AN547 - Why you need high performance, ultra-high SNR MEMS microphones

Table of contents

1   Abstract

The popularity of automatic speech recognition systems and the use of video content to share information and experiences is increasing dramatically. The performance and quality of the microphones used to capture sound must be high to ensure great user experiences. Critical factors include noise, distortion, frequency response and component matching. This application note concentrates on signal to noise ratio (SNR) and acoustic overload point (AOP) and explains the benefits of having high microphone performance in speech recognition and audio / video capturing systems.
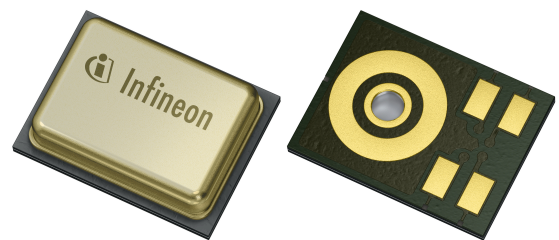
2 Signal to Noise Ratio (SNR)

Noise in the output of a microphone can be defined as any signals which are not the intended input source and is generally regarded to be an undesired element of the output signal. The higher the noise level, the more it reduces the audio signal quality. Noise can be external to the microphone or it can originate in the microphone itself. People usually hear microphone self-noise as a hiss that affects perceived sound quality. For algorithms, noise deteriorates the fidelity of the signal, thereby reducing system performance.

The noise of a microphone can be expressed in different ways:.

- **Self-noise** (Vrms, dBV, dBFS) is the rms noise voltage generated by the microphone itself when it is not excited by an external sound.
- **Signal to Noise Ratio**, **SNR** (dB), describes the self-noise of the microphone relative to the intended input signal. SNR is usually measured using a standardized acoustic input signal to represent the wanted sound, a 94 dBSPL (1 Pa) sine wave.
- **Equivalent Input Noise**, **EIN** (dBSPL), is the (imaginary) acoustic noise level coming into the microphone that is equivalent to the electrical noise level at the output of the microphone.

Electrical SNR (dBV)

- *Electrical output – Self-noise*

Acoustical SNR (dBSPL)

- *Acoustical input – EIN*

3 Acoustic Overload Point (AOP)

All real-life audio transducers are non-linear systems in that they add content to the signal that passes through them. In the case of distortion, the added content lies at the harmonics of the frequencies that are present in the original signal. Distortion is typically measured as **Total Harmonic Distortion**, **THD** (THD+N if self-noise is included). It is the ratio of the energy in the signal harmonics (typically second through fifth) to the energy in the fundamental frequency when the microphone is excited by a sine wave. The test signal is typically a 1 kHz sine signal at a relatively high sound pressure level (SPL), often 94 dBSPL or higher. THD is given as a percentage (%).

Acoustic Overload Point, **AOP** is commonly defined as the sound pressure level at which the THD exceeds 10%. The unit of AOP is dBSPL.

In most cases it is beneficial and important to preserve the original form and content of the sound incoming to the microphone(s). Adding content, such as distortion, to the original signal is likely to sound unpleasant to the person listening to the captured sound. The more added energy there is (i.e. the higher the THD), the worse the perceived audio quality will be. Distortion is also likely to confuse algorithms such as speech recognition system which carry out very detailed analysis of the contents of the incoming signal.

