaptics in the Smart Home

Executive Summary

Delivering a smart digital assistant is a now a topic that dominates the boardrooms of many major IoT companies. As voice activation and artificial intelligence (AI) transform the smart device market, engineering teams strive to deliver extraordinary experiences to customers through conservational speech.

Smart speakers are now more than just a widespread curiosity in increasing numbers of households. They are a key platform for driving digital assistant adoption in the smart device ecosystem.

Synaptics®, an industry leader at the forefront of voice and speech processing and Digital Signal Processing (DSP) technology, enables customers to launch digital assistants using Synaptics AudioSmart® voice solutions.

Similar to Synaptics' innovative touch, display, and biometrics technologies, now widely recognized as industry standards, AudioSmart is regarded as the benchmark for far-field, voice-enabled IoT products. Market projections confirm that there is fertile ground for such solutions; data from IHS Markit establishes that the consumer IoT market will climb to just under $2.7 billion by the year 2020, up from $1.75 billion in 2016.

By establishing a market-leading AudioSmart presence in the Amazon® Alexa® Voice Services (AVS) ecosystem, Synaptics now leads the digital assistant market with devices that combine a competitive advantage with a smaller form factor and a consumer-friendly price point.

The next stage of the smart audio revolution has arrived, with bolder objectives, and Synaptics is in a distinct leadership position to provide solutions to the entire worldwide digital assistant industry.

This white paper describes Synaptics' AudioSmart solutions for voice applications and smart audio designs.

Overview

Synaptics ignited the human interface revolution. Our touch, display, voice/audio, and biometrics products are built on the company's storied research and development, extensive intellectual property and global partnerships. With solutions designed to optimize user experience with smart devices, Cloud, mobile, PC, and automotive industries, Synaptics combines ease of use, functionality, and aesthetics to empower products to help make users' digital lives more productive, secure, and enjoyable.

Synaptics, an audio and imaging innovation leader, combines its significant technology portfolio in DSP and mixed-signal audio designs with embedded software, to deliver high-end silicon solutions that enrich and expand audio and imaging capabilities in enterprise and consumer products.

This document describes the following solutions:

- AudioSmart: Synaptics' intelligent algorithms enable high-performance audio processing in real-world conditions with the following features:
  - o Noise Reduction (NR) that is optimized for Automated Speech Recognition (ASR)
  - o Clear voice quality for Voice over Internet Protocol (VoIP) communication

- Smart Source Pickup (SSP): Synaptics' proprietary noise suppression algorithms include the following capabilities:
  - o Blind Source Separation (BSS) techniques that are modified for real-world robustness
  - o No beam-forming SSP uses spatial representation to target speech and noise sources, reducing stationary and semi non-stationary, interference (referred to as "babble" noise)
  - o Speech can be at any angle relative to the microphones
  - o Noise can be at any angle relative to the microphones, even in the same direction as the speech

AudioSmart Voice and Speech Processing

Voice processing is essential in consumer electronic devices that offer voice communication and hands-free control through Automatic Speech Recognition (ASR). Synaptics' AudioSmart solution provides high-quality, far-field voice processing that allows clear voice communication and accurate voice control, even when the user is five meters (5m) or more from the device. The AudioSmart experience isolates the voice source of interest, and is independent of its location regardless of loud background noise.

High-quality, far-field voice processing is increasingly important for the following applications:

- Digital Assistants
- Smart TVs
- Windows-based PCs, laptops, and convertibles
- Android-based smart phones, tablets, laptops, and wearable computers
- Industrial and Smart Home applications, including networked safety and warning devices, appliances, and Heating, Ventilating, and Air Conditioning (HVAC) equipment

Consumers expect the freedom to communicate and control their devices through a broader range of connected devices. Synaptics' solutions enable clear voice communication and accurate speech control that consumers expect. Users can now have an intelligible video chat with friends on a tablet in a noisy coffee shop, or control and communicate hands-free with a Smart phone while driving, despite surrounding wind and road noise. Using voice in far-field conditions to enhance the user interface is a trend that is currently seen in everything from gaming consoles to cars and TVs. Widespread adoption is advancing in other areas, such as consumer purchases and digital commerce, health and fitness devices, and homes completely controlled by voice.

