From hearing to understanding: MEMS microphones as a foundation of robotic perception

As robotics and AI continue to converge, the ability of robots to accurately perceive and interpret their acoustic environment becomes increasingly critical. High-performance MEMS (Micro-Electro-Mechanical Systems) microphones serve as the auditory foundation for advanced human-robot interaction, enabling machines to not just hear, but truly understand their surroundings.

Introduction

As robots leave structured environments, acoustic conditions become more complex: reverberation, overlapping conversations, machinery noise, and unpredictable background sounds.

High-performance microphones and microphone arrays are the foundation for enabling robots to operate reliably in these real world acoustic landscapes. On top spatial audio perception – made possible through multi microphone configurations – is increasingly seen as a safety critical feature.

As humanoid robots and robotic systems become more integrated into our daily lives, from service robots in healthcare to collaborative robots in manufacturing, their ability to accurately capture and process sound directly impacts their functionality and user experience. MEMS microphones with high SNR, wide bandwidth, and low distortion play a key role in enabling natural human machine interactions (HMI).

Why robotics need high quality audio

The integration of AI into robotic platforms has elevated audio capture from a peripheral function to a core system requirement. The shift from keyword-based command recognition to context-aware conversational AI places stringent new demands on MEMS microphone performance: accurately reproducing not only speech content, but the fine acoustic detail that encodes emotion, intent, and biometric identity.

AI relies on the audio data it receives to understand and interact with humans. The clearer the audio, the better AI can understand us and our environment. This is particularly critical as humanoid robots are expected to interact naturally with humans in complex, real-world situations where acoustic conditions are far from ideal.

Our human auditory system enables us to localize sounds in a 360° environment, identify speakers as well as to focus on a single conversation in the presence of surrounding noise.

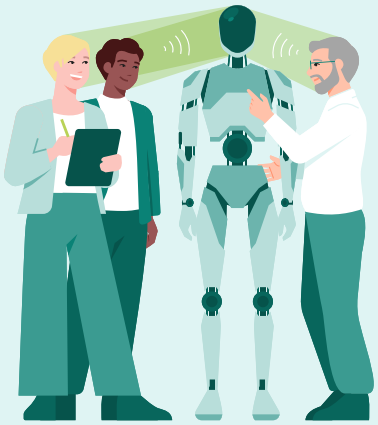
To equip humanoids and other robotic systems with similar capabilities in terms of audio perception, high performance microphones are needed. One of the most critical specifications for MEMS microphones is the Signal-to-Noise Ratio (SNR). SNR measures a microphone's ability to distinguish desired sounds – such as human speech, and more subtle audio cues – from its self-noise. In robotic systems high SNR microphones are essential for:

- **Advanced speech recognition:** Enabling robots to understand commands even in far field
- **Natural human-robot interaction:** Capturing the nuances of human communication
- **Speaker differentiation:** Allowing AI to more easily tell the difference between different speakers, reducing misunderstandings and improving the quality of interactions
- **Speaker identification and authorization:** Ensuring the AI performs sensitive commands only for authorized speakers (e.g., shopping, payments, security controls)

Although current MEMS microphones have not yet matched the signal-to-noise ratio of the human ear, they are a vital technology that enables AI-driven human-machine interaction. Better audio input is essential for making HMI suitable for everyday use. High SNR audio improves voice biometrics and anti-spoofing capabilities, minimizing false acceptances and rejections in noisy real-world settings, at a distance and in challenging acoustic environments.

Multi-microphone arrays: Enabling human-like acoustic orientation

To provide humanoids with the same level of acoustic understanding as humans, single microphone solutions are insufficient. Just as humans use two ears to localize sound sources and focus on specific speakers in noisy environments, humanoid robots require multi-microphone arrays to achieve comparable environmental understanding. This spatial awareness underpins safe navigation, natural interaction and situational understanding.



