MARCH 2024

EE Times EUROPE

ALSO IN THIS ISSUE

COVER STORY:

Advancements in Generative AI Audio: The Crucial Role of High-SNR MEMS Microphones

PAGE 6

COVER STORY

Advancements in Generative AI Audio: The Crucial Role of High-SNR MEMS Microphones

BY SOFIA CUCALON, Product Manager for Consumer Sensors, and **JULIAN KORNPROBST**, MEMS Sensor System Competence Team Business Lead, Infineon Technologies

The integration of artificial intelligence has undoubtedly transformed our daily lives, with text and image-generating tools that can produce incredibly realistic content. However, AI's impact is felt beyond just visual and written media, with audio applications like speech-to-text (STT) and natural language processing (NLP) also benefiting from this technology. Can the new levels of quality in audio applications be attributed solely to the latest large language model–based AI generation? Or does hardware still play a vital role in these developments? Specifically, what is the contribution of high-signal-to-noise–ratio (SNR) microelectromechanical-systems (MEMS) microphones to this new quality of human-machine interfaces that are poised to change our daily lives? In this article, we will explore these questions and delve into the crucial role high-SNR MEMS microphones play in the development of cutting-edge audio applications like text-to-speech (TTS) and NLP.

According to Qualcomm's 2023 State of Sound Report,[1] the time spent wearing headphones per day is increasing. More and more people are working in public places like cafés and using headphones to block out background noise, either for peace and quiet or for meetings. In their spare time, people want to wear the same headphones to play games, listen to music or audiobooks or communicate with friends. Due to the longer wearing time, besides comfort, audio quality is becoming a key purchase criterion. The study found that an increasing number of people are interested in "premium audio features" when buying headphones, such as spatial audio, clear voice calls and lower audio latency. Seventy-three percent of respondents said that the sound quality of their devices should improve with each purchase—up from 67% the previous year.

Important audio features in consumer electronics as well as in cars are voice recognition and voice generation. For several years now, voice assistants including Siri and Alexa have been simplifying handling and enabling new applications, such as smart-home control via voice commands. Today, a wide variety of devices are equipped with integrated voice assistants, from smartphones (**Figure 1**) and headphones to smart TVs, smart speakers, smart-home units, laptops and tablets. Voice assistants are also increasingly being used in cars to control various features without the driver having to take their hands off the wheel. SAR predicts that the market for all devices with integrated voice assistants will grow to 3 billion units sold per year by 2028, with a CAGR of 5%.[2]

THE PROMISE OF AI IN AUDIO

However, current systems are still a long way from being perfect. Speech recognition still fails due to accents, linguistic imperfections or simple background noise. The voice output still sounds very technical and clearly differs from real voices.

This is where the latest generation of AI promises nothing short of a technical revolution, which will be felt in all human-machine interactions. The advantages of generative AI audio do not end with voice assistants and hence a better understanding of human intentions. Generating artificial voices that are virtually indistinguishable from real human voices enables better accessibility for the visually impaired, for example. It can improve the user experience on various digital platforms and offers new possibilities in the entertainment sector or customer support.

A key application of generative AI audio is speech-to-text, the conversion of spoken language into text. The use of AI enables high speed and accuracy. Together with its counterpart, text-to-speech, STT has many potential applications in consumer electron-

Figure 1: *Voice assistants integrated in smartphones benefit from increased speech recognition based on high-SNR MEMS microphones.*

ics, such as in laptops or smartphones—the integration of voice assistants but also the automatic transcription of meetings. In a meeting, AI-based applications can summarize who said what and what points were made, capturing the spirit of the discussion, and as the meeting progresses, you can check in on points made by different people and ensure that everyone's viewpoints are considered.

NLP AND EXPRESSIVE VOICE GENERATION

NLP is a basic building block for generative voice AI. The aim is to understand the meaning of spoken language, regardless of accents, colloquial expressions, blurred pronunciation and other differences between spoken and written language. Recognizing opinions and emotions based on the speed of speech, intonation and tone of voice is also part of NLP. Because human voices have a tremendous range, the audio recording for NLP must capture the pure voice as accurately as possible, with minimal background noise, chatter and other external influences. In other words, the microphones and signal processing contribute significantly to the quality of NLP.

For excellent speech recognition, the AI must be trained with recordings of as many different human voices as possible. Only then can it handle the subtleties of speech and understand the spoken text.

