Design Parameters for ZeroPower Listening™ using VM1010

Application Note
Rev1.0

Author: Udaynag Pisipati

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

AN4-Rev1.0
JAN2019

Table of Contents

Table of Figures

Introduction

This Application Note is intended to address the selection of design parameters to integrate Vesper's VM1010 Wake on Sound (WoS) microphone with ZeroPower Listening™ (ZPL™) technology into always-on always listening systems. The document combined with the Application note AN2 – "Using VM1010 Wake-on-Sound Microphone and ZeroPower Listening™ technology" provides necessary details to integrate VM1010 into any application. While the analysis provided in this application note is applicable to ZPL™ systems located in an indoor living room environment, the guidelines could also be used to interpret the design integration of VM1010 into any other application.

ZPL™ design parameters

ZPL™ applications can be broadly categorized into two sets –

1) Voice activated devices that wake up on user interaction and utterance of a specific wake word such as "Alexa", "OK Google" or "Hey Siri". Handsfree TV remotes, headsets, hearables etc. fall into this category.
2) Sound or event detection applications where VM1010 wakes up to a specific activity of interest such as a glass break at the front door, a baby cry or a smoke alarm activated in a living room. Examples of such products include security cameras, video doorbells, smoke alarms etc.

Battery life optimization on these applications is a result of careful selection of two ZPL™ design parameters - Acoustic threshold and Hold time. These two metrics provide a balance between device performance and battery life thereby minimizing false wakeups while preserving the wakeword detection accuracy of the device. Appropriate selection of

these metrics therefore requires careful analysis of speech and background noise levels encountered in the usecases under consideration.

Acoustic threshold

ZPL™ uses a peak detect mechanism to trigger the microphone to ambient sound levels between 65 – 89 dBSPL acoustic level called the Acoustic threshold. This 24-dB range of threshold provides developers with the flexibility to fine tune the listening level of the microphone for a wide variety of usecases. Selection of Acoustic threshold depends on the following factors –

a. predominant noise types and their sound pressure levels
b. Typical Sound level of the event to be detected
c. Typical speech levels encountered at the microphone of the device
d. Device location and proximity of the user to the device

A list of background noises common to several ZPL™ applications are provided in the table below. Careful analysis of these noise types and their sound levels captured at the real-life location of the ZPL™ device provides valuable insight into the appropriate acoustic threshold that gives optimum battery life for the device. Once a WoS Threshold is chosen, the metric can be configured on the device using an external resistor as shown in application note AN2.

TV Remotes	Indoor/outdoor Security systems	Headsets/hearables
TV playback	Baby cry	Subway
Microwave Oven	Dog bark	Driving windows open
Music playback	Window glass break	Driving windows closed
Conversational speech	Smoke/CO alarm	Driving HVAC on

Wakeword utterance	Door closing	Pub
	Doorbell ring	Wind noise
	Emergency responses	Running/walking/physical activity

Table 1: Common sound profiles encountered in ZPL™ applications

Hold time

Hold time is the time in which ZPL™ microphone switches back to listening mode from a normal mode after waking up to a sound activity. This metric must be programmed at the system level, so the DSP can monitor the activity in the background and put the device back to sleep if needed. A system with short hold time improves the battery life by putting the microphone in WoS mode more often. On the other hand, a system with long hold time could mean a shorter battery life since the microphone would already be out of the WoS mode when the keyword is spoken. The system would therefore be in normal mode longer than necessary wasting battery life.

Choice of hold time is also dependent on whether the ZPL™ device performs voice activation or event detection. Key selection criteria for hold time include the following –

a. Duration of the event including the wakeup time of the system
b. Length of voice utterance following the wakeword

For voice activation applications where the wake word is followed by a longer voice command, hold time is usually longer compared to event detection application. Therefore, a ZPL™ device with voice activation could have a hold time upto 30 seconds to accommodate the user command as well as the device response. However, a device that detects the sound of a smoke alarm could go back to WoS mode within 10 seconds or less, provided the DSP captures enough audio stream to analyze the sound.