Value of beamforming in robotics applications

Multi-microphone arrays allow robotic systems to apply beamforming algorithms that spatially filter audio, enhancing target sounds while suppressing noise and interference. Following initial keyword detection via omnidirectional capture, adaptive beamforming enables:

- **Precise speaker and sound source localization:** Pinpointing the origin of speech or acoustic events in three-dimensional space
- **Focused listening:** Steering directional listening beams toward the target speaker while attenuating competing sound sources
- **Enhanced situational awareness:** Building a real-time spatial map of the acoustic environment – critical for safe, context-aware human-robot interaction

Infineon XENSIV™ Audio Shield: The platform for evaluation of XENSIV™ MEMS microphone arrays

Platforms such as Infineon's XENSIV™ Audio Shield simplify the evaluation of array geometries and enable developers to prototype beamforming and localization algorithms early in the design cycle.

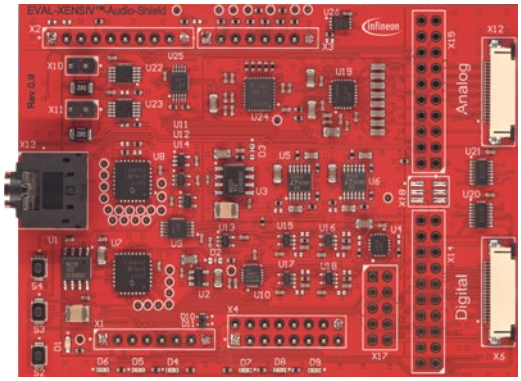


Figure 1 Top view XENISV™ Audio Shield

The XENSIV™ Audio Shield, shown in figure 1, enables comprehensive evaluation of all available XENSIV™ MEMS microphones in combination with PSOC™ 6 or ARDUINO™ platforms. It is compatible with all existing Infineon's XENSIV™ MEMS microphone flex kits and supports direct connection of up to 4 microphones on flex, with an option to interface up to 4 analog or 8 digital microphones in custom array designs.

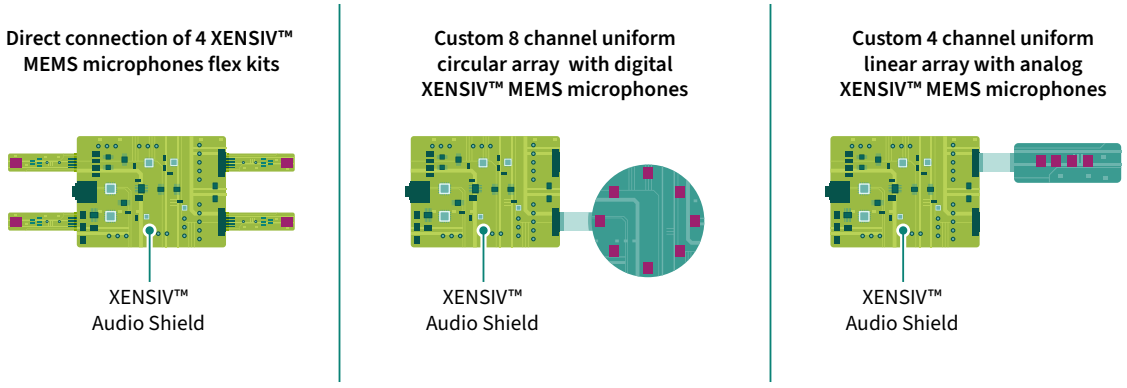


Table 1 Example array configurations using XENSIV™ Audio Shield

The bottom side of the PCB has 4 flex connectors for XENSIV™ MEMS microphone flex evaluation kits. They can be used for both analog and digital microphones.

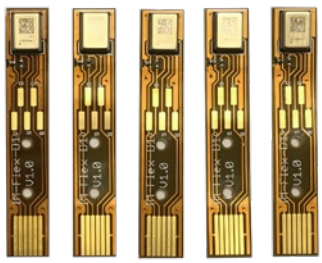


Figure 2 XENSIV™ MEMS microphone flex kit

The Infineon's XENSIV™ Audio Shield includes an Arduino interface, a high-performance audio codec as well as a headphone jack, expanding its flexibility for various audio development setups. It is also designed to work seamlessly with selected PSOC™ 6 Pioneer Kits, including the PSOC™ 62S2 Evaluation Kit (CY8CEVAL 062S2) and the PSOC™ 62S2 Wi-Fi Bluetooth® Pioneer Kit (CY8CKIT 062S2 43012). It is fully compatible with ModusToolbox™ and comes with entry-level code examples.

