Conexant AudioSmart^{TM} Voice + Speech Processing

White Paper

WHP-203968
February 2014

Revision History

Revision	Date	Description
C	February 12, 2014	Revision C Release

Acknowledgements

Initiated and released by the Conexant EngineeringSmart for Sound (ES^2) Team, this document was developed with support from across the organization and in direct collaboration with the following:

Key Contributors

Trausti Thormundsson (*Conexant*)
Vineet Ganju *(Conexant)*
Whit Hutson (*Conexant*)

Technical Reviewers

Whit Hutson (*Conexant*)
Vineet Ganju *(Conexant)*

The Conexant ES^{2}Team recognizes their efforts in helping to ensure delivery of an accurate and comprehensive technical resource to support the broader audio community and systematically extinguish interfering noises from everywhere.

Feedback

Please send comments or suggestions about this document to the Conexant ES^2 Team feedback alias ES^2_feedback@conexant.com.

Contents

Figures

Voice + Speech Processing

Conexant and AudioSmart^{TM} Voice + Speech Processing

Voice processing is important in consumer electronic devices that offer voice communication and hands-free control through speech recognition. Conexant's AudioSmart^{TM} Solution provides high-quality, far-field voice processing that allows clear voice communication and accurate voice control, even when the user is 4 meters or more from the device. ***Conexant's AudioSmart^{TM} experience isolates the voice source of interest, independent of its location and in spite of strong background noise.***

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

- 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 HVAC equipment

These days, consumers expect the freedom to communicate and control through a broader range of connected devices. Conexant's solution enables the 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 with a Smart Phone hands-free 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. This trend is poised for widespread adoption in other areas, such as fitness devices and homes completely controlled by voice.

Conexant's AudioSmart^{TM} solutions offer the optimum mixed-signal and DSP technology for high-fidelity voice and audio processing. Conexant recognizes that voice is a natural extension of the user interface and is the first to offer a solution featuring two microphones with flexible placement options. Conexant's AudioSmart^{TM}'s ability to deliver a high hit rate and excellent voice clarity at a distance of up to four meters is another industry first.

Conexant AudioSmart$^{TM's}$ far-field voice processing includes the following key features:

- Two-microphone solution with flexible microphone placement options for lower costs
- High hit-rate speech recognition and voice clarity, even when user is up to four meters from the device
- Excellent voice quality, independent of the user's angle relative to the device or device orientation (i.e., portrait, 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 Conexant technologies designed to overcome these problems and enhance the voice + speech-driven experience. Enabling robust voice communication and speech recognition in a real world environment is challenging. The environment around us is filled with sounds and noises that interfere with the desired speech, and the devices that we are using can themselves be playing back audio content while we are speaking to them. If a user is some distance from the device being used, the speech will further be 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 degrade the performance of speech recognition systems.

Introduction

Conexant has over 20 years experience developing voice processing algorithms, starting with speakerphone solutions included in voice band modems. In 2011, Conexant shifted its focus to develop Natural Language Pre-Processing for far-field communication and speech recognition with the goal of allowing people to communicate with devices at a distance, to truly enable the promise of seamless speech control and communication with virtual assistants.

Background

Most voice processing solutions applied in mobile devices today rely on some variant of algorithms that utilize level difference between two microphones (ILD). This works very well for phones in handset mode, where 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. However, when the user is 0.5 meters or more from the phone, these types of algorithm solutions break down. Another common class of voice processing solutions rely on beam-forming, such as the Minimum Variance Distortionless Response (MVDR).The drawback of beam-forming solutions is that they are very ineffective when using only two microphones and require a large number of microphones to perform well. This makes the solution too expensive to implement in consumer devices. In addition, beam-forming cannot handle interference that comes from the same direction as the desired speech.

To overcome these limitations, Conexant has developed a new framework for pre-processing based on Blind Source Separation (BSS). Our algorithm framework is based on the state-of-art multichannel BSS theory, where we use statistical independence to decompose the acoustic scene into its atomic sound components.

