Far-Field Voice Input Processing

Meaghan Ziemba, Editor, WDD

Conexant has created a new set of advanced algorithms to help deliver a clear and easily understandable signal from a far-field source.

We have all experienced the sense of frustration when losing the one remote that controls everything from the television's surround sound to the blinds in the bedroom.

What if we were able to control all the gadgets around us by simply talking to them? No buttons to push; just tell them to do what we want – true hands-free operation.

Past attempts of providing this functionality have not been successful due to the difficulty with far-field voice input processing (FFVIP). When a microphone is close to the mouth of the individual talking, the quality of the audio tends to be better and louder than the surrounding noise. This area in the neighborhood of the microphone is known as the near-field. The challenge arises when the user moves further away from the microphone – into the 'far-field'.

There are several techniques available to distinguish the speech of the individual talking (near-field talker) from far away disturbances. One such technique uses spatial difference between multiple microphones, as well as taking advantage of the level difference to distinguish the higher-level voice from the noise, and combining that with statistical algorithms to deliver high-quality, voice signal output.

According to Newport Beach, C A-based Conexant Systems, the voice signal can be distorted by numerous sources of noise in the far-field, which are captured by the microphone that is 12 to 15 feet away from the speaker's mouth. Conexant has created a new set of advanced algorithms to help deliver a clear and easily understandable signal from a far-field source.

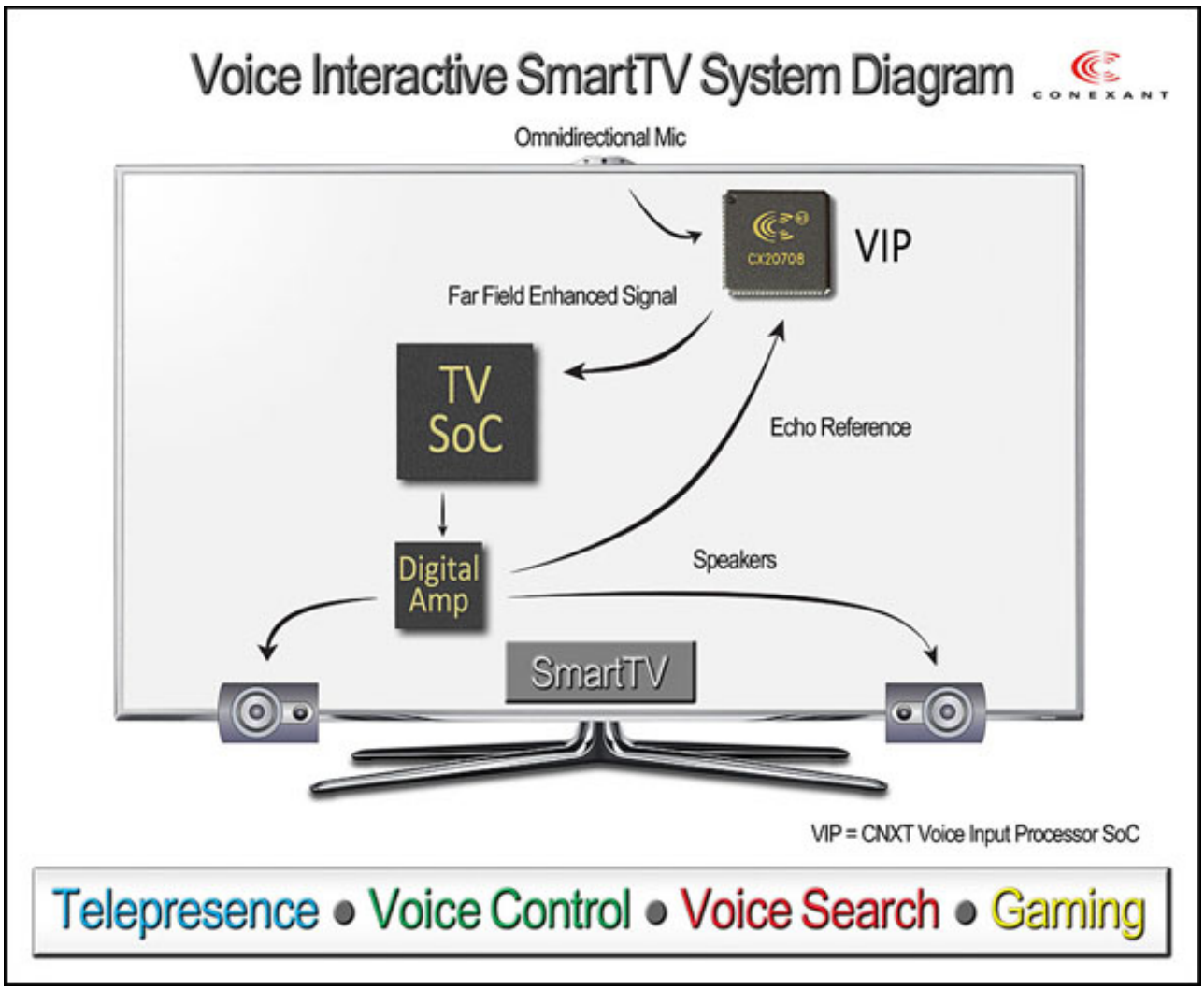
“In the far-field, users are located at distance from the microphone, and Conexant has mainly been focusing in this area, where the distance can be anywhere from 1 to 7 meters from the microphone,” explains Saleel Awsare, VP and GM of audio for Conexant. “Examples include the Kinect system from Microsoft and the Voice Controlled Smart TV from Samsung.”

