From Noise to Nuance: MEMS Microphones and the Evolution of AI Understanding

by Dr. Gunar Lorenz, Infineon Technologies

In our technology-filled world, high-quality audio isn't a luxury – it's a necessity. From virtual assistants to video calls, the way we capture and transmit sound has a direct impact on our experience. But capturing high-quality sound isn't technically easy, especially in the noisy environments of our daily lives. Infineon's advances in MEMS microphones bring AI-powered devices closer to human-like interaction by capturing audio with greater clarity and detail.

MEMS, or micro-electro-mechanical systems, are miniature devices that combine tiny mechanical and electrical components on a single chip. From airbags in cars to sensors in smartphones – they're used in countless applications. MEMS microphones, specifically, are tiny microphones that are built in applications to pick up sound with impressive clarity despite their small size. You'll find them in every modern smartphone, earbud and smart speaker, enabling these devices to "hear" what's going on around them.

Now – with AI becoming more and more integrated into our daily lives – the quality of audio input provided by MEMS microphones will become even more important. This is because AI relies on the audio data it receives to make sense of human speech, emotion and intent. The clearer the audio, the better AI can understand us. Therefore, as we explore the future of AI, we must also consider how MEMS microphones will need to evolve to keep pace with the increasingly sophisticated AI technologies.

Why Voice Matters in AI's Evolution

When we as human beings speak, we convey much more than just the words themselves. Tone of voice, resonance space, and speaking behaviors such as volume, pace, and intonation make up a significant part of our communication and influence how messages are perceived. On average, it's estimated that only about 7% of communication is based on the actual words, while the other 93% is non-verbal, including tone of the voice (38%) and body language (55%)[1]. Our voices carry subtle cues about our emotions, intentions, and even our surroundings. For instance, a hurried, loud voice might indicate urgency, while a soft, steady tone might suggest calmness. Just as humans naturally interpret these cues, advanced AI aims to pick up on them to provide responses that feel intuitive and "human".

Imagine an AI assistant like Siri or a co-working machine that truly understands our needs, and not just the 7% of our communication containing words. For example, when we are angry, we may just want someone to listen and rant with us without giving unasked solutions. (Empathetic) humans can identify acoustic scenes like these through experience – but how would a machine "see" or feel this?

Now if we enable AI to "listen better", it will open the door to transform machines into a digital friend or digital colleague, or from a talking wiki to an effective and even personalized companion. For AI to reach this level, it must recognize vocal cues like emotion, tone, and urgency. For instance, if someone is calling from a noisy café or walking down a street, AI should be able to pick up on that background noise and adjust its response accordingly. The setting (e.g. an airport) can give hints on the intent of the question, just as "quick answer needed". This kind of understanding would allow AI to be more than just a helper – it could become a supportive presence, adapting responses to fit each situation.

One of the biggest challenges, however, is that AI often struggles to understand audio in real-life environments filled with background noise and interference. While MEMS microphones, the primary technology capturing sound in most consumer devices, are continuously improving, there's still a gap between human hearing and what these microphones can capture. Closing this gap is essential if we want AI to understand us as naturally as another human would.

The Art of Hearing: How Humans Perceive Sound

Human hearing is a remarkable result of evolution, enabling us to focus on a conversation in a noisy room or fully appreciate the richness of music. Our ears naturally tune into important sounds while filtering out distractions. For instance, we are particularly sensitive to specific frequencies, especially between 500 Hz and 6 kHz, where most speech information resides. Another important factor is the Sound Pressure Level (SPL), or simply 'volume' in layman's terms. SPL describes the amplitude of pressure variation caused by a sound wave, expressed in decibels (dB), relative to a reference pressure, typically the threshold of human hearing

However, even human hearing has its limits. Especially as we get older, our hearing gradually declines, often making it harder to hear high-pitched sounds or to understand speech in noisy environments. MEMS microphones face similar limitations. They can pick up a wide range of sounds but aren't yet on par with the human ear in capturing nuanced or quiet sounds, particularly in noisy places. When compared to human hearing, most currently used MEMS microphones perform at a level that's roughly comparable to the hearing abilities of a 60-year-old man[2], especially when it comes to detecting subtle details in speech.