The two ZPL™ parameters can also work in tandem to enable battery life savings in extreme cases and still provide battery life savings compared to alternate listening solutions. For example, in usecase where there is continuous speech activity in the background, choosing a longer hold time could penalize the system heavily for false wakeups. However, a combination of higher Acoustic threshold and short hold time will reduce the false alarms and significantly improve the battery life of the device. Figure 1 shows the relation between the two metrics and the battery life of the device. For a given hold time, say a value between 7-30 seconds, battery life of the device can be improved by increasing the Acoustic threshold to a higher value. In other words, a higher acoustic threshold of 71 dB can compensate the battery life degradation when the hold time is increased from 7 seconds to 30 seconds.

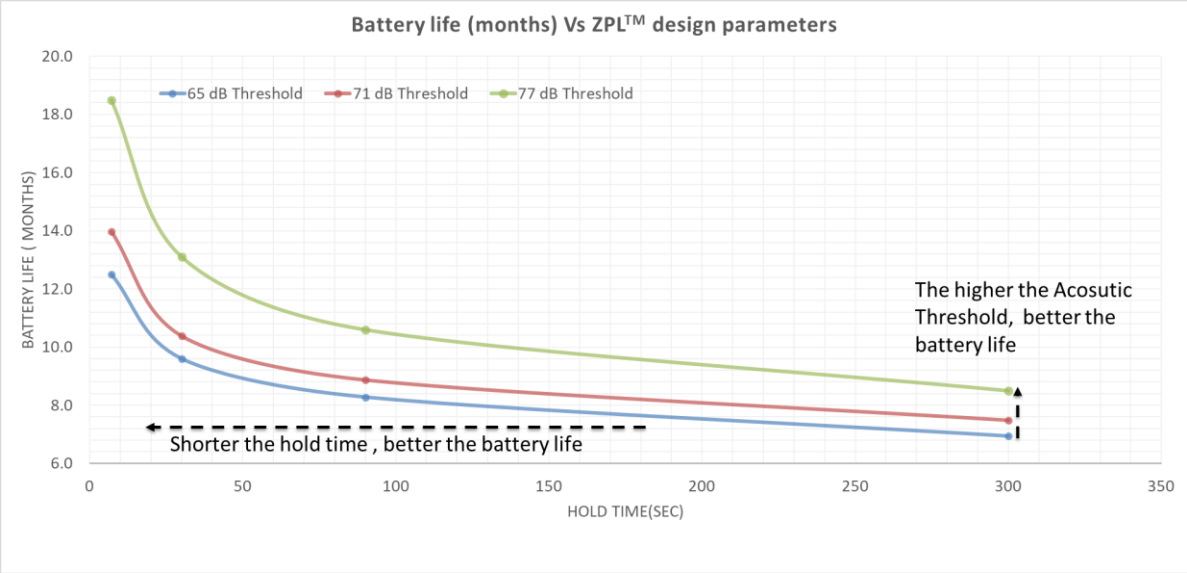

Figure 1: Battery life (months) Vs. ZPLTM design parameters

ZPL™ design parameters for living room applications

This section describes the selection of ZPL™ design parameters and their impact on WoS time and battery life performance for two different usecases of VM1010 microphone –

- Handsfree TV Voice Remote that is typically located on a coffee table in a living room 1-2 meters from the TV and
- Indoor Security camera looking at the front door in the living room

To facilitate the battery life performance study on the above usecases, Vesper designed data loggers with VM1010 microphone at three different threshold levels – 65 dB, 71 dB and 77 dB. These data loggers were used to record data in living rooms during a 24-hour period on a weekday as well as a weekend. 700 hours of data recorded from 24 different households is then used to analyze the frequency of wake on sound triggers in a living room environment with hold times varying form 7 seconds upto 5 minutes. Following subsections provide the results from the study along with a description of the design parameters.