AudioSmart solutions offer the optimum mixed-signal and DSP technology for high-fidelity voice and audio processing. Synaptics recognizes that voice is a natural extension of the UI, and is the first to offer a solution featuring two microphones with flexible placement options. AudioSmart's ability to deliver a high hit rate (the number of accurately recognized words or commands) and excellent voice clarity at a distance up to 5m is another industry first.

AudioSmart's far-field voice processing includes the following key features:

- Two-microphone solution with flexible microphone placement options for a low price point

- Four-microphone solution with Smart Source Locator for contextual awareness

- High hit-rate ASR and voice clarity, even when the user is positioned up to 5m from the device

- Excellent voice quality, independent of the user's angle relative to the device or the device's orientation (portrait and landscape)

Problem Statement

This white paper describes the challenges associated with useful voice and speech capture in the real world, as well as the attributes and applications of Synaptics' technologies designed to overcome these problems and enhance the voice and speech-driven experience.

Enabling robust voice communication and ASR in a real-world environment is challenging. The environment is filled with sounds and noises that interfere with speech, and the devices being used can themselves be playing back audio content while the user is speaking. If a user is some distance from the device being used, the speech is further colored by reverberation caused by the speech reflecting off surfaces in the user's surroundings. All these effects combine to degrade voice clarity for communications and diminish the performance of ASR systems.

Voice Processing Technology Background

Most voice processing solutions applied in mobile devices today rely on some variant of algorithms that use Interaural Level Difference (ILD) between two microphones. Although this solution works well for phones in handset mode when the phone is held next to the user's face and one microphone is much closer to the user's mouth than the other microphone, these types of algorithm solutions break down when the user is 0.5m or more from the phone.

Another common class of voice processing solutions relies on beam-forming, such as the Minimum Variance Distortionless Response (MVDR). The drawback of beam-forming solutions is that they are highly ineffective when using only two microphones, and require an increased number of microphones to perform well. This solution is too expensive to implement in consumer devices. Additionally, beam-forming solutions cannot handle interference that comes from the same direction as the speech.

To overcome these limitations, Synaptics has developed a new framework for pre-processing based on BSS. Synaptics' algorithm framework is based on the state-of-art multichannel BSS theory, and uses statistical independence to deconstruct the acoustic scene to its atomic sound components.

Synaptics' voice processing solution has solved the many challenges of far-field communication (0.25m to 5m):

- Interferences can be at the same level or higher than the preferred voice signal.

- The effect of reverb that distorts the signal is dramatically reduced, improving speech clarity and Speech Recognition (SR) performance.

- The dynamic range of the preferred signal and interference is dramatically increased.

- Full-duplex communication is possible in the presence of multichannel playback.

- The orientation, distance, and location are independent.

Voice Processing Technology Applied

Smart Source Pickup – Solving the Far-Field Problem

Traditionally, it has been extremely challenging to deliver clear communication and robust SR when the user is far away from a microphone. Figure 1 shows a classical MVDR beam-forming solution, which has many disadvantages:

- In reverberant conditions, there is considerable residual speech in the noise estimation channel, which causes speech distortion in the adaptive post-filter stage. The only way to overcome this within the MVDR framework is to add more microphones, which adds to the overall cost of the solution.

- The MVDR solution requires that the location of the preferred speaker be constantly tracked. A mistake in the tracking causes speech distortion or cancellation of speech.

- A beam-forming solution cannot remove an interfering noise source that has the same direction of arrival to the microphones array as the preferred source.

