Advantages of Vesper's Digital Microphone in Voice Interface Design

White Paper
Rev1.0
Author: Udaynag Pisipati

Vesper Technologies Inc.
77 Summer St, 8th floor
Boston, MA - 02110

WP4-Rev1.0
10/1/2019

Table of Contents

Introduction

As the smart device wars rage, many sensor companies aim to build the best microphone arrays for voice activated devices. With the proliferation of devices taking the consumer market by storm, such players are in high demand. Advanced sensor companies are proving to be the innovators in the space, introducing novel technologies to extend battery life, improve far-field audio quality and withstand environmental contaminants.

Advanced sensor company, Vesper, is one such company. In fact, its newest microphone, the VM3000, is a Pulse Density Modulation (PDM) output microphone with an in-built Analog to Digital Converter (ADC) that converts the incoming analog signal captured by the MEMS sensor into digital domain. Compared to analog microphones that require an external ADC, digital microphones provide savings in Bill of Materials (BOM), design footprint as well as higher immunity to noise from radio frequency and electromagnetic interference. While the PDM output is the most widely used digital interface for MEMS microphones in the market today, Vesper's VM3000 offers unique differentiation for embedded voice interfaces compared to its capacitive counterparts. How does the PDM microphone's piezoelectric MEMS construction translate to system level performance advantages for voice interface designs? This white paper aims to explore this question with technical insights gathered from performance studies at Vesper.

Piezo Advantages

Robustness

Let's begin with the robustness of the VM3000 microphone. Built on the same piezoelectric MEMS construction as other analog microphones from Vesper, VM3000 uses a single layer piezoelectric crystal, offering robustness to dust, water, moisture and other environmental particles. In a capacitive MEMS microphone with a dual plate design, any foreign particles entering the MEMS sensor can get trapped in between the plates, therefore degrading the sensitivity and frequency response over time. As the individual microphones in the array drift in performance, beamforming algorithms perceive this difference as directionality cues steering the beam away from the expected orientation. However, the single layer durable construction of Vesper mics makes them immune to these sensitivity drifts and provides stable arrays without the need for a mesh. Figure 1 shows the frequency response comparison between the two different MEMS microphones before and after dust exposure measured at an independent test facility.

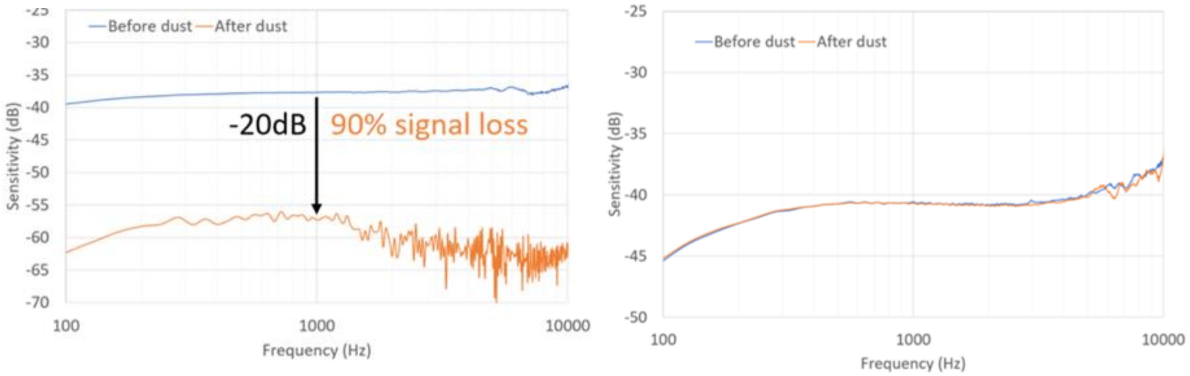

Figure 1: Frequency Response Measurement of Capacitive (Left) and Vesper (Right) microphone pre- and post-dust exposure, according to JDEC standard

One might ask, why is this crucial now since capacitive MEMS microphones have existed for years? The answer lies in the increased adaption of microphone arrays in voice activation devices such as smart speakers, security cameras, headsets, smartphones, etc. Now that both analog and digital versions of robust piezo microphones are available, developers can be assured of the long-term stability of their arrays irrespective of the interface available on their chipsets. For arrays that require multiple microphones with a protective mesh or membrane, VM3000 turns out to be a boon given Vesper products with IP57 rating do not require a mesh/membrane, enabling simpler design, unaltered performance as well as BOM savings.