MEMS MICROPHONES FOR AUDIO AI

As for NLP, the audio AI can do its job optimally only if the appropriate hardware is deployed. Everything starts by converting sound waves generated by human speech into an electrical signal. The perfection of this conversion will affect the comprehension of the recorded signal. Any loss or degradation will have an impact on the accuracy of STT.

As the first component in the audio chain, microphones play a critical role when designing an audio AI device. MEMS microphones are unrivaled: They deliver high performance and low power consumption in a very small form factor and can therefore be easily integrated into a wider variety of devices.

MEMS microphones consist of three building blocks (**Figure 2**). First is the actual sensing element, the microelectromechanical system: Sound waves move a membrane, which forms a capacitor with the backplane. The resulting changes in capacitance generate the electrical signal. The second building block, the ASIC, contains the charge pump for the membrane, amplifier stages, a low-dropout

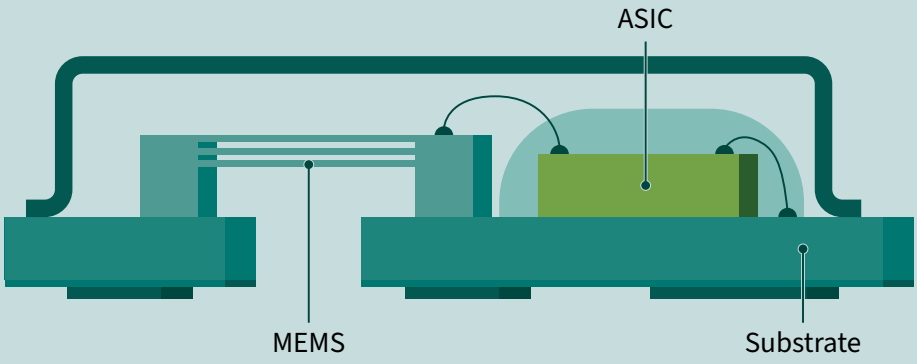

Figure 2: *Block diagram of a MEMS microphone.*

regulator (LDO) for clean power supply and the calibration logic. These parts are integrated into the third building block, the package. The package protects the component, shields it and forms an acoustical back volume.

To recognize the subtleties of speech, even under difficult conditions, such as background noise, accents or non-optimal distance of the speaker to the microphone, the key characteristic of the microphone is the SNR, which describes the difference between the inherent self-noise of the microphone and a standard reference signal. All elements of the microphone (MEMS, ASIC, package and sound ports) contribute to the self-noise.

ADVANTAGES OF XENSIV™ MEMS MICROPHONES FOR AUDIO AI

As mentioned above, audio AI devices require microphones with high SNR for accurate speech recognition. Infineon has a long track record in the development of high-performance MEMS microphones.[3] Sealed Dual Membrane (SDM) is Infineon's revolutionary MEMS microphone technology that utilizes two membranes and a charged stator to create a sealed low-pressure cavity (**Figure 3**) and a differential output signal. The architecture enables ultra-high SNR (up to 75 dB) and very low distortions and delivers high ingress protection (IP57) at a microphone level.

The XENSIV™ IM73A135 from Infineon thus achieves an SNR of 73 dB, one of the best values for a MEMS microphone in the industry, making it ideal for demanding applications, such as audio AI. A 4 × 3-mm² package allows miniaturization of the sound-capture unit and enables easy integration of voice AI technology into a wide range of devices, from laptops and conference phones to smart speakers and smartphones.

Another advantage of XENSIV™ MEMS microphones is their low energy consumption. With different operating modes to save energy, they contribute to the power efficiency of the final devices. As many of the devices with generative voice AI are portable and battery-powered, this is particularly important to achieve longer battery life.

Thanks to their compact size, cost-efficiency and low power consumption, multiple microphones can be used in one device. This allows background noise to be detected and reduced to enable better speech recognition. Beamforming algorithms can also be employed to isolate and capture specific speakers from background noise, again allowing for better voice recognition.

In a world that values improved audio quality, the advantages of MEMS microphones are also reflected in the market figures. The market for high-SNR MEMS microphones is growing significantly faster than for microphones with a lower SNR. For example, Omdia expects a CAGR of 8.7% in the consumer sector for MEMS microphones with an SNR above 64 dB, with unit sales of almost 3 billion by 2027.[4]

Infineon has been anticipating this trend for some time and is continuously working on increasingly high-performance MEMS microphones for audio AI applications, among others. In addition to the already remarkable 73-dB SNR, devices with higher SNR and even lower power consumption will follow soon.

