SYNTIANT®

# MEMS Microphone SNR Impact on Voice AI Features in Consumer Electronics Devices

**Larry Chien,** Senior Application Engineer, Consumer MEMS Microphone
**Pokai Jen**, Senior Manager, Applications Engineering
**Nikolay Skovorodnikov**, Senior Manager, Applications Engineering

A MEMS microphone is a standard way to capture sound on a consumer electronic device. Voice input is becoming more and more adopted by users not only for communication but also for voice commands and interfaces across various devices and form-factors. This trend is particularly accelerated with the introduction of AI-enabled assistants on mobile, wearable and IOT devices. A growing number of devices rely entirely on voice input without any alternative method for entering commands, such as a keyboard. The value of high signal-to-noise (SNR) MEMS microphones is demonstrated in context of AI-enabled voice applications. Legacy metrics, such as recording quality and MOS (Mean Opinion Score), are compared to the performance achieved by AI voice transcription algorithms. It is demonstrated that high SNR MEMS microphone provides better experience across all metrics in the critical range of sound pressure levels commonly occurring in everyday use cases.

## Content



Syntiant Corp.
7555 Irvine Center Drive, Suite 200
Irvine, CA 92618

Phone: +1 949-774-4887
info@syntiant.com
www.syntiant.com

## Chapter 1. High SNR MEMS Microphone

SNR is one of the most important parameters for all electronic components, and the microphone is no exception. A high SNR microphone has a low self-noise level, which means the microphone can capture subtle nuances even at very low sound pressure. Theoretically, recording with a microphone with a self-noise (EIN) that is 15dB below the ambient noise level results in less than a 0.1dB increase in noise (Fig.1). For instance, to record in 38dBSPL environment one can choose a microphone with EIN (38-15=) 23dBSPL which is equivalent to (94-23=) 71dB SNR. **[1]**

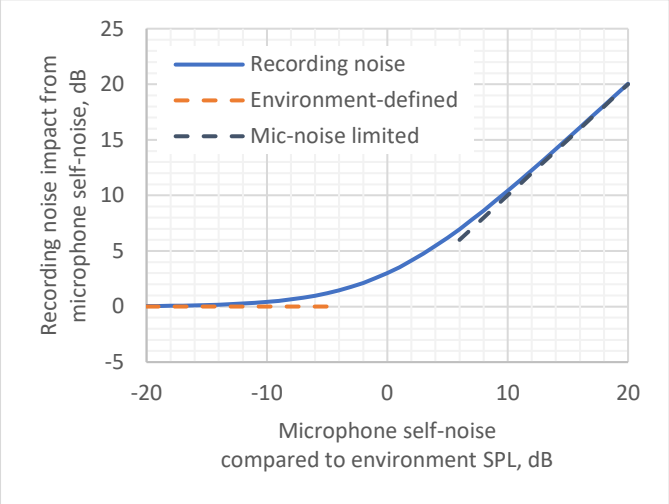


**Figure 1. Microphone contribution to recording noise.**

The intuitive benefit of High SNR microphone is the ability to capture sound without adding any noticeable noise to the recording. If SNR is insufficient, users would hear a hiss or hum in the background. Important reference point: if microphone self-noise equals background noise, total noise in the recording goes up by 3dB.

Nowadays, audio recording noise floor in consumer device does not equal MEMS microphone noise floor. Instead, sensor sets a starting point for system noise. Signal processing and integration features implemented in the device usually come with a noise penalty. Environmental barriers protecting microphone porthole from external contamination lead to SNR reduction. Convenient placement of the microphone inside the device requires long acoustic ducts/ports that can resonate in audible range and lead to noise increase. If an analog microphone is adopted in the design, additional ADC is required for signal conversion, which also adds noise. Signal processing features also may come with noise increases. The most common multi-mic technique is beamforming. This algorithm allows directionally selective sound pickup but also comes with a noise increase tradeoff, especially in low frequency range.