Startup Time

Like other piezo microphones, VM3000 does not require a charge pump to generate the bias voltage. Therefore, VM3000 can wake up from sleep within 200 µsec as well as switch between the different power modes within 100 μsec in contrast to the 10-millisecond startup as well as mode switch latency of capacitive microphones. This 50x advantage in startup time allows voice activation systems to wake up fast enough to capture complete wake words. Slow startup of capacitive microphones limits the overall system wakeup time, impacting the wake word detection performance. Ultra-fast startup time advantage is more prominent when used in combination with Vesper's Zero Power Listening microphone, VM1010. For a visual interpretation of startup time, consider a simplified block diagram of the voice activation system as shown in Figure 2. In this design, VM3000 is combined with VM1010, an analog microphone that wakes up from ultra-low power sleep mode only when there is any sound activity exceeding a pre-defined sound pressure level. This wakeup is controlled by the processor using 2 GPIO control pins – Dout and Mode. In summary, when there is a sound activity in the environment, VM1010 switches from sleep mode to normal mode and outputs a HIGH on Dout. The Dout signal then drives the clock on the DSP/Codec, which in turn sets the CLK on VM3000 to wake up the microphone from sleep.

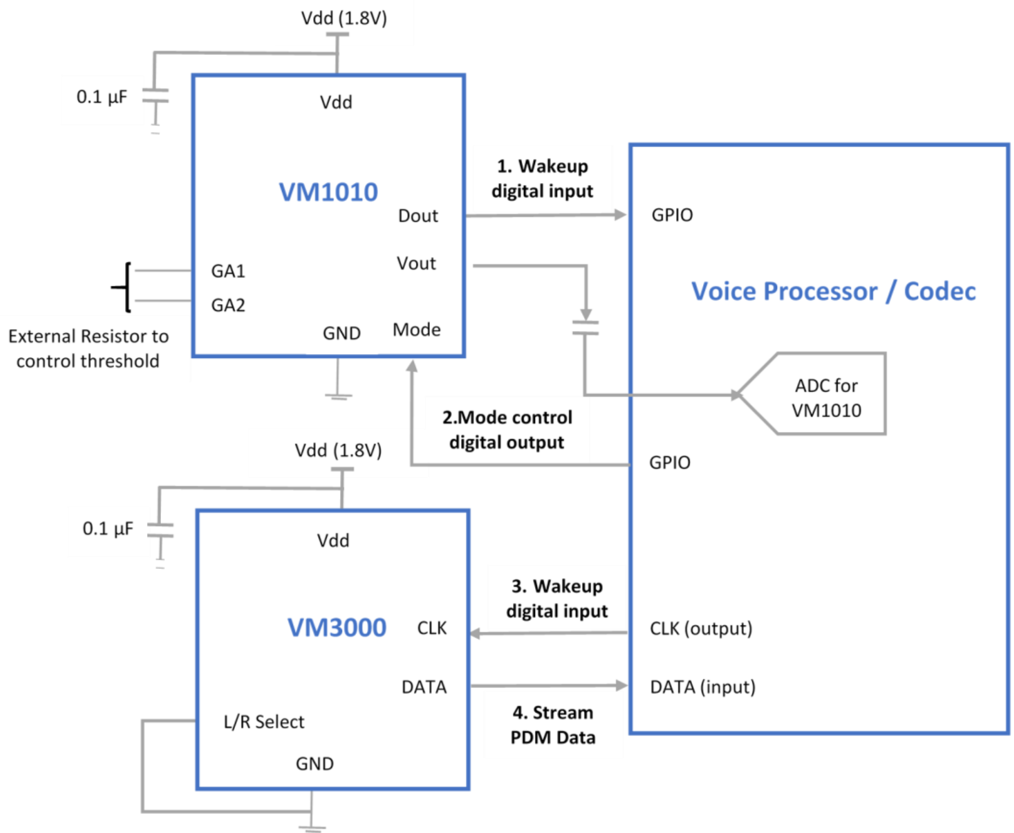

Figure 2: Simplified voice activation system with Vesper's ZPL microphone combined with VM3000 digital microphone

The above system is also modified to measure startup time on both VM3000 and a capacitive microphone with comparable acoustic specification. Figure 3 and Figure 4 shows the state transition of Mode/Dout pins on VM1010 along with CLK and PDM data pins on a digital microphone in each case. As the VM1010 hears the wake word, Dout goes from low to high and the processor sets the mode pin within 8 µsec. The PDM clock signal starts at 16 µsec (shown in Figure 3) to switch the digital mic out of sleep mode and the PDM data is available on the DATA pin within 200 µsec as shown in Figure 4.