These limitations make beam-forming an impractical solution.
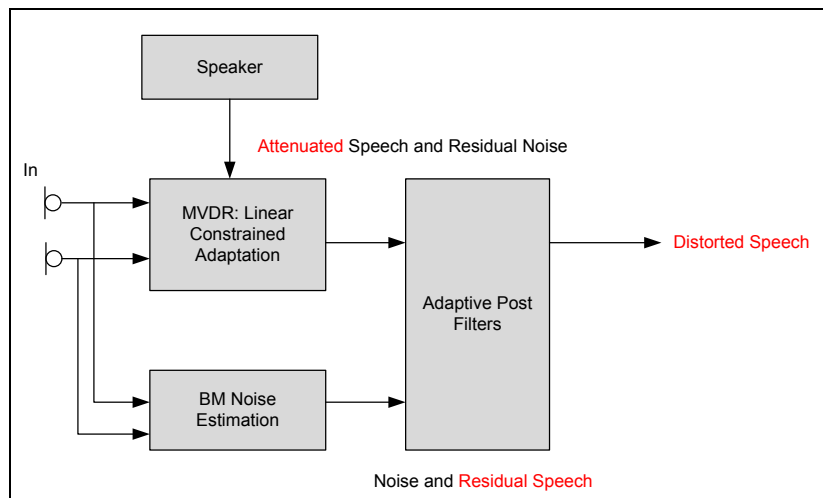
Figure 1. Typical MVDR beam former

To overcome the limitation of beam-forming and ILD solutions, Synaptics has developed a new algorithm framework using BSS. Traditional BSS approaches have suffered from a lack of robustness in real-world conditions. Synaptics has been able to overcome these issues and develop a solution based on constrained Independent Component Analysis (ICA), as shown in Figure 2.
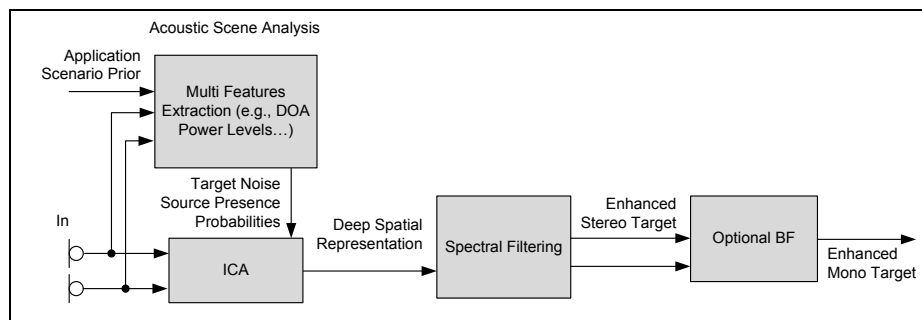

Figure 2. Synaptics' Smart Source Pickup solution

This algorithm performs dynamic acoustic scene analysis that produces multiple features used to condition the ICA adaptation. The features include estimation of number of acoustic sources, direction of arrival estimation, and classification of sources into interference, speech sources, and various statistical measures.

The ICA produces a "deep" spatial representation of the target sources (users speaking) and noise sources, even in highly reverberant conditions, because reverberation is implicitly modeled in the filtering. The features and estimated spatial filters are used to control a statistically-based spectral filter that enhances the signal.

The enhanced signal can be a true stereo output, where spatial information in the preferred signal(s) is preserved while removing an unwanted signal from both channels. The output can also be dedicated as a true mono signal or a mono signal derived from the stereo signal through optional direct-sum beam-forming.

Both stationary and non-stationary noise cancellation are possible, and strict directional constraints or ILD constraints are not limitations.

The Synaptics SSP framework creates new possibilities that were previously unattainable with two-microphone solutions. The SSP solution delivers high signal-to-noise improvement, both in near-field and far-field, without noticeable speech distortion. The preferred speech can be at any angle relative to the microphones. The noise can also be at any angle relative to the microphones, even in the same direction as the preferred speech.
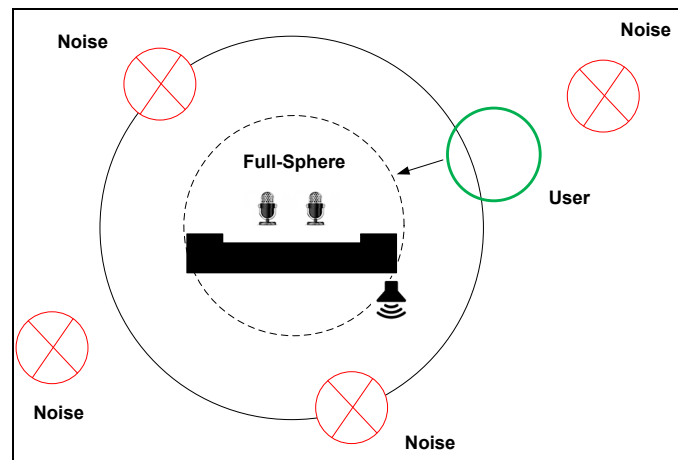