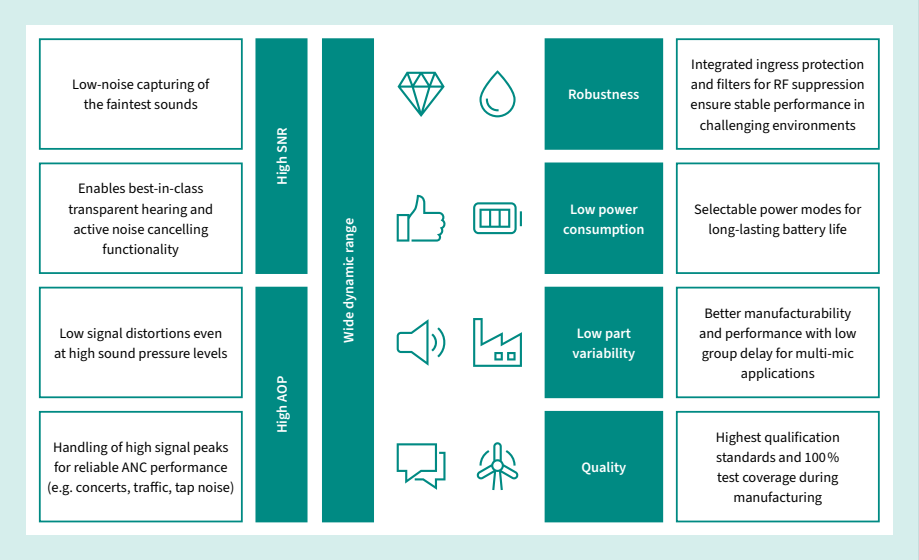

Figure 4: *Key value indicators of XENSIV™ MEMS microphones. Find out more at* www.infineon.com/mems.

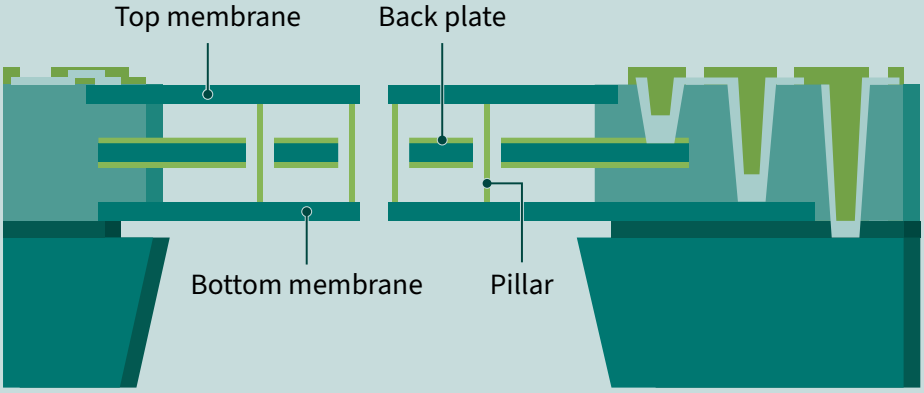

Figure 3: *SDM technology utilizes two membranes and a charged stator to create a sealed, low-pressure cavity and a differential output signal, resulting in ultra-high SNR and very low distortions.*

CONCLUSION

In the realm of generative AI audio, the integration of high-SNR MEMS microphones plays a pivotal role. As AI transforms audio applications like STT, MEMS microphones contribute by capturing nuanced voice data. This advancement enhances voice recognition, making it more natural and applicable in various domains, from consumer electronics to accessibility features for the visually impaired. With the advantages of excellent MEMS microphones, audio AI will open up further applications in the coming years, including voice cloning, emotion recognition and more.

Infineon Technologies develops and produces all the building blocks of MEMS microphones in-house. The company can easily identify the optimal combination of MEMS, ASIC and package to achieve the best possible performance for every application. This paves the way for improved user experiences and broader applications in the evolving landscape of voice AI. ■

REFERENCES

[1] Qualcomm Technologies Inc. (2023). "The 2023 State of Sound Report."
[2] SAR Insight & Consulting. Voice assistant platform forecasts, 2023.
[3] Infineon Technologies. *www.infineon.com/mems*
[4] Omdia. (2023). "MEMS Microphone Report."