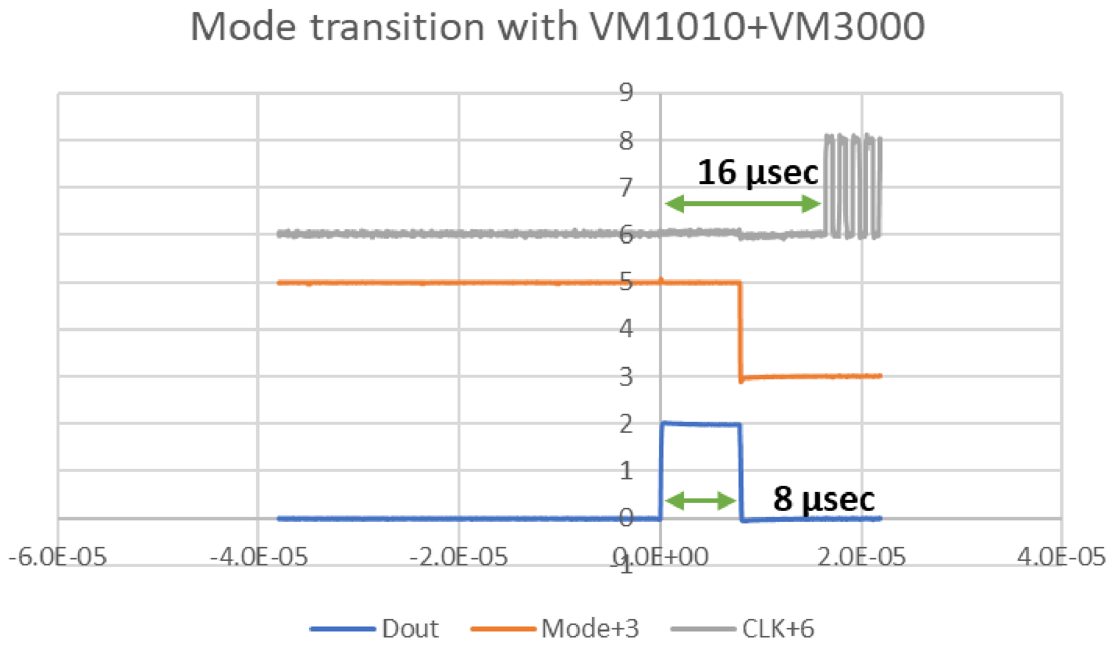

Figure 3: Mode transition with VM1010 and a digital microphone (shown separately for granularity)

