

In the last two years, consumers have become increasingly discerning when it comes to audio quality and expect to enjoy full compatibility between their devices, regardless of the brand.

A report commissioned by Qualcomm – ‘The State of Play’ – found consumers were willing to pay for products that met those requirements and that, in an already crowded market, manufacturers and developers who looked to address those needs would benefit.

Improving sound quality and compatibility are providing a vital competitive edge, but while audio technology has come a long way in a relatively short time, consumer expectations continue to grow apace.

Qualcomm Technologies recently used a developers’ conference in China to unveil what it described as its next generation direct digital feedback amplifier (DDFA) audio technology.

Targeted at high-resolution audio devices, such as wireless speakers, soundbars, networked audio and headphone amplifiers, DDFA is intended to replace more traditional class D amplifiers with an all-digital pulse width modulator, while its closed-loop architecture means that it can compensate for nonlinearities of power supplies and output stages. The result is that it can deliver much higher fidelity audio and greater design flexibility, while retaining the advantages traditionally associated with class D amplifiers.

Anthony Murray, general manager, voice and music, with Qualcomm Technologies, said the ability to deliver an improved sound experience was a major focus for audio technologists.

“DDFA is intended to help customers who are having to meet the needs of increasingly audio-savvy consumers,” he suggested.

“While DDFA has previously been used by premium audio manufacturers, this new release

The interface of choice



Major changes are happening in the audio market, but a key driver has been the emergence of smart assistants and speakers. By **Neil Tyler**.

means many more OEMs will be able to integrate high-performance amplifier capability into their products.

“The audio market is being transformed, whether that’s how consumers access their music to high resolution audio.

“We’re also seeing a dramatic increase in the demand for voice technology on the back of the emergence of smart assistants and speakers, such as the Home and Echo,” he says.

Voice recognition

Demand for voice recognition technology is growing and, while it first appeared in analogue form back in the 1950s, it’s a technology that holds

Below: The AcuEdge Development Kit from Microsemi can deliver enhanced audio processing capabilities

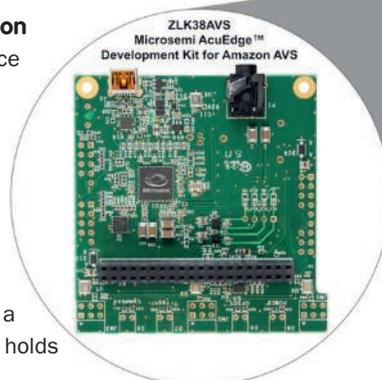
the prospect of changing the way in which we interact with our electronic devices.

Voice has traditionally been seen as a very complex technology and only available as a solution to those companies with deep pockets and the ‘super expertise’ to implement it.

Today, it’s a technology that is more widely available. Word error rates have fallen significantly and machines are quickly becoming as accurate as humans.

“As a result, voice is expected to encourage a host of new products, as well as improve existing devices,” according to Murray.

Sales of voice-activated digital assistants, such as Google Home or the market leading Amazon Echo, have been running at around



5million units per year, and market analysts are forecasting annual sales of more than 10m in 2017. The overall smart speaker/digital assistance market could reach 47m units by 2022.

While one significant company, Apple, has been accused of being late to the market, it has recently started work on a Siri-based device. Reports suggest it will be a voice-controlled speaker capable of responding to basic commands and queries but which, according to Bloomberg News, will have one key differentiator from those products already on the market – audio quality.

It is suggested that the speaker will offer ‘virtual surround sound technology’, using one speaker to create the impression that the sound is coming from several devices.

While Amazon and Google lead the market, Microsoft, IBM and Nuance are developing voice controlled products whose quality has improved significantly, allowing users to talk more naturally. Meanwhile, many companies are looking to enable complete residential IoT solutions.

“Automatic Speech Recognition (ASR) is defined as the ability of a machine or program to receive and interpret dictation or to understand and carry out spoken commands,” explains Shahin Sadeghi, Microsemi’s director of marketing and applications. “ASR Assist uses audio enhancement to target the human-to-machine voice interface specifically.”