Conexant has developed a suite of products that guarantee the far-field voice signals are clean to enable far-field applications, such as a Skype call while watching television, voice search for content, etc.

The algorithms that Conexant has developed for far-field processing help to suppress certain surrounding noises in the environment, and focus on the dominant voice signal in the room and ensure that the speech recognition rates are very high.

One noise that Conexant’s FFVIP technology addresses is reverberation. “Imagine sitting in a room watching television,” suggests Conexant’s CEO Sailesh Chittipeddi. “You have walls in the room that your voice bounces off of – a phenomenon called reverberation – and you need to make sure that you take care of that in terms of how your voice gets fractured.”

The voice signal that bounces off the walls often takes hundreds of milliseconds to die down, and normally the human auditory system sorts these reverberations out so they are not noticeable. But, if the sound is recorded by a microphone it becomes very pronounced and sounds as if the voice signal is being spoken from a large empty barrel.



“Another thing you have to deal with is the echo cancellation aspect of it,” explains Chittipeddi. “This is multiple-channel, acoustic echo cancellation.” Conexant’s algorithms use advanced adaptive filters to estimate the echo and perform statistical estimation of which frequency bands contain echo, and which contain the desired voice. Sophisticated control algorithms tie these together to produce a natural-sounding, echo-free voice signal.

Other background noises, such as air conditioners, vacuum cleaners, or children’s voices, need to be eliminated in order to provide true FFVIP. Conexant’s algorithms can process not just stationary noises, but time-varying noises as well, such as passing cars, airplanes, home appliances, and office machines. “Conexant’s algorithms for background noise suppression, suppresses anything that is clearly not a voice, by zooming into that person, and allowing speech recognition without impacting real voice signals,” says Awsare.

Finally, with its gain level adjustment algorithm, Conexant is able to keep the voice signal at a constant level independent of the distance to the microphone. “Whether you are speaking loudly or softly, it is able to adjust the gain level needs without changing the characteristics of the voice signal,” explains Awsare.

“Our whole objective for real-room, living environments, is to process the voice without the use of other devices like a remote control,” explains Chittipeddi. “How often do we spend our time searching for these remotes? Our algorithms for FFVIP

work around those kinds of things."

With its collection of audio-processing algorithms, Conexant's FFVIP technology removes a major stumbling block for the integration of accurate voice control and speech recognition capabilities in the leading consumer electronic devices of today.

Source URL (retrieved on *04/20/2014 - 3:36am*):
http://www.wirelessdesignmag.com/articles/2013/05/far-field-voice-input-processing