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

- Interferences can be at the same level or higher than the desired voice signal
- The effect of reverb that distorts the signal is reduced, dramatically improving speech clarity and speech recognition performance
- Dynamic range of desired signal and interference can be greatly increased
- Full duplex communication is possible in the presence of multichannel playback
- Orientation, distance, and location independence

Definitions

AudioSmart™: Conexant software solution that enables high performance audio processing in real-world conditions and enables the following features:

- Noise reduction optimized for speech recognition
- Clear voice quality for VoIP communication

Smart Source Pickup (SSP): Conexant proprietary noise suppression algorithm

- Based on a Blind Source Separation technique, modified for real-world robustness
- ***No*** Beamforming: SSP uses spatial representation of target speech and noise sources to reduce stationary and semi non-stationary interference (i.e., babble noise)
- Speech can be at any angle relative to the microphones
- Noise can be at any angle relative to the microphones, even in the same direction as the speech

Technology in Application

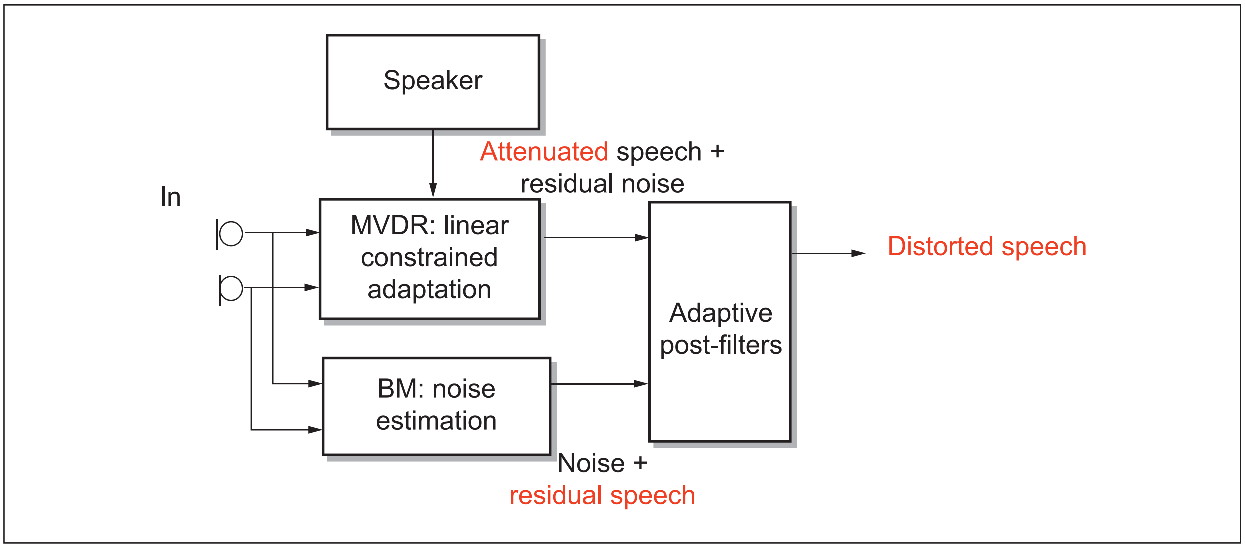
Smart Source Pickup, Solving the Far-field Problem

Traditionally, it has been very challenging to enable clear communication and robust speech recognition when the user is far away from the microphones. A classical MVDR beam-forming solution is shown in Figure 1. This type of beam-forming solution 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, adding to the overall cost of the solution.
- The MVDR solution requires that the location of the desired 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 desired source.

These limitations make beam-forming an impractical solution.

