

Time to debottleneck your audio chain

New Infineon XENSIV™ MEMS microphone makes high-quality audio capturing possible with any device

Marcel Knecht, Marketing Manager MEMS Microphones, Infineon Technologies

The popularity of voice user interfaces and the usage of audio recording to share information and experiences are increasing dramatically. However, the poor performance of microphones often limits the potential of today's cutting-edge devices. Now with its recently launched XENSIV™ MEMS microphone IM69D130, Infineon introduces a new performance class for high-performance digital MEMS microphones that overcomes existing audio chain limitations. The IM69D130 is designed for applications where low self-noise (high SNR), wide dynamic range, low distortions and a high acoustic overload point are required, allowing highly accurate voice recognition where it was not possible before.

The market for micro electrical-mechanical system (MEMS) microphone technology has grown rapidly in recent years. They are primarily used in various applications such as mobile phones, cameras, security systems, digital voice assistants, robots, televisions, laptops, headphones, smart home appliances and automotive applications.

MEMS microphones offer many unique advantages compared to the traditional electret condenser microphone (ECM) technology. They are smaller in the same performance class as well as designed and trimmed to have matched amplitude and phase between multi microphone audio signals. Additionally, MEMS technology is resistant to high temperatures and reflow soldering, allowing an automated PCB assembly process.



Figure 1: High performance microphones are essential to many modern applications

Voice user interface for optimized speech recognition

Sound is captured for digital algorithm systems very differently to the way it is received by human ears, and therefore sound quality targets also differ. The signal does not necessarily need

to sound natural as long as it is optimized for the algorithms. Regardless of the use case, it is always important that the signal stays free of disturbances, distortion and noise.

Automatic speech recognition (ASR) is the process of automatically transcribing a speech signal into written words. ASR transcription accuracy is getting incredibly close to human levels, which is present-day at approximately 95 percent. So far with ASR, this level has been possible to achieve only in laboratories where the ambient conditions are more favorable.

The rationale behind the creation of any voice-based user interface should always be focused on making the experience reliable and simple for the end user. In order to deliver on this, system designers must consider real-life use cases such as how far away the user is likely to be from the microphone and how much background noise is expected to be present. Only then is it possible to design accordingly to for the best possible system performance.

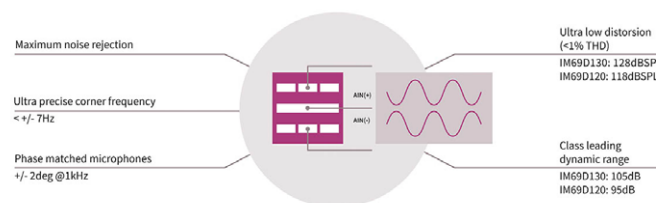


Figure 2: Infineon's dual-backplate technology is based on a miniature symmetrical design

Speech recognition in real-life environments, particularly when the speaker is at a distance, usually involves significant acoustic challenges such as background noise, reverberations, echo cancellation and microphone positioning. For this reason it is not enough to just have a good speech recognition engine. Every element in the system should be performing at the highest possible levels to prevent a quality bottleneck. The microphone's job is to provide the speech recognition system with the best possible input signal. High input signal quality helps the ASR system analyze the incoming sound and find the characteristics in it which enable the speech content to be recognized. Critical parameters for this include noise, distortion, frequency response, and phase.

Speech recognition systems in noisy environments can be significantly improved if the microphone technology has a high acoustic overload point (AOP). Sometimes the speech signal itself is not loud and there are other sounds causing interference. For example, speakers standing close to the microphones in speech controlled home entertainment systems and digital assistants which may output loud music or spoken information.

High AOP helps keep distortion low and improve the cancellation of noise and echoes. The larger the distance to the speech source is, the lower the signal to noise ratio of the signal

Glossary / Microphone basics

As a transducer, a microphone simply converts sound pressure waves into electrical signals. However, not all microphones offer the same performance and several parameters are key in determining suitability for an application.

Signal to noise ratio (SNR): Noise in the output of the microphone is simply any signals that are not from the intended input source. Noise can be in the environment, or generated within the microphone itself and the higher the noise level, the lower the audio signal quality. There are several different ways of expressing noise:

Self-Noise: this is the noise generated by the microphone itself when there is no sound signal present – normally expressed in V_{rms} , dBV or dBFS.

Equivalent Input Noise (EIN): this is the imaginary acoustic noise level that equates to the electrical noise level at the output of the microphone – expressed in dB SPL.