Acoustic threshold

Figure 2 below shows the percentage normal mode time on a TV remote for various acoustic thresholds with the hold time set to 7 seconds. The minimum threshold of 65 dB provides close to 77% of WoS time (100 - % Normal mode time) at the TV remote location. As the acoustic threshold is increased, the amount of WoS time and hence the total time a device is in sleep mode increases, thereby improving the battery life. But how high or low an acoustic threshold should be selected to obtain a balance between accurate wake word detection and false positives?

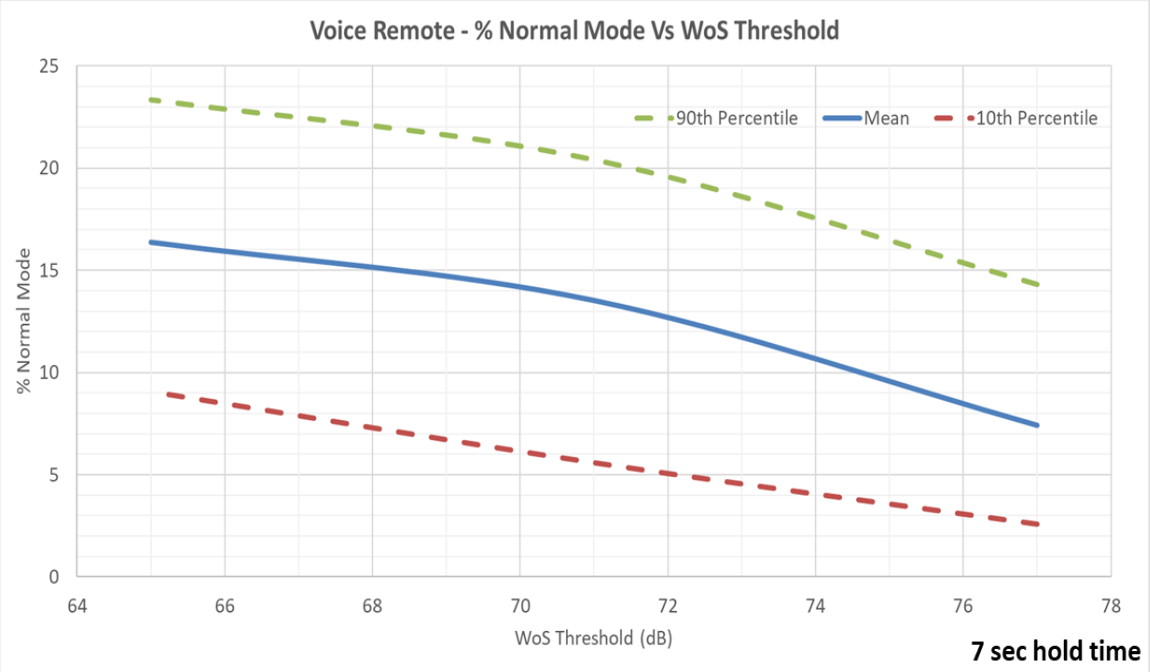

Figure 2: % Normal mode Vs. WoS Threshold for TV Voice Remote

A 65 dB Threshold on VM1010 is well above the noise level encountered in a quiet living room without the television powered on. On the other hand, usual conversational levels at the ZPL™ microphone on the coffee table would be in the low 70 dBSPL. Setting the acoustic threshold at a higher level might require the user to shout at the remote control for voice activation and therefore compromises user experience.

Our data in Figure 2 shows that the savings in normal mode time and hence the battery life savings are only a few percent at higher threshold, say, 71 dB compared to 65 dB minimum threshold setting. Therefore, for a handsfree TV remote, Acoustic threshold close to 65 dB level is a recommended setting that provides the tradeoff between false wakeups and optimal battery life. Even in the case where a TV is powered on and running at a nominal volume level, a 65-dB threshold could be optimal given the user is already speaking louder in a noisy environment according to the Lombard effect.