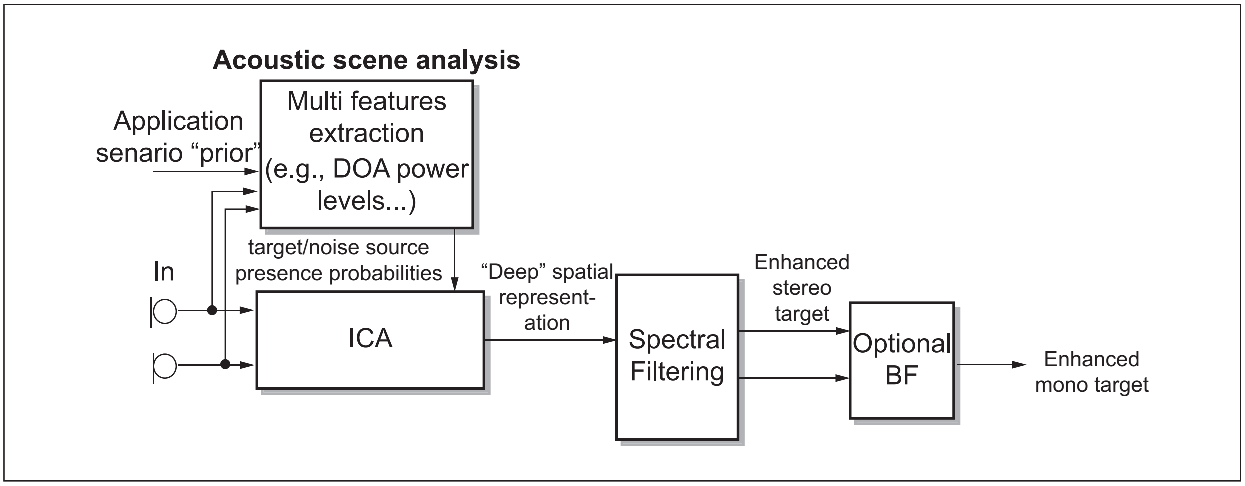
Figure 1. MVDR Beam-former



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

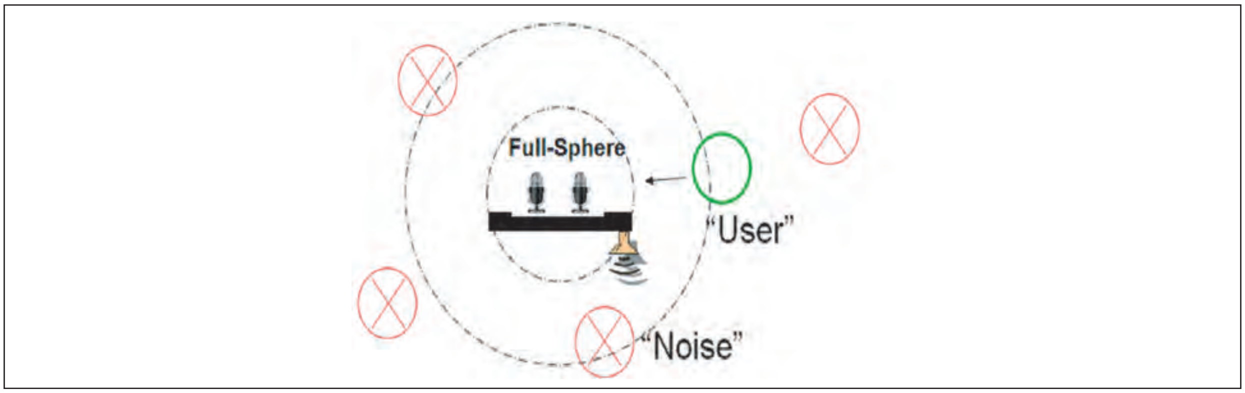
The ICA produces a "deep" spatial representation of the target sources and the noise sources, even in highly reverberant conditions, because reverberation is implicitly modeled in the filtering. The features and estimated spatial filters are used to control a statistically based spectral filter that enhances the signal. The enhanced signal can be a true stereo output, where spatial information in the desired signal/signals is preserved while removing 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 an optional direct sum beam-forming. Both stationary and non-stationary noise cancellation are possible, and strict directional constraints or inter-channel level difference (ILD) constraints are not limitations.