The goal of audio chain design in consumer electronic devices is to provide the best recording quality with all additional integration and signal processing features. Overall additional noise penalty from mic noise to total system noise can be as high as 10+dB. Choosing a high SNR mic with the lowest noise allows maximum flexibility in the additional audio features while maintaining superior audio quality.

In this study, data is presented based on single-mic recordings in a lab environment. In practical applications, however, the effective SNR of all microphone models is typically reduced due to the contribution of system noise.

Important note on microphone SNR specification: usually noise is measured from 20Hz to 20kHz unless mentioned otherwise. While this is a convenient metric allowing quick calculation and model comparison, it has drawbacks. For speech applications detecting and transcribing AI voice commands, it might be misleading to use 20kHz noise bandwidth as a metric because vast majority of speech energy is under 8kHz **[1]**. For precise analysis, it is important to take into consideration the spectral noise shape of the microphone and changes to it from integration and processing steps.

## Chapter 2. ViSQOL MOS Comparison

An Objective Listening Evaluation Tool can assess the quality of audio or speech recordings without relying on human evaluation. It is a fast and cost-effective method with more consistent results than Subjective Evaluation. There are several Evaluation Tools like PESQ (Perceptual Evaluation of Speech Quality), POLQA (Perceptual Objective Listening Quality Analysis) and ViSQOL (Visual Quality Objective Listening). ViSQOL was chosen because it is focused on capturing the subtleties of speech and predicts human perception of audio. The MOS rating ranges from 0-5. **[2]** In this article, microphones with different SNR specifications were used to record 20 Harvard sentences. **[3]** The volume of speech is ranges from normal speaking volume (65dBSPL) down to a whisper (30dBSPL). The recording files were not processed before the MOS calculation to focus on the impact of the microphone SNR. Based on the MOS result, high SNR microphone has a decisive advantage. (Fig.2)

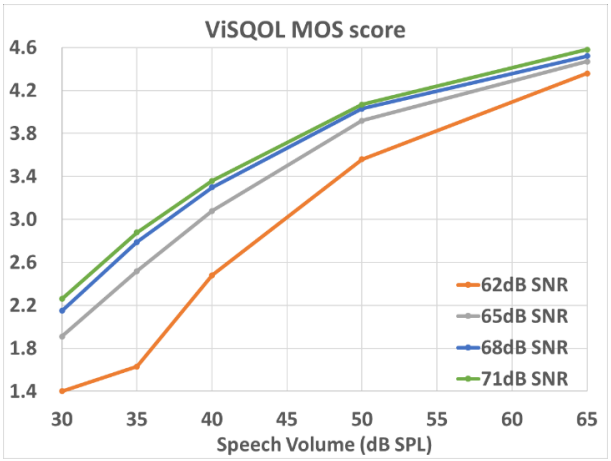


**Figure 2. ViSQOL MOS for Each Mic Model**

High and Low SNR microphones show significant differences especially at lower speech volumes. The score of 71 dB SNR mic is 35% higher than that of the 62 dB SNR mic at 40 dB SPL. At 35dB SPL, the 71 dB SNR mic score is 77% higher than the 62 dB SNR mic. Lower SNR microphone did not provide sufficient room between speech signal and its self-noise. It leads to muffled, unclear recording and fine details of speech are lost in the microphone noise floor. (Fig.3).