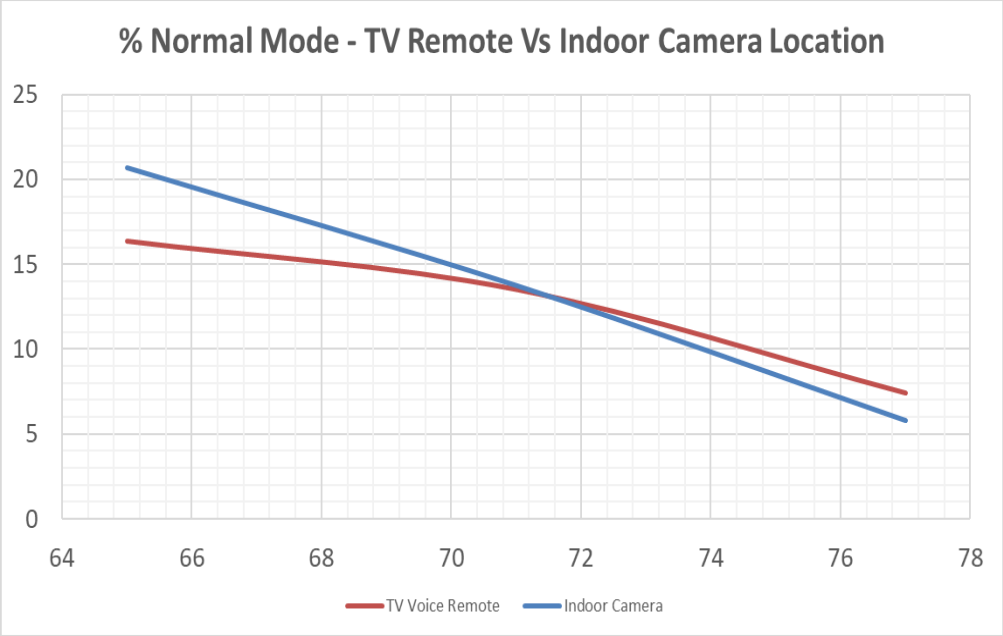

Figure 3: % Normal mode Vs. WoS Threshold for TV Voice Remote Vs. Indoor Camera

Data from indoor camera location in Figure 3 also shows similar performance characteristics as the handsfree TV remote scenario. However, for the thresholds between 65-71 dB, this usecase shows more normal mode time which could be attributed to the location of the device in a living room. Indoor camera usecase is generally looking for a background event rather than wake word. The threshold is therefore chosen based on the sound profile of interest. But, the typical sound level of these profiles such as smoke alarm, dog barks etc. are typically high. Therefore, VM1010 can be set to higher Acoustic thresholds such as 77 dB to optimize battery life. The selected threshold should still fall below the sound level of the event to be detected.

A comparison of battery life savings for the two different usecase is shown in Figure 4