Figure 2. Smart Source Pickup (SSP)



The SSP framework from Conexant has opened up new possibilities previously unattainable with two-microphone solutions. The solution delivers high signal-to-noise improvement, both in near-field and far-field, without noticeable speech distortion. The desired 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 desired speech.

Figure 3. SSP Framework



This makes the algorithm orientation, distance, and location independent (and the solution highly cost-effective, because only two microphones are necessary). This property is valuable for mobile devices like tablets and cell phones, where noise reduction and voice enhancement performance are independent from the way the user holds and rotates the device. For PCs, TVs, or other devices, this gives the user the freedom to communicate with the device from any location or direction.

If desired, the algorithm can be configured to privacy mode, where only a user from a desired direction is allowed, and everything outside this direction is removed. In privacy mode, a noise source from the direction of the desired user is still removed.

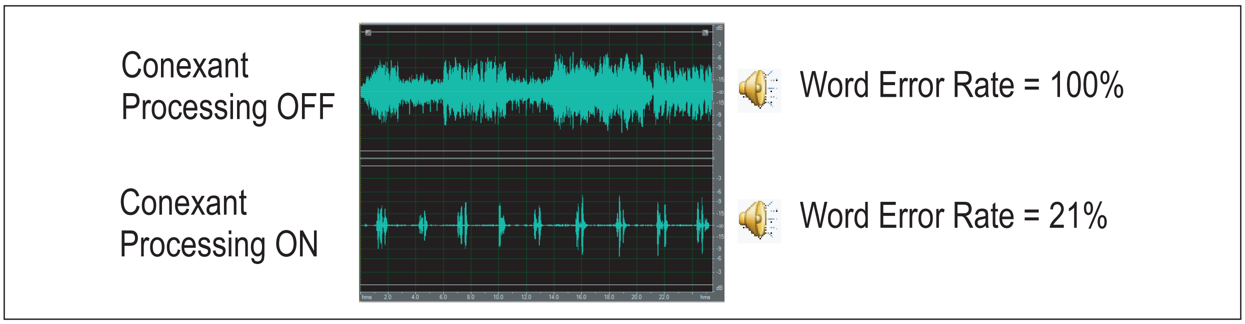
Real World Results

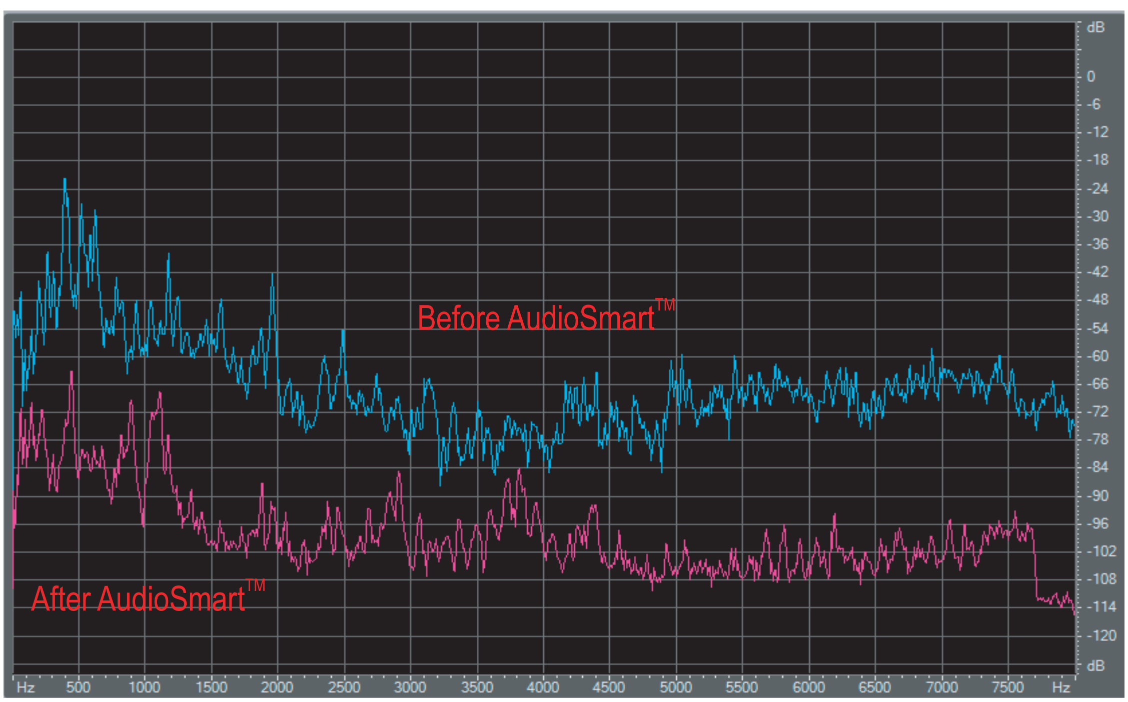
Conexant delivers!