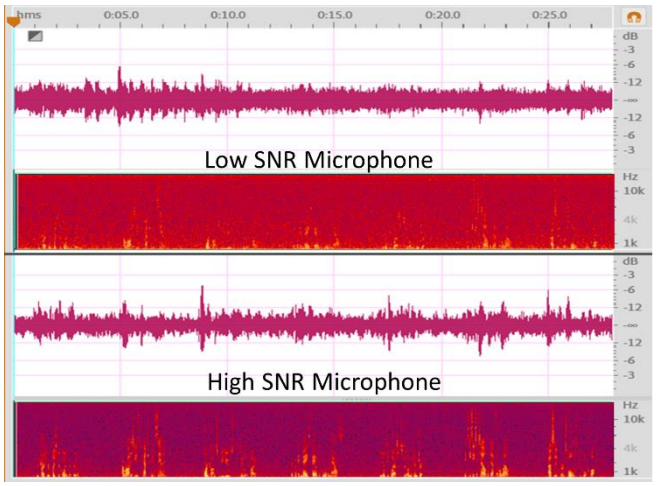


**Figure 3. Spectral Frequency Display at 35dB SPL**

## Chapter 3. AI application: Transcription

Any voice-controlled AI feature starts with a verbal command. The first task for the algorithm is to detect the command and correctly convert it into a text format. This process is called transcription. Next step, a text command is processed by a large language model (LLM) to produce an answer. Usually, the audio reply is generated based on the text output of an LLM. Transcription accuracy for various MEMS microphones is presented in this chapter.

Google Speech-to-Text (STT) is a cloud-based service that can transcribe speech from audio files. 20 Harvard sentences were recorded and uploaded to STT. The Speech-to-Text accuracy is calculated by sentences and words, respectively. See results for sentence recognition accuracy (Fig.4). At a speech volume of 65dBSPL, the microphone with High/Normal SNR can transcribe all the sentences perfectly while the Low SNR microphone misrecognizes one sentence. At 40dBSPL, the High SNR microphone still achieves 85% accuracy, but the Low-SNR mic is only at 60%. The most significant

difference is observed at 35dBSPL. In this case, the High SNR microphone maintains an 85% accuracy, but the Low SNR microphone's accuracy drops significantly to 25%. Based on word recognition results (Fig.5), the accuracy rate is more than 93% at all SPL circumstances if the microphone SNR is more than 68dB.

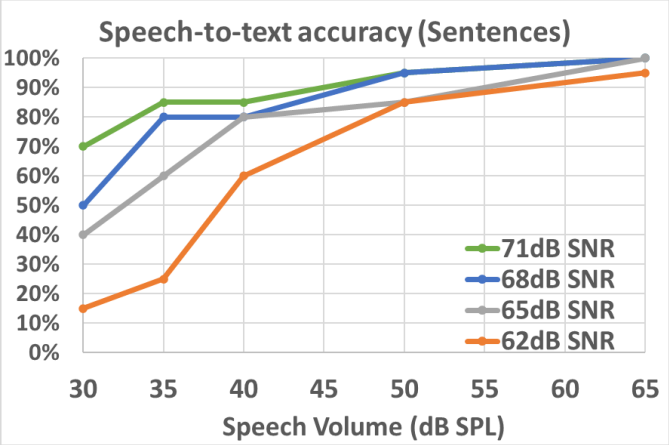


**Figure 4. Recognition Accuracy (20 Sentences)**

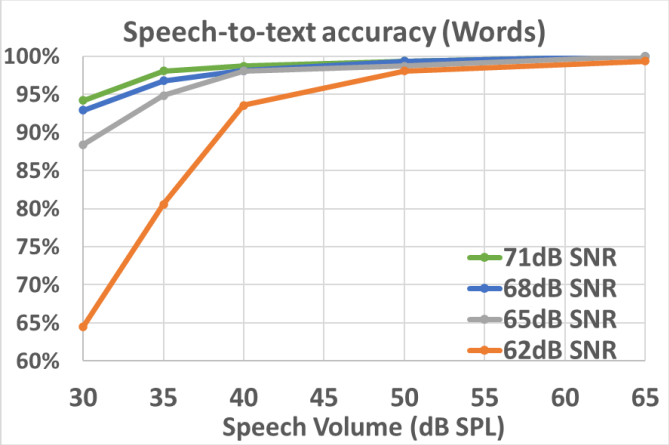


**Figure 5. Recognition Accuracy (Words)**

Comparison of overall words recognition accuracy, microphones with High SNR surpass Low SNR by 0.6% at 65dBSPL, 1.3% at 50dBSPL, 5.2% at 40dBSPL, 17.5% at 35dBSPL, and 29.7% at 30dBSPL (Fig.6). The 1% difference of the Word Error Rate (WER) can't be overlooked. For example, take a 10-minute speech with approximately 1,500 words. The speaker's volume fluctuates between 50-60dBSPL, so 10-20 words would fail to be transcribed with a Low SNR recording.