Speech or voice recognition is a computer software program or hardware device with the ability to decode the human voice and uses ASR software.

“Most of these programs require the device to be trained to recognise a voice so that it can convert that speech into data accurately,” Sadeghi says.

A variety of voice recognition systems, including speaker dependent systems, require a series of words and phrases to be

pre-recorded. But there are also: speaker independent systems, in which no training is required; discrete speech recognition systems, in which the users may have to pause between words to enable the speech recognition software to identify each separate word; and continuous speech recognition, where the system can understand a normal rate of speaking. It’s this latter space where there has been real improvement in recent years.

“The performance of voice controlled products is markedly better,” says Sadeghi. “Key drivers have been set top boxes and smart TVs, where voice commands can be complex.

“But whether it’s Amazon, Google or Apple, all the big players are now enabling more complete ecosystems.”

The rise of ‘smart speakers’ has accelerated over the past few years and they are now seen, according to Sadeghi, as hubs for smart homes. But voice recognition and voice direction is not only targeting the consumer market.

Increasingly, voice is being used across industries ranging from healthcare and telecommunications to the military and personal computing. It is being used in a growing number of commercial settings to improve the performance of employees and being used to transform data into speech to instruct workers.



“The performance of voice controlled products is markedly better. Key drivers have been set top boxes and smart TVs, where voice commands can be complex.”
Shahin Sadeghi

Below: Companies like Microsemi, Qualcomm and XMOS want to help manufacturers accelerate the commercialisation of smart speakers



Voice is no longer seen as ‘research’ or as an interesting ‘science project’; demand for voice recognition technology is surging and opening up a host of opportunities for smaller businesses and a growing number of companies are releasing development boards, processors and low power microphones to meet that demand.

In June, Microsemi unveiled its AcuEdge Development Kit for the Amazon Alexa Voice Service (AVS).

“AcuEdge can deliver enhanced audio processing and is intended to improve voice recognition rates in adverse or difficult audio environments,” Sadeghi explains. “We are seeing demand for voice recognition accelerating and are looking to provide a much faster and easier way to prototype devices and to assist in the development of human to machine (H2M) applications being driven by the IoT and automated assistance markets.”

The kit will enable third party developers and manufacturers to evaluate and incorporate Alexa functionality into any H2M application and does this by interfacing with Microsemi’s Timberwolf ZL38063 multi-mic audio processor.

Qualcomm Technologies has also unveiled what it calls its Smart Audio Platform. “We want to help manufacturers accelerate the development and commercialisation of smart and networked speakers,” Murray explained.

“The platform is flexible. It offers two SoC options based on our APQ8009 and APQ8017 devices and, when combined with a range of software configurations, means that OEMs will be able to create smart speaker systems for multiple products and categories.

“It’s an integrated platform that brings together processing capability, a variety of connectivity options, voice user interfaces and high end audio technologies. We are finding that users are demanding highly intuitive smart speakers.”

A world of gadgets

Voice-activated gadgets can range from those with basic sound-control capabilities to smarter ones controlled by a combination of voice and gestures. Children-friendly gadgets are also finding their way on to the market, says Murray.

The market is broad and includes devices as varied as: smart helmets, using voice to control display systems; remote controls and alarm clocks that can be controlled via various voice commands; voice activated smart phones and lights; and the electronic assistants and applications mentioned earlier.

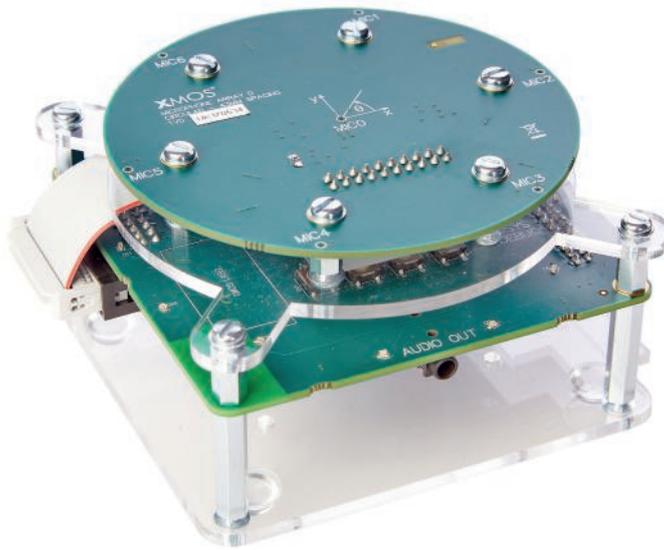
“Voice-control is by far the simplest means of guiding technology and we are seeing more devices becoming available to this type of technology,” according to Murray.