Excellent Speech Recognition Performance

Worst case conditions:

- Commands are not understandable to human ears
- Natural language voice commands at microphones: 60 dB
- Playback volume (echo) at microphones: 72 dB
- Side interference (e.g., babble noise) at microphones: 50 dB





Around 30 dB reduction in echo and interference!

Condition	Non-Conexant Processing Google Cloud ASR (Word Error Rate)	Conexant 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%

*Tablet was taken into various real world environments:

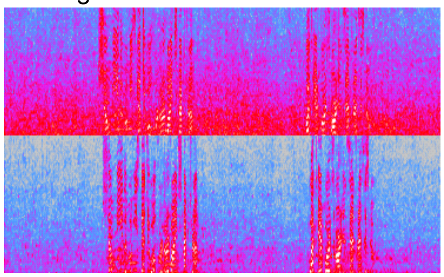
- SR was measured with existing tablet software solution
- Recordings were taken with tablet in same environments and processed offline with Conexant software and hit rates measured

*Tablet was held chest high, about 0.5 m away from the mouth

Dramatic Voice Communication Improvement

Before and After Conexant Processing in the Real World

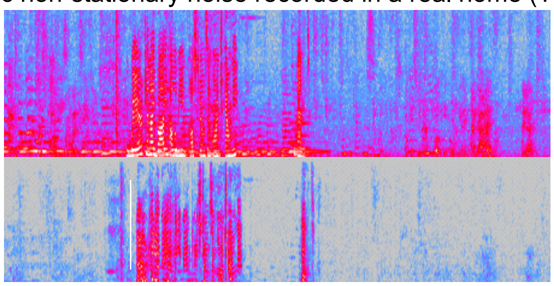
Church Hall: real-world strong diffuse babble noise recorded in a church hall



Before Conexant Processing

After Conexant Processing

Noisy Home: Domestic non-stationary noise recorded in a real home (TV sound, piano playing, kids talking…)