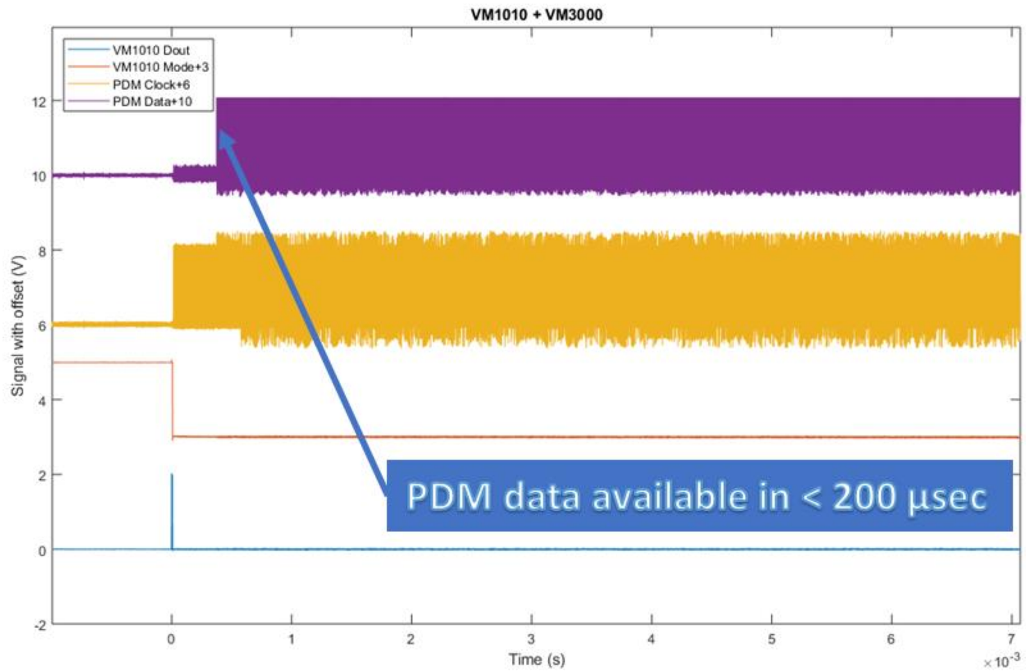

Figure 4: VM1010 + VM3000 timing as measured on an oscilloscope

Figure 5 shows that the PDM data on capacitive microphone is only available 10 milliseconds after the CLK goes high. Therefore, the DSP/Codec must wait for this duration to receive the PDM data and perform additional wake word processing. From an implementation perspective, this means that the DSP must keep the capacitive microphone in one of the low power modes instead of switching back to standby mode (50 µA typical current), since it cannot wake up fast enough to serve a wake word detection request. In contrast, Vesper's system can remain in sleep mode for extended periods with just under 1 µA current, while at the same time switching to high performance mode in less than 100 µsec as needed.

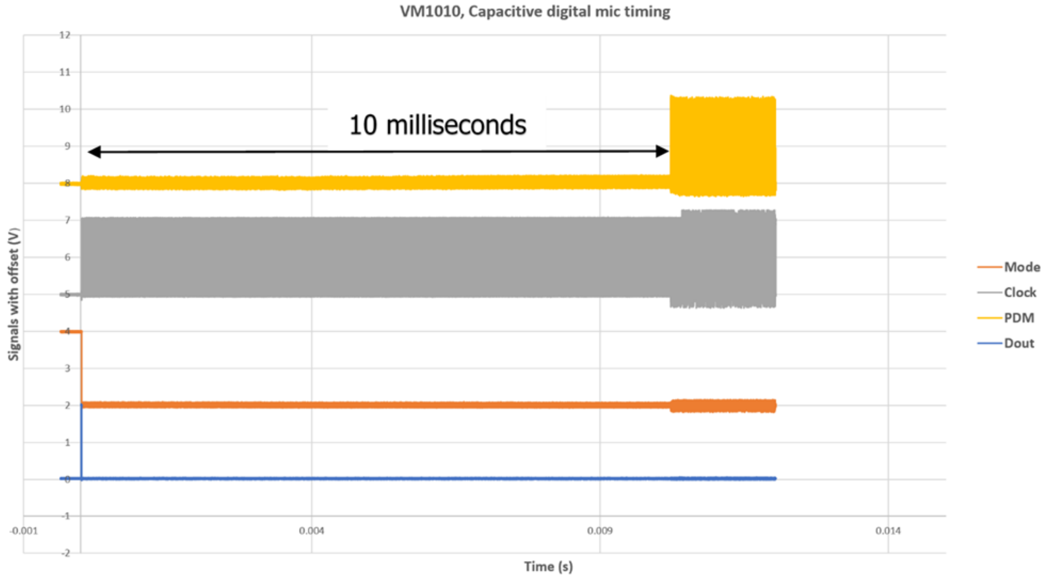

Figure 5: VM1010 and capacitive MEMS digital microphone timing

Ultra-low Power Sleep mode

Digital microphones also operate in multiple modes where a Codec or an application processor can define the state of the microphone using the clock frequency. VM3000 offers four such power modes – sleep(<1μA), standby (145μA), low power (400μA) and normal (700μA) mode, depending on the clock frequency of operation. Figure 6 shows the state diagram of the microphone.