Signal to noise ratio (SNR): This is a measure of the self-noise of the microphone relative to the intended or desired input signal, expressed in dB.

Total Harmonic Distortion (THD): Microphones, in common with any other real world transducers are non-linear, meaning that they introduce distortion. In the case of distortion, the added content lies in the harmonics (typically the second through fifth). THD is the ratio of the energy in these harmonics, to the energy in the fundamental and is expressed as a percentage.

Acoustic Overload Point (AOP): Commonly defined as the point where THD exceeds 10%, although for more demanding applications, the AOP is sometimes expressed as the point where THD exceeds 1%.

being fed to the ASR algorithm will be. Therefore, microphone SNR should be higher when the intended capturing distance is greater.

Audio and video capturing and human-to-human communication quality can be also improved by excluding unwanted sounds from the signal. The goal is to increase SNR, which in this case is the ratio of the wanted sound signal to any unwanted ambient sounds, or noise. Noise cancellation and directionality can be achieved by using multiple microphones in combination with algorithms.

Directional microphone systems, such as beam forming, can concentrate the sensitivity of the microphones towards the desired direction and highlight the desired sound sources. Unwanted sounds can also be canceled based on parameters such as level differences between two microphones. Blind source separation is a more sophisticated noise reduction system. It enables canceling noise independent of orientation, distance, and location. All these noise cancellation methods benefit from the accuracy and quality of the signal they receive. The microphone should have high SNR, low distortion, flat frequency response, which also improves phase response, and low group delay.

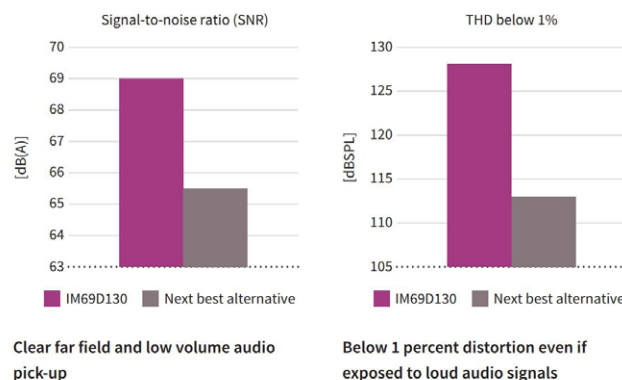


Figure 3: Clear audio signals even at high sound pressure levels with the new IM69D130

The benchmark for audio and voice pickup

Infineon's dual-backplate MEMS technology is based on a miniaturized symmetrical microphone design, similar to those utilized in studio condenser microphones. This design results in high linearity of the output signal within a dynamic range of 105 dB. The microphone noise floor is at 25 dB[A] (69 dB[A] SNR) and distortion does not exceed 1 percent, even at sound pressure levels of 128 dB SPL (AOP 130 dB SPL). This means that distortion free voice command capturing is possible even while music is being played from a loudspeaker.

The flat frequency response (28 Hz low-frequency roll-off) and tight manufacturing tolerance result in close phase matching of the microphones, which is vital for array applications.

With sensitivity (± 1 dB) and phase ($\pm 2^\circ$ @ 1 kHz) matching, the IM69D130 supports ultra-precise audio beam forming, allowing advanced audio and voice algorithms with high performance and far field voice user interfaces.

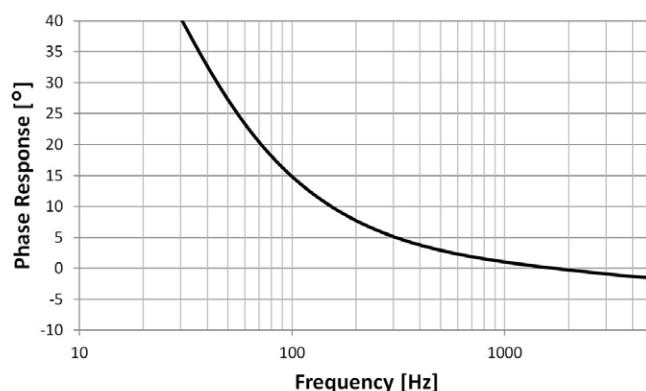


Figure 4: The IM69D130 sets a new benchmark for audio and voice pickup (IM69D130 typical phase response is depicted)

Its digital pulse density modulation interface means there is no need for analog components. This reduces the need for RF signal protection on the circuit board and the amount of required data lines in multi-microphone designs.

Combined with advanced audio signal processing algorithms, Infineon's IM69D130 provides a premium digital audio raw data signal that can handle even the most demanding voice recognition scenarios such as far field or whispered voice pick-up.