3 Acoustic Overload Point (AOP)
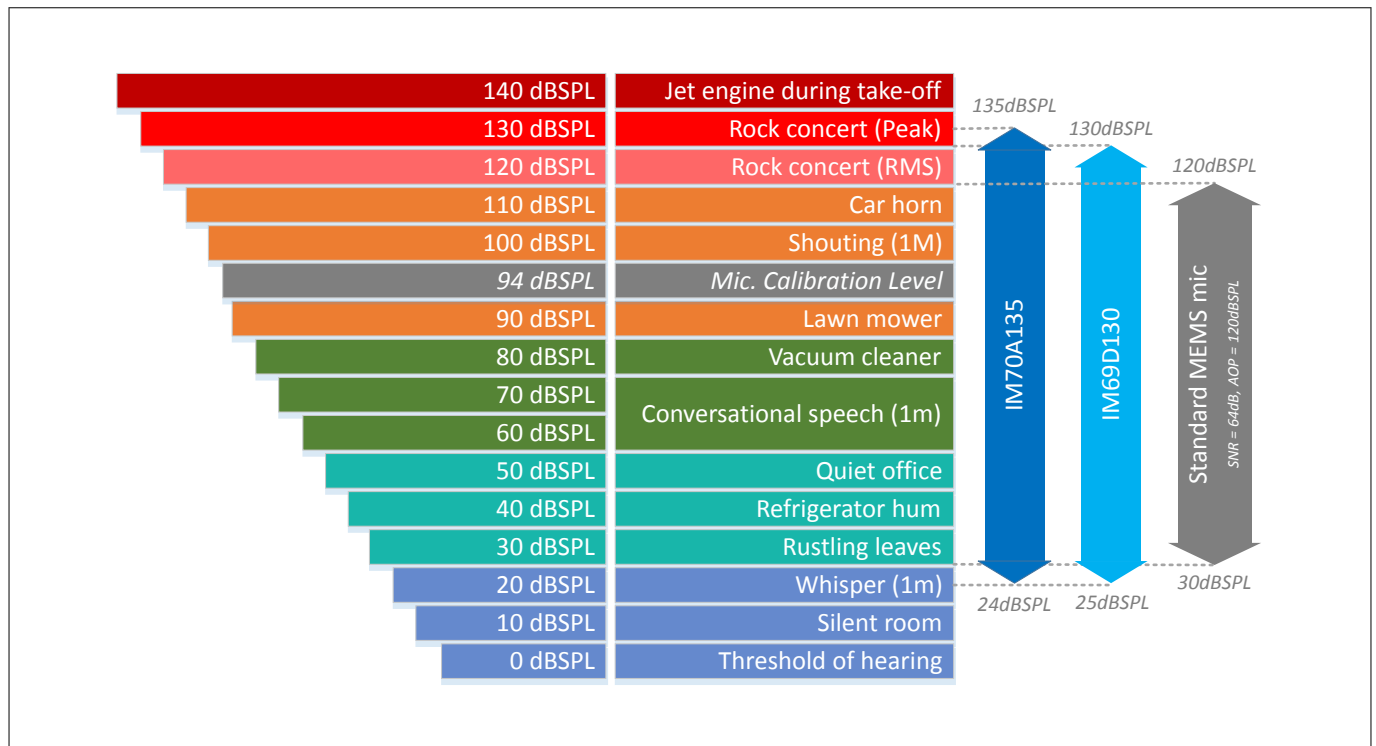
Figure 1   Dynamic range of IM69D130 and IM70A135 vs standard MEMS microphones

4 Importance of microphone performance for recordings

4.1 Signal to Noise Ratio importance for recordings

The goal of audio / video recording is to capture the incoming sound from the subject and to reproduce it in the output of the microphone system. When the recording is intended for human ears, it is desirable for the electrical output signal to match the acoustic signal as closely as possible, providing a "natural" sounding recording. The microphone and its SNR are critical parts of the sound capturing signal chain which affects the quality of audio recordings. Some typical use cases are presented in the table below.

Use Case	Details and Challenges
Home video	Typically the home is a quiet environment where microphone noise can easily become dominant. Varying capturing and playback conditions and equipment.
Children	Filmed objects are mobile and have soft (quiet) voices.
Social media	High video quality requirements to maximize viewer engagement.
Professional videos	Job applications, job interviews, talent introductions, presentations, etc. High video quality is crucial to differentiate an applicant or business from others.
Music	High sound quality is important to ensure a natural sounding recording. Varying capturing and playback conditions are challenging.
Performances	E.g. school plays can be challenging: quiet voices, long distances, ambient noise.
Nature	Recorded sounds can be at low or very low sound pressure levels.
Surveillance	The captured sounds can be quiet and coming in from long distances.

In free field, sound pressure halves (reduces by 6 dB) for every doubling of distance. The further the captured sound source is, the quieter the acoustic signal that reaches the microphone. As the self-noise of a microphone is practically constant, a reduction in incoming signal level causes a reduction in the SNR of the output signal of the microphone. Typically, a weak signal has to be amplified to bring it up to an appropriate level for the device signal path. Amplifying the signal also amplifies the noise present in the output. The more amplification there is, the higher the risk is that the noise will rise to a level at which it degrades the quality of the captured signal significantly.

A high microphone SNR helps keep the noise floor inaudible even when the signal is amplified. The longer the capturing distance, the lower the microphone self-noise should be to avoid problems. This is especially critical when the distance is long and the sound source itself is quiet. As sound pressure attenuates by 6 dB per doubling of distance, using a microphone with a 6 dB higher SNR can enable doubling the capturing distance without degradation in signal quality.

POLQA (Perceptual Objective Listening Quality Assessment) is an ITU-T standard model that uses digital speech analysis to objectively determine the quality and intelligibility of a recorded speech signal. Microphones with high SNR perform clearly better in POLQA tests and result in superior speech intelligibility. Signals of the same level are more intelligible when recorded with a higher SNR microphone.

Playback conditions and video picture quality affect the perceived noise level.

- *Ambient noise level in the playback environment*
- *Playback volume*
- *Quality of listening equipment (e.g. noise and frequency response)*
- *High video quality demands high sound quality to avoid degrading the overall audio / video quality*

4.2 Acoustic Overload Point importance for recordings

Just like SNR, AOP is an important audio / video quality factor. Distortion can very easily render a video recording useless. There are many smartphone videos online which have been shot in pop/rock concerts and are unwatchable due to badly distorted audio. High AOP improves sound quality if the incoming sound pressure level of the intended sound (or of disturbances) is high or very high. High AOP helps a microphone system handle very high signal peaks that may appear in the incoming acoustic signal even if the average sound pressure level is not very high. See some typical use cases in the table below.