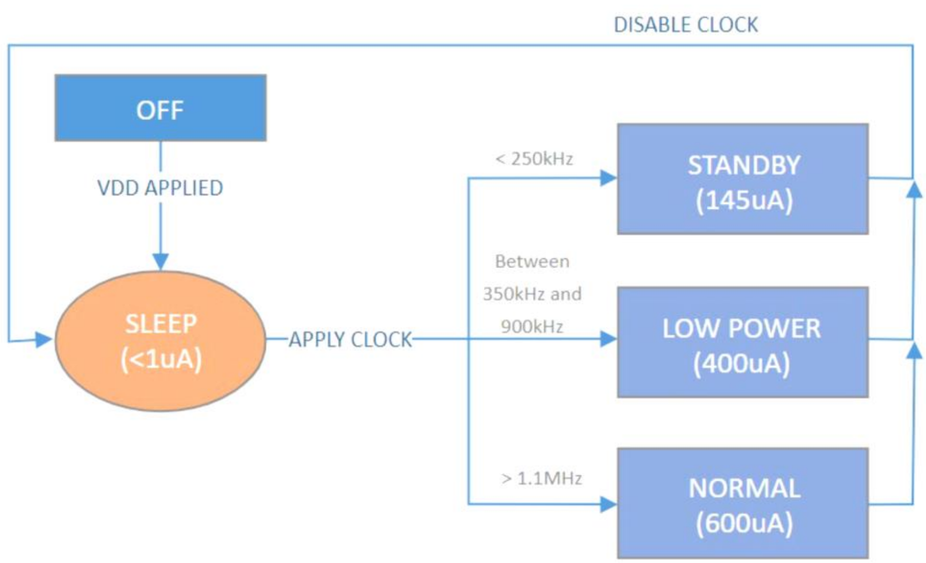

Figure 6: State diagram of Vesper VM3000 digital MEMS microphone

It is worth noting that the ultra-low current consumption of 0.35 μA on VM3000 is the best deep sleep mode current available on the current MEMS microphones. A capacitive MEMS microphone requires a charge pump to power up the microphone which itself consumes a significant amount of power in sleep mode. Piezo microphones, on the other hand do not require a charge pump, offering the best in class deep sleep power consumption together with ultra-fast mode switching time. This ultra-low power sleep mode together with fast mode switching time allows the microphone to be used in ultra-low power sleep mode and still not lose any audio information when the device is woken up.

System Level Advantages

Effective System SNR

The effective SNR available for the audio signal processing algorithms is more critical than the component level SNR specification of a microphone. Many designs with capacitive MEMS microphones require the use of a mesh/membrane to achieve water ingress protection up to 1-meter immersion, also called IPx7 rating. A mesh/membrane is typically attached to the acoustic port as a protective element from environmental contaminants in the designs. The downside of using a mesh/membrane is the drop-in sensitivity of the MEMS microphone which results in the degradation of effective Signal to Noise Ratio (SNR) of the system. How much does an acoustic mesh/membrane impact the Signal to Noise Ratio (SNR) of a system built with capacitive MEMS microphones? How does the robustness advantage of Vesper's piezo mics translate into acoustic performance? These are the most common and intriguing questions posed by our customers since our Piezoelectric MEMS microphones first hit the market a few years ago. To answer this question, let us consider the component level specs of capacitive MEMS microphone with Vesper's latest digital microphone, VM3000 shown in the Table below.

Metric	VM3000	Capacitive MEMS mic
Microphone SNR	62.5 dB	65 dB
Minimum attenuation from mesh assembly	N/A	4 dB
Effective SNR	62.5 dB	61 dB
Performance over time	Consistent sensitivity	Mesh degrades performance over time

Table assumes a minimum attenuation of 4 dB from the mesh assembly which includes the typical insertion loss of the mesh and the surrounding components such as gaskets that hold the mesh in place. Comparison shows that Vesper microphones provide equivalent system SNR due to their inherent robustness compared to a capacitive MEMS microphone with mesh attached.

Our analysis above is also proven up by a design study conducted by a team of acoustic experts in terrestrial wildlife that compared piezo mics with no mesh to capacitive microphones with mesh. The study focused on "assembling cheap, high performance MEMS microphones for recording terrestrial wildlife: the Sonitor system" [1] with the goal to build microphone configurations to record wildlife at a fraction of the cost of commercial microphones. The study quantified the acoustic mesh's tradeoff between weatherproofing and sound attenuation using a) Vesper mic without a mesh and b) Capacitive mic with two different mesh designs. GAW112 is only rated for IP4x but no guaranteed water ingress protection. GAW325 provides IP67 rating, perhaps something that is most suitable to use in a consumer electronics device with capacitive microphones for IP67 protection. Results from study on the microphone assemblies shown in Figure 7 below indicates performance degradation of 2-3 dB with GAW112 and 6-8 dB with the GAW325 mesh.