Figure 3. Smart source pickup (SSP) framework

The SSP framework makes the algorithm orientation, distance, and location independent — and the solution highly cost-effective, because only two microphones are necessary. This property of the SSP framework is valuable for mobile devices, such as tablets and cell phones, in which NR and voice enhancement performance are independent from the way the user holds and rotates the device. For PCs, TVs, or other devices, this property gives the user the freedom to communicate with the device from any location or direction.

If desired, the algorithm can be set to privacy mode: only the user from the preferred direction is allowed — everything other than the preferred direction is removed. In privacy mode, noise sources from the direction of the preferred user are still filtered out.

Real World Results… *Synaptics Delivers!*

Superior Speech Recognition Performance

Worst-Case Conditions for Speech Recognition

- Commands are not understandable to human ears.

- Natural language voice commands at microphones: 60 dB.

- Playback volume (echo) at microphones: 72 dB.

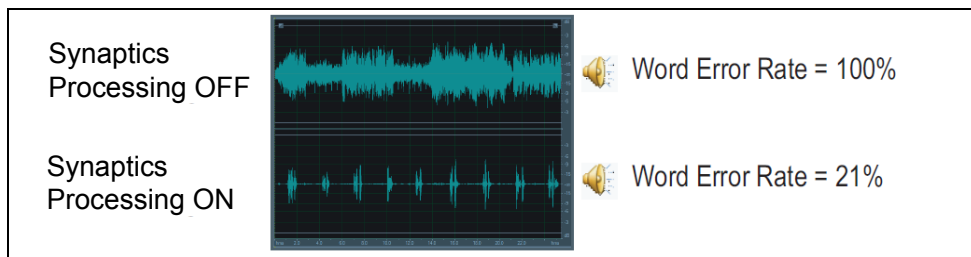- Side interference (such as "babble" noise) at microphones: 50 dB.

Figure 4. Synaptics processing dramatically reduces the word error rate

As illustrated in Figure 5, there is approximately a 30 dB reduction in echo and interference. An extraordinary improvement for the user experience.
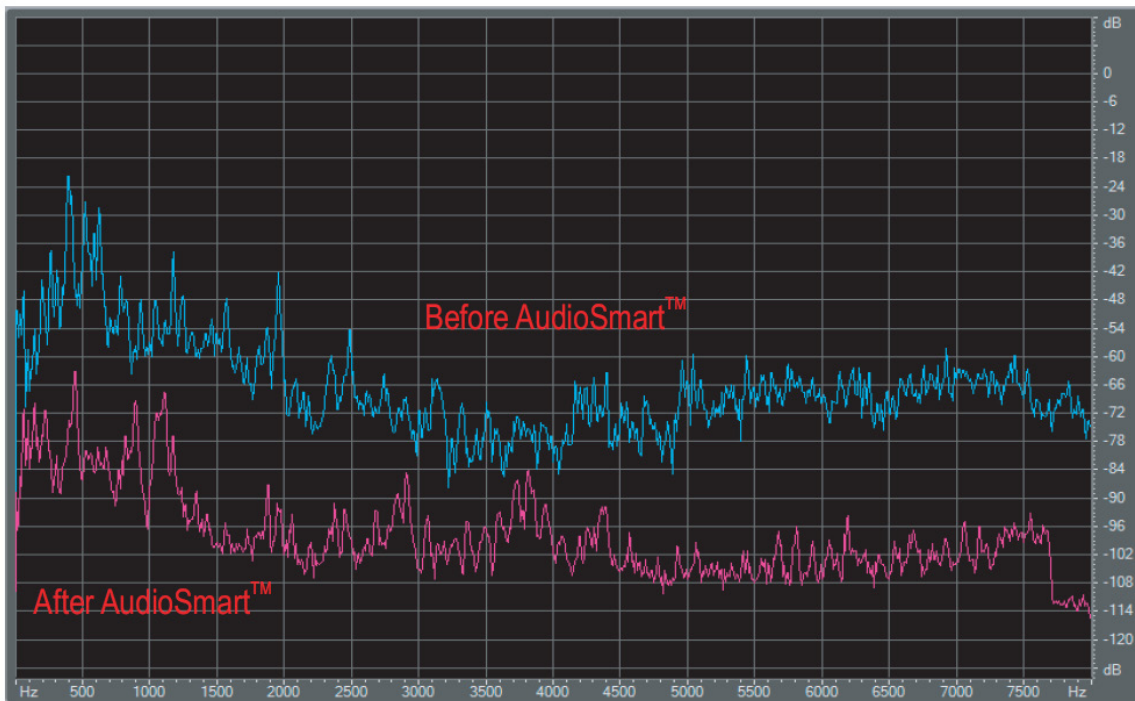
Figure 5. Approximately 30 dB reduction in echo and interference with AudioSmart