Qualcomm’s integrated voice solution has been designed to deliver advanced multi-mic far-field voice capabilities that can combine highly responsive voice activation with beamforming technologies.

“The platform’s voice software incorporates echo-cancellation, noise suppression and ‘barge-in’ capabilities, supporting a more reliable voice interface in loud or noisy environments even when users are a distance from the smart speaker,” says Murray.

As with most electronic devices, a key challenge remains power. “When it comes to speech recognition, one thing that designers are confronting –and which, for many, is a radically new challenge – is that their designs need to support ‘active all day’ speakers,” says Murray.

“That is a real technical challenge and battery life remains a serious issue. While playback times are now running at eight hours or more – double what they were a few years ago – the trade-offs have been considerable. We are currently looking at not only new battery chemistries, but also the actual shape of batteries. While I can’t say



more at this stage, new products in the pipeline will reduce power requirements significantly.”

Vesper, a sensor specialist, has developed what it believes is the first MEMS microphone that can bring voice activation to battery-powered consumer devices, and does so by drawing nearly zero power.

The company’s ZeroPower Listening MEMS microphone, the VM1010, is a tiny piezoelectric MEMS microphone that allows product designers to offer touchless user interfaces to consumers and consumes just 6µA while in listening mode, waiting for a keyword.

Power issues

“Consumers want to interact with their electronic devices using just their voice and there is an explosion in voice-enabled consumer electronics, including smart speakers, smart earbuds, and diverse IoT products,” said Matt Crowley, the company’s CEO. “The problem is that speech consumes a lot of power, 1000µW or more, and that is why most voice-activated devices have to be plugged in.”

The microphone uses a variety of ‘unique’ materials that convert sound energy directly into electrical energy, then uses that sound energy to wake devices from sleep.

“It consumes a negligible

Above: According to research from Gartner, there could be as many as 500 smart devices in the average home by 2025. XMOS believes its development kit will help to create the voice interfaces needed to control them

amount of power and that’s a huge advantage when it comes to using devices like hearables, wearables, smart speakers, and event-detection systems. We’ve been able to extend battery life by 10 times,” Crowley claimed.

Voice processors

UK based XMOS, which supplies a number of voice solutions for IoT products, recently launched the XVF3000 family of voice processors, which provide far-field voice capture using arrays of MEMS microphones.

The company also took the opportunity to make available the VocalFusion Speaker development kit, which includes an XVF3000 processor card and a four-mic circular microphone array. The kit provides a quick way to start developing far-field voice capture applications.

“These are the first in a range of voice capture products from XMOS,” said Mark Lippett, CEO. “XVF voice processors will open up a host of new possibilities for designers looking to deliver high performance voice capture.”

XVF3000 devices include speech enhancement algorithms, along with an adaptive beamformer, which uses signals from four microphones to track a talker as they move. There is also high performance full-duplex, acoustic echo cancellation.

“Crucially, XVF3000 devices can be integrated with an applications processor or host PC via either USB, for data and control, or a combination of I2S and I2C. That means developers can quickly add custom voice and audio processing.”

It is becoming increasingly apparent that the possibilities for voice interfaces are growing dramatically, especially, when according to Gartner, there will be 500 smart devices in every home by 2025.

Because today’s solutions don’t scale, there will be a need for a new type of interface and voice looks like it could be the new click.

“Voice will open up a host of new possibilities for designers looking to deliver high performance voice capture.”
Mark Lippett