Use Case	Details and Challenges
Pop/rock music concerts	Concerts are typically loud. High sound quality is a key enabler for good and natural sounding performance recordings.
Sports events	Either the sport (e.g. motorsports) or the crowd (e.g. ice hockey arena) is very loud.
Traffic	Lots of low frequency noise.
Wind	Wind is a common cause for poor sound quality in audio / video recordings shot outdoors. High AOP can help with certain kinds of wind conditions.

Up until a few years ago the standard level for consumer electronics device microphone AOP was between 110 and 120 dBSPL. In the recent past, the requirements for AOP have moved up. In order to ensure sound quality and speech recognition performance which satisfy customers, a device designer should choose significantly better microphones that have AOPs closer to the 130 dBSPL mark, or higher.

At lower sound pressure levels, it makes more sense to look at lower THD levels than the 10% specified for AOP. In addition to having high AOP, it is also important that the THD stays low, below 2%, up to high enough sound pressure levels for the intended applications (for example, up to 120 dBSPL).

5 Importance of microphone performance for speech recognition

In the case of systems where the captured sound is intended for algorithms, the sound quality goals may be different to when the signal is for human ears. The signal does not necessarily have 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 clean of disturbances, artifacts, distortion and noise.

Automatic speech recognition (ASR) is the task of automatically transcribing a speech signal into written words. Transcription accuracies are getting closer to the human level, which is at approximately 95%. However, so far achieving this level has been possible only in laboratories where the ambient conditions are favorable.

Speech recognition in real-life environments and at a distance involves some significant acoustic challenges such as background noise, reverberations, echo cancellation and microphone positioning. It is not enough to just have a good speech recognition engine. Every element in the system should be performing at a high standard 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 that enable recognizing the speech content. Critical parameters are noise, distortion, frequency response and phase.

High AOP can help speech recognition systems in loud environments. Sometimes the speech signal itself is not loud but there are other disturbances present. For example, there are speakers 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 longer the distance to the speech source, the lower the signal to noise ratio of the signal being fed to the ASR algorithm. Therefore, microphone SNR should be the higher when the intended capturing distance is longer.

6 Importance of microphone performance for noise cancellation algorithms

A key function for speech recognition systems is being able to ignore the sounds and noises which are not the speech to be transcribed. Audio / video capturing and human-to-human communication quality can also be 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 the unwanted ambient sounds (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 high quality of the signal they receive. The microphone should have high SNR, low distortion, flat frequency response (also improves phase response) and low group delay.

In order to optimize the functionality of noise cancellation algorithms, the microphones used in the system should have identical properties. The role of microphone to microphone matching is critical. The less variance there is in sensitivity, phase behavior and latency from microphone to microphone, the better.

7 Summary

From 2005 to 2015 the SNRs of state-of-the-art microphones in mass market consumer electronics devices improved from below 60 dB up to about 65 dB. With the requirements set by new high-performance speech recognition systems and other capturing use cases, even 65 dB is no longer enough. Current high-end microphones are approaching 70 dB SNR.

High microphone performance is a key enabler for high speech recognition and audio capturing quality. The performance of technologies such as automatic speech recognition algorithms and cameras are improving rapidly and the user experience expectations of device buyers are rising. It is important to avoid microphones becoming improvement bottlenecks.

Luckily there are high performance microphones available. Noise performance has improved significantly in the last few years. SNR is rising beyond the 70 dB level and quality degrading distortion is becoming a thing of the past with AOP reaching the 130 dBSPL mark. This level of microphone performance helps devices give satisfying user experiences to even the most demanding customers.

7.1   List of abbreviations

SNR: signal to noise ratio

EIN: equivalent input noise

THD: total harmonic distortion

AOP: acoustic overload point

ASR: automatic speech recognition

SPL: sound pressure level

dB: decibel

dB(A): decibel, A-weighted

dBV: decibels relative to 1 volt

dBSPL: decibels, sound pressure level

Pa: Pascal, unit of pressure

CE: consumer electronics

Trademarks

All referenced product or service names and trademarks are the property of their respective owners.

IMPORTANT NOTICE

The information contained in this application note is given as a hint for the implementation of the product only and shall in no event be regarded as a description or warranty of a certain functionality, condition or quality of the product. Before implementation of the product, the recipient of this application note must verify any function and other technical information given herein in the real application. Infineon Technologies hereby disclaims any and all warranties and liabilities of any kind (including without limitation warranties of non-infringement of intellectual property rights of any third party) with respect to any and all information given in this application note.

The data contained in this document is exclusively intended for technically trained staff. It is the responsibility of customer's technical departments to evaluate the suitability of the product for the intended application and the completeness of the product information given in this document with respect to such application.

WARNINGS

Due to technical requirements products may contain dangerous substances. For information on the types in question please contact your nearest Infineon Technologies office.

Except as otherwise explicitly approved by Infineon Technologies in a written document signed by authorized representatives of Infineon Technologies, Infineon Technologies' products may not be used in any applications where a failure of the product or any consequences of the use thereof can reasonably be expected to result in personal injury