In Table 1, a tablet was:

- Taken into various real-world environments
    - The SR was measured with an existing tablet software solution
    - Recordings were taken with the tablet in the same environments, processed offline with Synaptics software, and hit rates were measured
- Held chest high, about 0.5m away from the mouth

Table 1. Google Cloud Automatic Speech Recognition (ASR) word error rate processing

Condition	Non-Synaptics Processing Google Cloud ASR (Word Error Rate)	Synaptics Processing Google Cloud ASR (Word Error Rate)
Home (Quiet)	24.5%	1.3%
Home (Noisy)	40.4%	0.7%
Church Hall (Noisy)	27.2%	14.6%
Playground (Day 1)	21.1%	4.0%
Playground (Day 2)	22.1%	9.3%
Restaurant (Set 1)	53.6%	7.3%
Restaurant (Set 2)	26.4%	5.0%

Dramatic Voice Communication Improvement

Before and After Synaptics Processing in the Real World

Figure 6 shows the before and after processing by Synaptics for real-world "babble" noise recorded in a public hall.
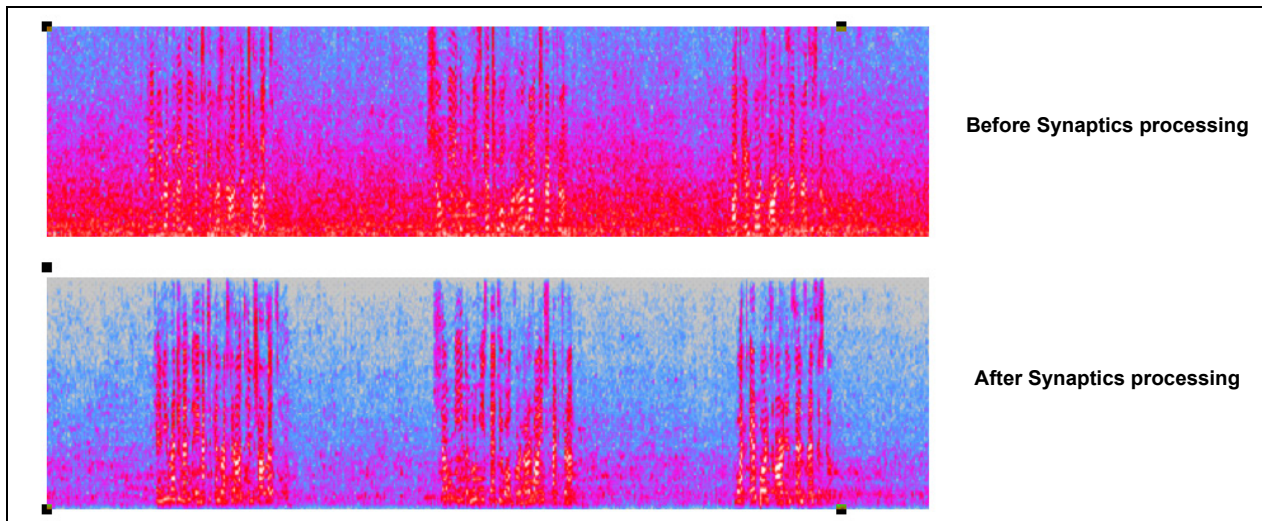
Figure 6. Before and after processing by Synaptics for real-world "babble" noise reduction