Before Conexant Processing

After Conexant Processing

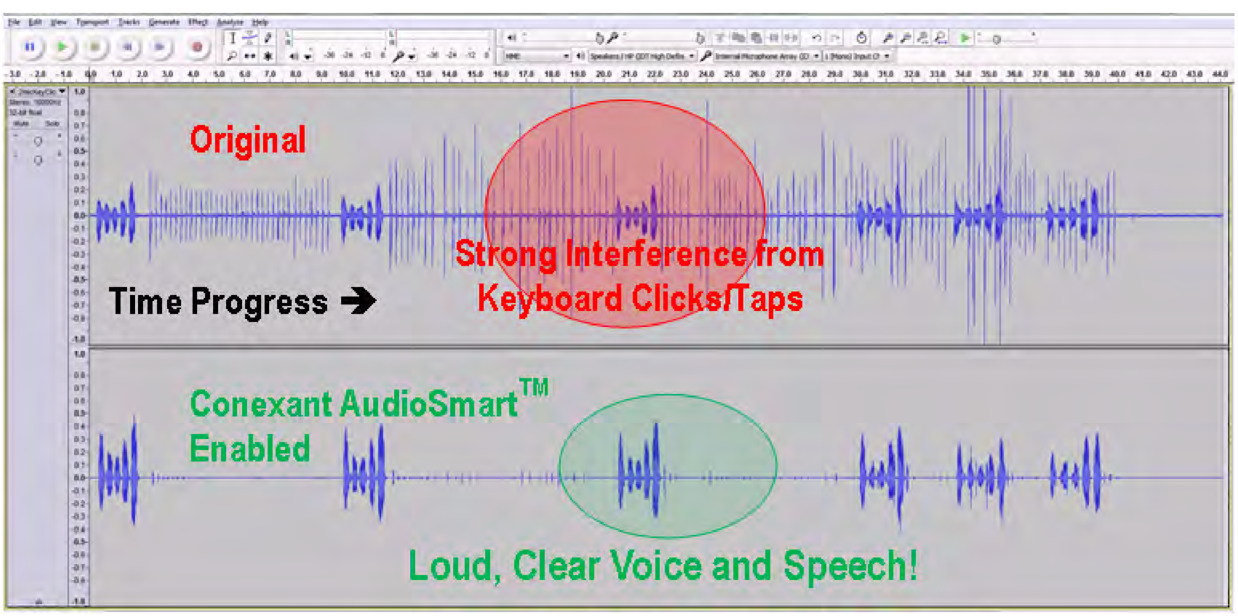
Keystroke Noise Suppression

Conexant offers advanced DSP technology to dramatically reduce annoying “key click” noise distractions during PC-based voice communications.

- Uses proprietary Conexant algorithms
- ***Can be attached to capture and render endpoints!***

  –Local User benefits (external/remote keystroke noise removed from inbound audio signal)

  –Remote User benefits (local machine keystroke noise removed from outbound audio signal)



Implementation Options

Flexible Technology Components Enable Lower Overall System Cost

Portable, modular software components

Conexant AudioSmart™ algorithms run on Conexant DSP and 3rd party cores. Algorithms are available as Windows APO plug-ins for x86, or Android plugins for x86 and ARM. AudioSmart™ embedded applications are also supported.

Low-power, high-performance audio CODEC ICs

Voice/Speech/Audio CODEC ICs optimized for high voice quality and low power.

Integrated, turnkey DSP CODEC processors

Voice/Speech/Audio CODEC ICs with integrated voice processing enable lower implementation effort.

Conexant AudioSmart^{TM} Software Solutions

Conexant offers both capture (record) and render (playback) DSP algorithms for Speech + Voice and Audio Experience Enhancement. These effects are available in three pre-packaged AudioSmart^{TM} configurations, each optimized for a particular user experience and price point.

AudioSmart^{TM} Package	Target Market	User Experience Benefits	Catalog
AutoSmart^{TM} **Classic** **Lowest Price**	Optimized for Average Consumers	• **Clear quality for VoIP** ✓ Skype calls sound good with noisy background • **Noise-reduced Speech Input** ✓ Improved Google Speech Engine Accuracy with Noisy Background • **Basic Speaker Correction and Playback Enhancement** ✓ Playback is free of distortion, sounds balanced	**Android: CX3801** **Windows: CX3800** **Windows: + Android: CX3802**
AutoSmart^{TM} **PRO** **Premium Price Point**	Optimized for the Modern Corporate Workforce	• **Most Advanced VoIP Experience** ✓ Conference Speakerphone Mode with mid- and far-field interaction for multiple users ✓ Commuter Mode with train noise, road noise, engine noise, and horn noise removal • **Most Advanced Noise-Reduced Speech Input (far- and mid-field)** ✓ Exceeds latest Intel ASR requirements ✓ Allows Speech Command and Control from across a noisy conference room • **Keyboard Click Noise Reduction** ✓ Filters Outbound and Inbound typing noise • **Automatic Loudness + Ear Fatigue Protection for Headphones/ Headsets** • **Improved Intelligibility of Voice Communications** ✓ Enhances Outbound and Inbound Voice signals • **Advanced Playback Enhancement** ✓ Wideband Voice Restoration makes conversation more natural ✓ AudioSmart^{TM} BOOST ensures listening experience is loud and clear • **Advanced Speaker Correction with Bass/Treble Enhancement for small form-factors**	**Windows: CX3803** **Android: CX3804** **Windows: + Android: CX3805**
AutoSmart^{TM} **MAX** **Mainstream Price Point**	Optimized for Multimedia and Gaming Enthusiasts	• **Improved VoIP for Gaming** ✓ Full duplex ✓ Party Mode with social background noise removal • **Advanced Noise-Reduced Speech Input (mid-field)** ✓ Exceeds latest Intel ASR requirements ✓ Allows Speech Command and Control from across a noisy room • **Vocal-removal for Karaoke** • **Voice morphing for Gaming and Karaoke**	**Windows: CX3806** **Android: CX3807** **Windows: + Android: CX3808**