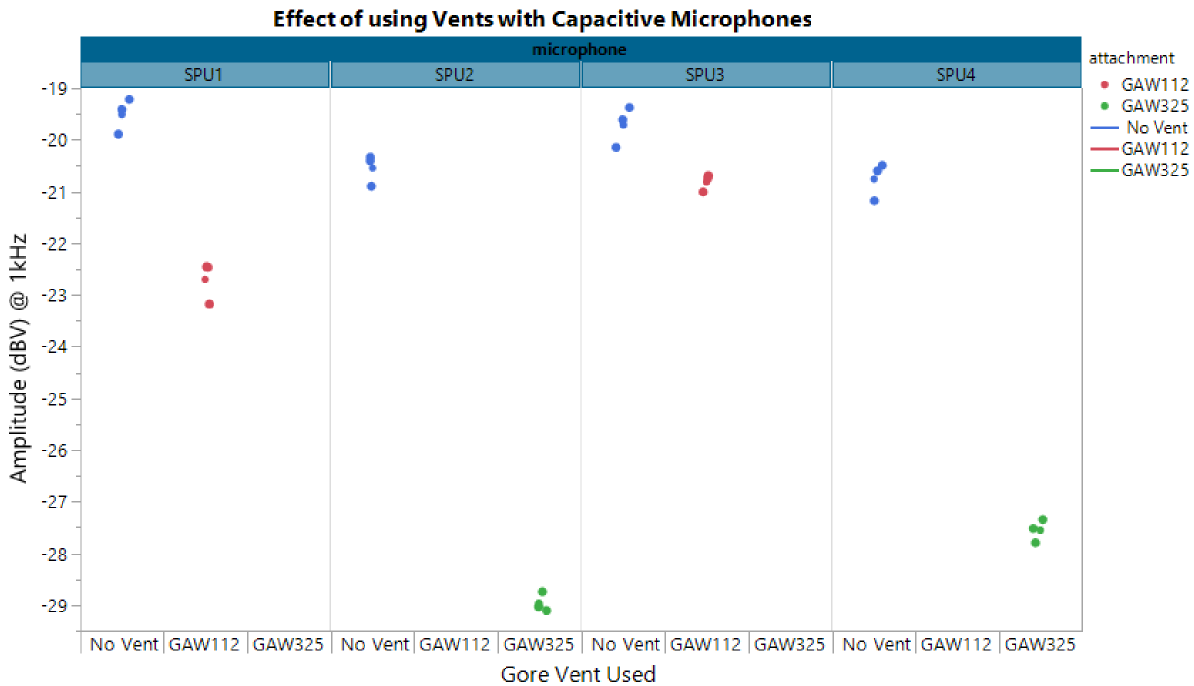

Figure 7: Effect of using Mesh with Capacitive Microphones. SPU1/SPU3 use GAW112, SPU2/SPU4 use GAW325 Mesh

Furthermore, all the three microphone configurations performed equally well under water and post immersion. Only difference is that Vesper microphones are directly exposed to water while capacitive microphones are covered with mesh during immersion.

Wake word detection performance

How does this startup time advantage translate into wake word detection performance? To investigate further, Vesper measured the wake word detection performance to evaluate the impact of startup time on False Rejection Rate (FRR). FRR is measured as the number of missed wake words out of all the 50 different wake word utterances spoken. Measurements are done inside an office-type room in a quiet environment with an ambient noise floor of 30 dBA. Head and Torso mouth simulator (HATS) is used to play back the utterances at different voice levels ranging from 75 dB-98 dBSPL. A 30 second interval is used between each utterance to ensure the complete system, including the VM1010 and VM3000, is switched to sleep mode after wakeup. A comparison of the FRR performance with VM3000 and capacitive microphone is shown in Figure 8.