Figure 7 shows the before and after processing by Synaptics for real-world domestic non-stationary noise (such as TV, piano playing, kids talking, and so forth) recorded in a home.
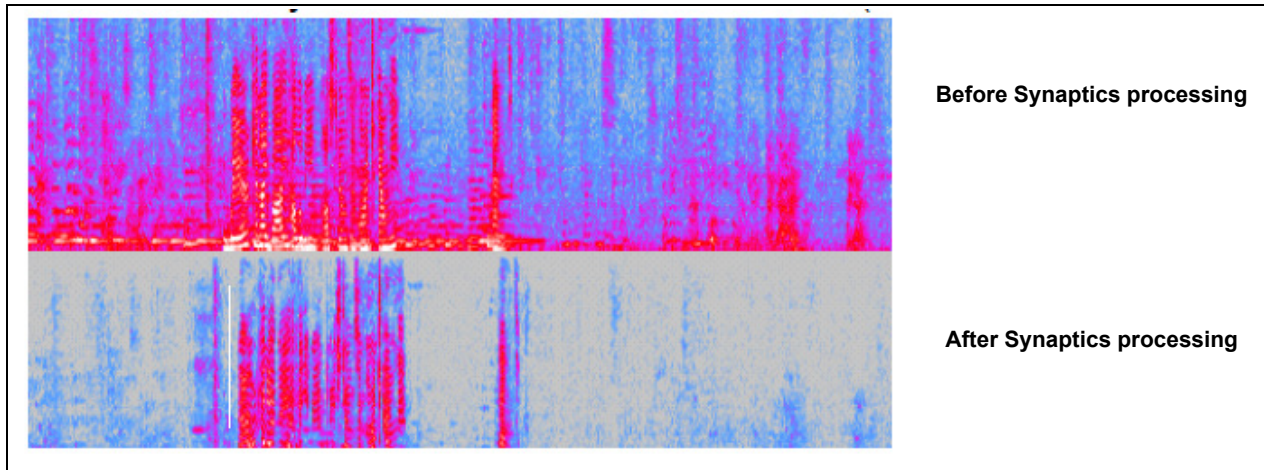

Before Synaptics processing

After Synaptics processing

Figure 7. Before and after processing by Synaptics for real-world domestic non-stationary noise reduction

Keystroke Noise Suppression

Synaptics offers advanced DSP technology to dramatically reduce annoying "key click" noise distractions during PC-based voice communication:

- It uses proprietary Synaptics algorithms.

- It can be attached to capture and render endpoints with the following advantages:

    o Local user benefits (external/remote keystroke noise removed from the inbound audio signal)