- Works with *all* CODECs
- Compatible with *all* Major Branded APOs and Plug-ins:
  –Works with Waves, DTS, Dolby, Beats, Sound Research, etc.
- Conexant fully supports 3rd party tuning + integration (end-to-end)
- Conexant provides compiled binaries and turnkey tuning parameters
- 3rd party or OEM integrates binary files into Windows audio driver package or Android image
- Conexant provides GUI or API for control integration with 3rd party GUI
- CODEC vendor WHQL-certifies final driver package for shipping under Windows
  –Conexant solution proven to pass Win 8.1 HCK with standard MS SysFX implementation. ZERO certification concerns
- Demos available now for OEM platforms

Application Spotlight: Smart TV and Conexant

In contrast with existing voice processing solutions that were created for the near-field phone application, Conexant's technology was created with the most challenging far- field environment in mind, the Smart TV. Leading the way to a smarter, voice-enabled future, Conexant has successfully demonstrated the power of its far-field voice input processing solutions for Smart TVs. The recently announced CX20921, incorporating Conexant's 3rd generation far-field voice processing technology, offers a highly accurate, robust way to control TVs with voice, and has been adopted in 2014 by leaders of the Smart TV industry.

Conexant has also brought the world's first far-field class microphone ADC to market. Proprietary algorithms for far-field processing help to suppress certain surrounding noises in the environment, placing the focus on the dominant voice signal in the room. This ensures an extremely high speech recognition rate and clearer voice communications when a user is at a distance from the device (within four meters). With highly accurate far-field speech recognition performance, Conexant goes one step further by enabling natural language speech control, wherein users are not tied to a specific set of commands that the end device can understand. Its high-performance voice processing allows users to speak to their TVs and computers in a normal fashion, using words and phrases of their choice for a truly smart audio experience.

Conexant's chip solution for Smart TVs and appliances is in production and available now. Conexant's software-only AudioSmart™ solution is also available for Windows tablets and PCs, Chromebooks, and Android devices of all kinds. Please visit www.conexant.com for more details.

About Conexant

Conexant Systems, Inc., an audio and imaging innovation leader, combines its significant IP portfolio in DSP and mixed-signal technology with embedded software, delivering highly innovative software and silicon solutions to enrich and expand audio and imaging capabilities. Both enterprise and consumer markets are addressed by Conexant. Products built in with the company's technology include PCs, tablet computers, smart phones, TVs, headsets, printers, video monitors, game consoles, and a variety of other devices.

Founded in 1999, Conexant is a privately-held fab-less semiconductor company headquartered in Irvine, California with offices worldwide and design centers in the U.S. and Asia. To learn more, please visit www.conexant.com, follow us on LinkedIn, and like us on Facebook.

www.conexant.com
General Information:
U.S. and Canada: (888) 855-4562
International: (949) 483-4600
Headquarters – Conexant Systems, Inc.
1901 Main Street, Suite 300
Irvine, CA 92614US

CONEXANT®