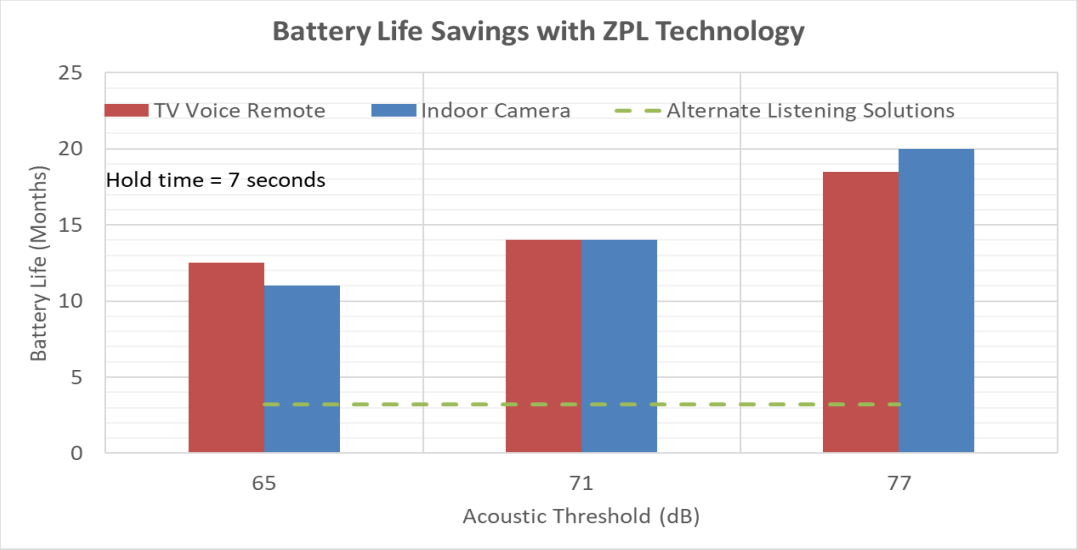

Figure 4: Battery life (months) Vs. WoS Threshold

Hold time

Figure 5 shows the amount of time VM1010 would be in normal mode at hold times from 7 seconds to 5 minutes for a given 65 dB acoustic threshold setting. 7 second hold time reduces the normal mode activity by half compared to 5-minute hold time. However, like the Acoustic threshold, selection of hold time also depends on the usecase. In a handsfree voice remote where the actual voice command such as "Play", "Pause", "Open House of Cards" combined with the wakeword is in the order of a few seconds, a hold time in the range 7 sec to 30 seconds is recommended. For a device with voice activation, the microphone should be in normal mode for at least the amount of time a required voice command would be uttered, but not too long that the system can consume additional resources. For a security camera application, hold times less than 10 seconds could be enough depending on the sound profile to be captured. Battery life analysis for the 2 usecases as a function of hold time is shown in Figure 6