    o Remote user benefits (local machine keystroke noise removed from the outbound audio signal)
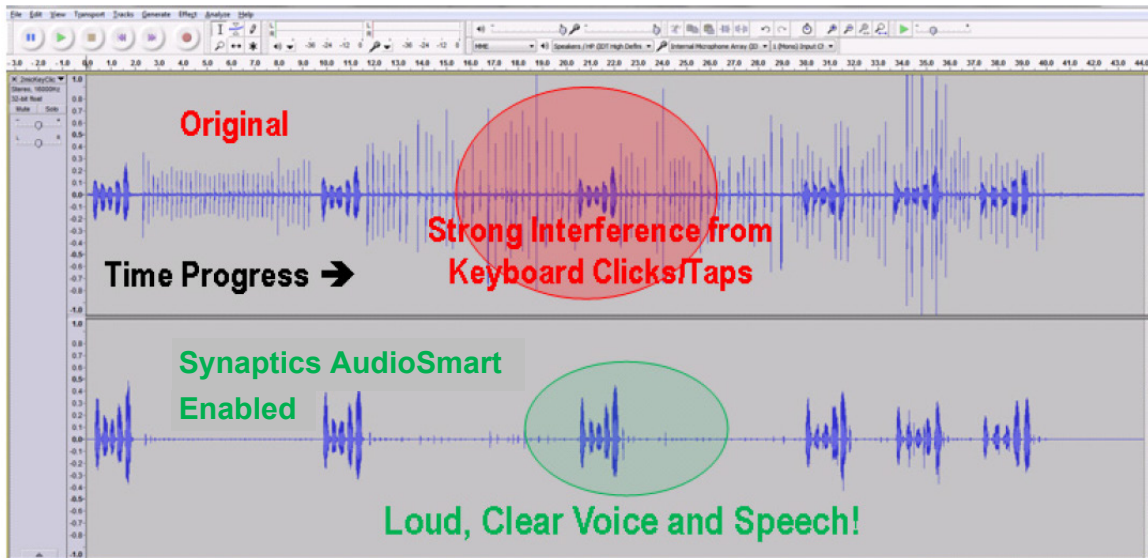
Figure 8. Synaptics dramatically reduces "key click" noise distraction during PC-based voice communication

Implementation Options

Flexible Technology Components Reduce Overall System Cost

Portable, Modular Software Components

Synaptics AudioSmart algorithms run on Synaptics DSP and third-party cores. Algorithms are available as Windows Audio Processing Object (APO) plug-ins for x86 or as Android plug-ins for x86 and ARM. AudioSmart-embedded applications are also supported.

Low-Power, High-Performance Audio Coder-Decoder (CODEC) Integrated Circuits (ICs)

Voice/speech/audio Coder-Decoder (CODEC) Integrated Circuits (ICs) are optimized for high voice quality and low power.

Integrated, Turnkey Digital Signal Processing (DSP) CODEC Processors

Voice/speech/audio CODEC ICs with integrated voice processing minimize implementation effort.

Application Spotlight: Smart TV, Amazon Alexa, and Synaptics

In contrast with existing voice processing solutions that were created for near-field phone applications, Synaptics' technology was created with the most challenging far- field environment in mind, the Smart TV. Leading the way to a smarter, voice-enabled future, Synaptics has proven the power of its far-field voice input processing solutions for Smart TVs.

The groundbreaking 2-mic CX20921 incorporated Synaptics' third generation, far-field voice processing technology, thus offering a highly accurate, robust method to control TVs with voice. This approach was adopted in 2014 by leaders of the Smart TV industry. Fast-forward to 2017, Amazon selected the Synaptics 2-mic CX20921 and 4-mic CX20924 devices to help bring the industry-changing Alexa Voice Services digital assistant to a broad ecosystem of Smart Home product developers in segments such as Smart Speakers, appliances, lighting, HVAC, networking, and home security.

Synaptics brought the world's first far-field class microphone Analog-to-Digital Converter (ADC) to market. This included proprietary algorithms for far-field processing to help suppress undesired surrounding noises in the environment, placing the focus on the dominant voice signal in the room. This processing ensures an extremely high ASR rate and clearer voice communications when a user is at a distance of 5m or less from the device.

With highly accurate far-field ASR performance, Synaptics goes one step further by enabling natural language speech control; users are not restricted to a specific set of commands that the device can understand. Synaptics' high-performance voice processing allows users to speak to their TVs, digital assistants, and computers in a normal fashion, using the words and phrases of their choosing for a truly smart audio experience.

Synaptics' chip solution for Smart TVs and Smart Home applications is in production and available now. The CX20921 and CX20924 EVKs are also available on Amazon's website. Synaptics' software-only AudioSmart solution is available for Windows tablets and PCs. Visit the Synaptics website at https://www.synaptics.com for additional information.

Copyright

Trademarks

Notice