This new shield enables fast evaluation of microphone performance and supports development of advanced multi-microphone applications such as beamforming, machine learning, and audio algorithm design for robotic applications and beyond.

In combination with Infineon's high-performance XENSIV™ MEMS microphones like the IM72D128V01 this solution is the perfect development platform for spatial audio and robotic HMI applications.

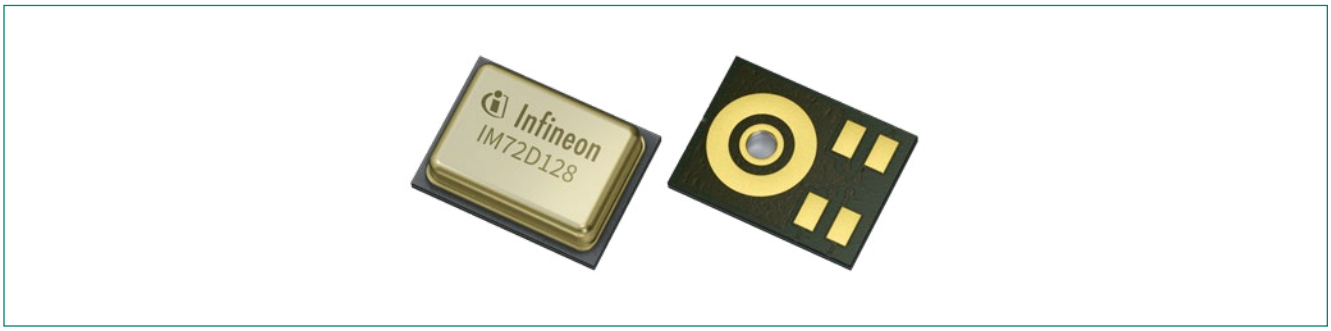


Figure 3 High-performance IM72D128 XENSIV™ MEMS microphone

Conclusion

High-performance MEMS microphones are not merely components – they are fundamental enablers of intelligent human-robot communication. Their compact form factor, combined with high SNR audio performance and scalability into multi-microphone arrays, makes them a critical building block for the next generation of humanoid and collaborative robotic platforms.

The ability to precisely localize speakers, isolate speech in multi-talker environments, and extract spatial audio information brings humanoid robots measurably closer to human-level acoustic perception. As robotic systems take on increasingly complex roles alongside humans, investment in superior MEMS microphone performance and advanced array architectures will be decisive – defining which robots can truly hear, understand, and seamlessly interact with the world around them.

Author

Manuela Heiss
Head of System Application Engineering MEMS Sensors at Infineon Technologies AG



Manuela Heiss is heading the system application engineering group for MEMS sensors at Infineon Technologies. Manuela joined Infineon in 2018 as a software engineer, followed by positions as system architect for IoT solutions at Infineon, Singapore. In 2022 Manuela established the system application engineering group for MEMS sensors at Infineon. Manuela and her team are focusing on enabling high performance acoustic solutions for consumer, industrial and automotive applications and thereby bridging the gap between technical innovation and customer-centric solutions.

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

© 2026 Infineon Technologies AG.
All rights reserved.

Public

Version: V1.0_EN
Date: 06/2026

Stay connected!

Scan QR code and explore offering
www.infineon.com

Important notice
Products are sold or provided and delivered by Infineon Technologies AG and its affiliates ("Infineon") subject to the terms and conditions of the frame supply contract or other written agreement(s) executed by a customer and Infineon or, in the absence of the foregoing, the applicable Sales Conditions of Infineon. General terms and conditions of a customer or deviations from applicable Sales Conditions of Infineon shall only be binding for Infineon if and to the extent Infineon has given its express written consent.

To the fullest extent permissible pursuant to applicable law, with respect to any information given in this document or in any associated documentation, Infineon disclaims all warranties and liabilities of any kind, whether express or implied, including but not limited to any warranties of merchantability, suitability of the products for the intended application or the specific use, or non-infringement of third-party rights.

Subject to the development and release of the products for series supply by Infineon, the technical specifications of the products are set forth in the relevant final datasheet provided by Infineon and, if any, agreed and signed specifications. Infineon's customers are required to evaluate the suitability of the products for the intended application or specific use.

The information given in this document is subject to change by Infineon at any time without notice.