Understanding how sound is perceived helps us recognize the challenge faced by current AI systems. If AI can only "hear" as well as an old person, it's bound to miss out on important information, such as emotional cues or context. If AI is to act as a true assistant—listening carefully, understanding intentions, and responding appropriately – it needs the best audio input possible. This is why future MEMS microphones need to capture sound with as much clarity and detail as possible to provide AI with high-quality data for interpretation.

Why MEMS Microphones Matter for Future AI

The human ear detects audio based on frequency and sound pressure level as outlined before, though its sensitivity declines with age. So a high-quality MEMS microphone must have a broad "dynamic range," meaning it can capture both very loud and very soft sounds without distortion. The quieter a microphone's self-noise, the clearer the captured sound will be. This is especially important for AI applications, as a clearer sound means more accurate interpretation by the AI system.

When talking about MEMS advancements, it is the "High Signal-to-Noise Ratio" (SNR) microphones that are essential for advanced AI-driven human-machine interaction, even if – so far – the latest microphones still cannot match the audio quality of the human ear. The SNR measures the microphone's ability to distinguish sounds from background noise, comparing the strength of the desired sound (e.g., a voice) to background noise. High SNR microphones therefore improve clarity, making them ideal for speech recognition tasks.

This way, high SNR microphones can particularly enhance various generative AI models used in short command recognition: In simpler tasks like so-called wake-word detection, where a specific word activates a device, like “Alexa” or “Hey Siri” – a high SNR provides a distinct signal, enabling quick response and reliable activation. For complex tasks, so-called “Large Language Models (LLMs)”, such as those powering voice assistants, can use linguistic context to interpret lower-quality audio. Trained on vast language data, they integrate text, audio, and visuals, leverage context, making speech-to-text more robust. These LLMs excel at recognizing intent, even when audio is imperfect. And finally, also Edge AI models (AI running on local devices, “on the edge”) benefit particularly from high SNR for clarity in command understanding.

Future of AI and MEMS: What’s Next?

Imagine AI assistants that do more than simply follow commands – they interpret tone, summarize meetings, and even prioritize follow-ups based on urgency. With advancements in MEMS microphone technology, such as Infineon’s, detecting emotions, analysing mood or linking spoken words in a conversation to a particular person (speaker diarization)[3] is increasingly possible. Improved audio input will enable AI to adapt responses, understand emotional context, and offer genuinely supportive interactions. Each innovation in MEMS technology brings us closer to AI that listens, understands, and empathizes naturally, making it an integral part of daily life and contributing to a brighter future. From telemedicine to virtual assistants in various fields – such as support for people with special needs, mental health applications, or education, including language learning – the potential is limitless.

Infineon’s microphone designs go beyond standard MEMS technology, featuring high SNR microphones with a focus on Single Backplate (SBP) and Sealed Dual Membrane (SDM) technologies to support a wide range of applications. Optimizing system design is key to capturing MEMS’ full potential.

Future AI applications powered by MEMS / high SNR microphones will revolutionize how devices interact with the world by enabling advanced environment recognition, speaker diarization3 and sound classification. Devices will filter and prioritize key information with unprecedented precision. For instance, smartphones might automatically adjust sound profiles based on music type, earbuds could detect urgent sirens, or smart glasses could identify anomalies in machinery sounds. In healthcare, AI could assist in monitoring patient conditions by detecting subtle changes in speech patterns or breathing und by understanding commands from doctors precisely, while in education, it could personalize learning experiences by adapting to students’ vocal tones or engagement levels. These advancements lay the foundation for AI that goes beyond simple responses. AI will perceive the nuances of our environment, understand context, and adapt to our needs with a level of engagement and empathy akin to human interaction, transforming many aspects of our lives and becoming an all-encompassing companion.

References

[1] online.utpb.edu/about-us/articles/communication/how-much-of-communication-is-nonverbal
[2] www.eetimes.com/ai-driven-interactions-powered-by-high-snr-mems-microphones
[3] en.wikipedia.org/wiki/Speaker_diarisation

Published by
Infineon Technologies AG
Am Campeon 1-15, 85579 Neubiberg
Germany

Public

Version: V1.0_EN
Date: 12/2024

Stay connected!

Scan QR code and explore offering
www.infineon.com