## Chapter 4. Device-level performance

The result presented in this study was collected in the lab using standalone MEMS microphones. The approximate threshold for AI algorithm failure and a rising word transcription error rate corresponds to the MEMS microphone noise floor (94dBSPL-SNR). For successful AI recognition, the output signal must be above the system's noise level. This ensures the algorithm can accurately process the input without interference.

Next, consider an equivalent scenario in a real consumer device. Additional features such as environmental protection and beamforming may attenuate the signal, effectively increasing the equivalent noise floor. In order to achieve similar voice recognition results, user would have to raise the voice and increase the SPL level. Figure 6 below illustrates the algorithm shift assuming a realistic 10dB noise penalty. With this signal level, the difference in speech recognition performance between mic models occurs becomes noticeable at SPL levels that are highly relevant to typical consumer use cases, such as comfortable soft speech levels. Imagine trying to use a voice assistant on a mobile phone in a living room or bedroom and having to repeat the command louder and louder until it is finally recognized. Adoption of high-SNR MEMS microphones can fix such an issue.

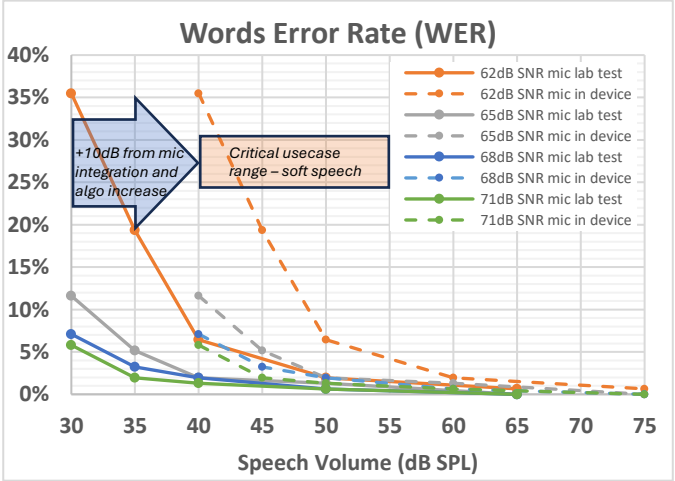


**Figure 6. Words Error Rate**

If the AI voice interface recognition rate is considered sufficient in target dBSPL range, high SNR can allow more flexibility with other audio features. Stronger environmental protection can be implemented to increase overall device robustness. Additionally, microphones can be located closer to each other and save space while preserving beamforming characteristics. This would require higher gains in the beamforming algorithm which adds more noise but high-SNR MEMS microphones can allow such design freedom.

## Chapter 5. Conclusion

Based on the test results with different SNR microphones, the high SNR microphone has a clear advantage for recording quality. We can not only observe the difference in the ViSQOL MOS but also hear an audible difference. Moreover, the accuracy rate of the speech recognition also shows the importance of High SNR microphone. The system's misrecognition of certain key words may lead to an interpretation that is completely different from the speaker's intent. Taking into consideration the increased system noise floor from additional integration and signal processing features, high-SNR MEMS microphones bring value for recording and AI-assistant use cases in the critical SPL range.

## References

**[1] Mikko Suvanto: The MEMS Microphone Book (2021)**

**[2] Measuring Speech Intelligibility and Perceived Audio Quality with STOI and ViSQOL- MATLAB Mathworks**

**[3] ETSI TECHNICAL SPECIFICATION 103 106 Standard**