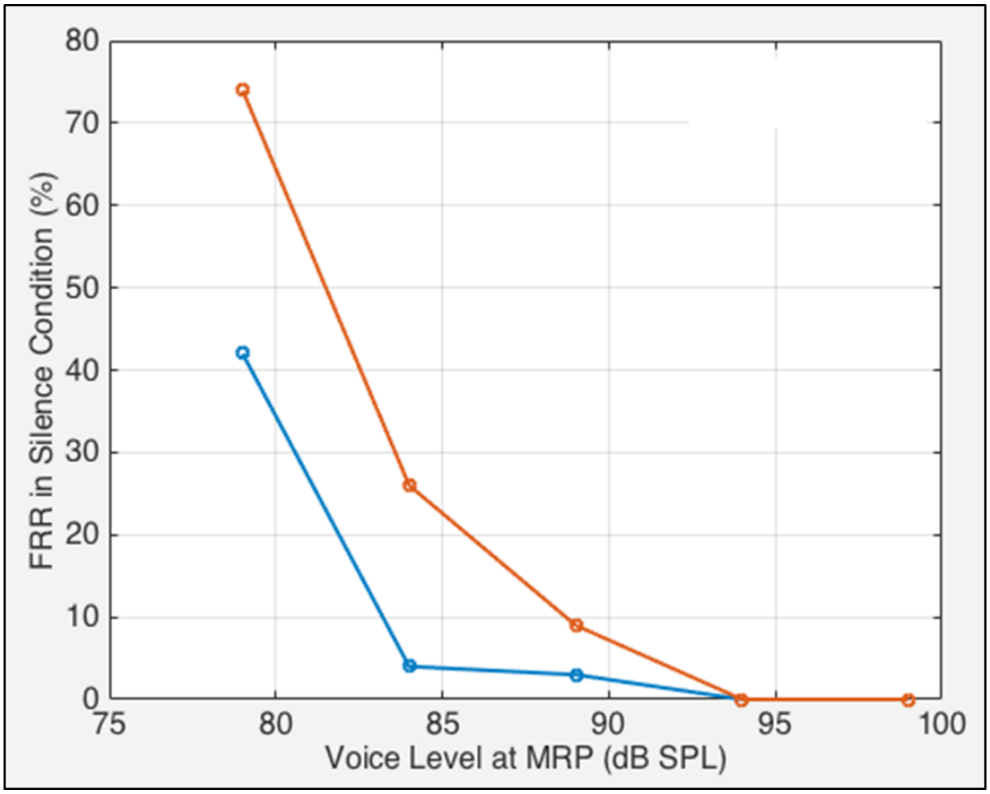

Figure 8: Wake word detection with Vesper (Blue) and Capacitive (Red) Microphones

At higher sound pressure levels, both microphones perform similar in FRR due to the high sound levels available at the microphone input. As the SPL levels decrease, the system with slow wakeup is penalized since the first syllable in the wake word is missed. The results indicate that VM3000 has up to 2x improvement in keyword accuracy compared to capacitive microphones, when used in combination with Vesper's ZeroPower Listening™ technology.

Complexity and Cost

Mesh or membrane adds to the overall cost of the system BOM. While the study cited above shows a unit price of more than a dollar for GAW325 mesh in 1K quantity, even a price one-fifth of that for volume production adds a significant cost to the assembly compared to the microphone component by itself. Below is a list of factors to consider while selecting components that require the protection of a mesh

- Cost of mesh can be equivalent or more than the cost of a microphone itself
- Higher assembly cost for a design with mesh
- Cost of additional testing or calibration required post mesh installation
- Added complexity and manufacturing yield issues with the mesh/membrane
- Impact on the long-term stability and durability of the design due to mesh contamination

Conclusion

While it's widely accepted that digital microphones offer better flexibility for designers in placement and routing of wiring around the codecs, Vesper's unique value proposition also provides additional design flexibility and unprecedented performance advantages. Evolution of the digital microphone portfolio at Vesper will be a great step forward towards the proliferation of robust and accurate voice activation devices that can operate even in harsh conditions.

For additional information on VM3000 and other Vesper products, please visit our website at www.vespermems.com or reach out to sales@vespermems.com.

Citations

[1] Darras K, Kolbrek B, Knorr A and Meyer V. Assembling cheap, high-performance microphones for recording terrestrial wildlife: the Sonitor system [version 1; peer review: 2 approved, 1 approved with reservations]. *F1000Research* 2018, 7:1984 (https://doi.org/10.12688/f1000research.17511.1)