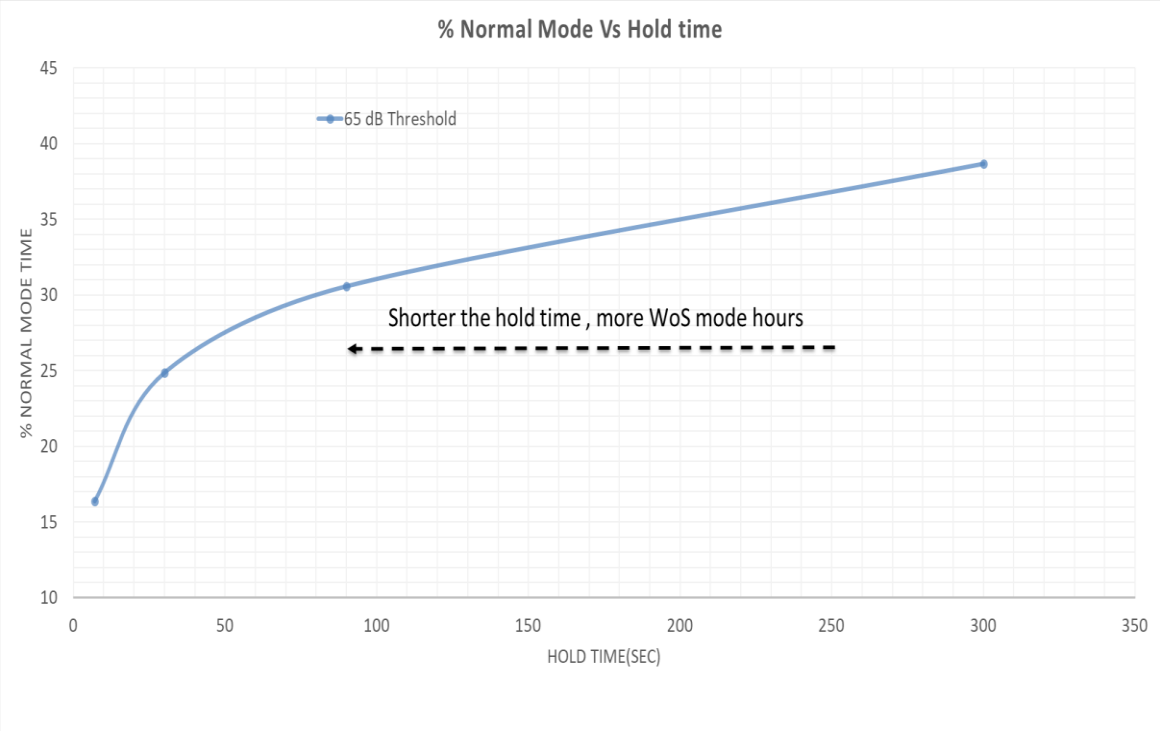

Figure 5: % Normal mode Vs. Hold time for 65 dB Acoustic threshold

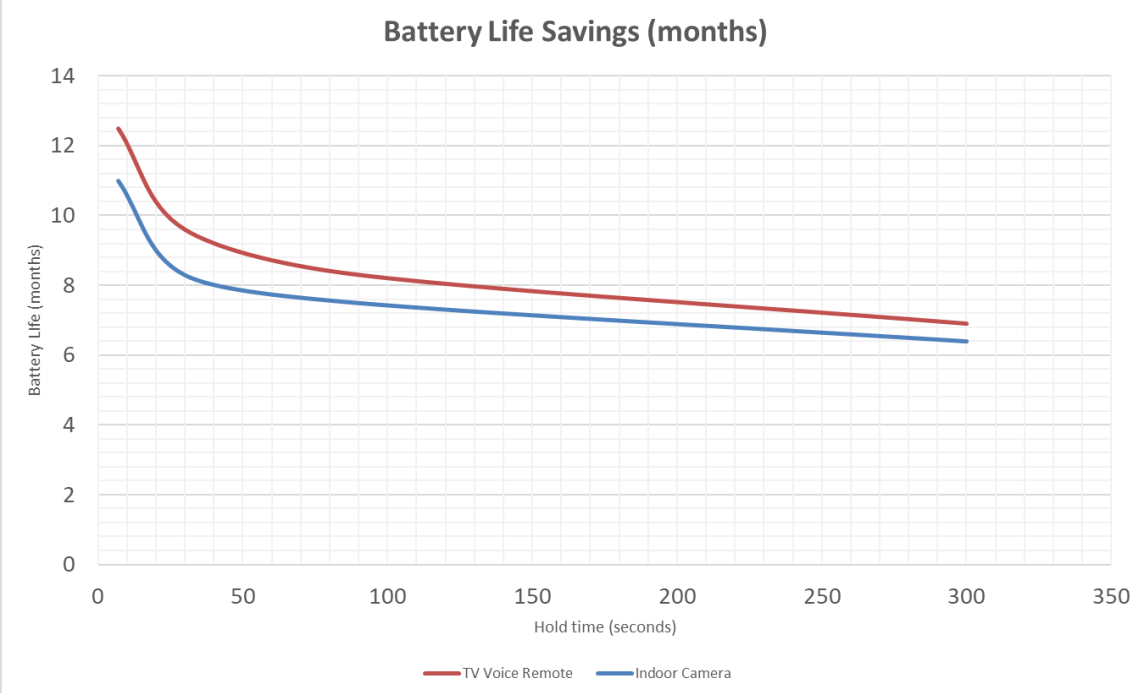

Figure 6: Battery life Vs. Hold time for 65 dB Acoustic threshold

Figure 7 summarizes the battery life savings from WoS microphone for a handsfree TV remote and indoor camera for the recommended settings on VM1010. ZPL™ provides 10x increase in standby battery life compared to alternate capacitive microphone-based solution. A 3x improvement in typical battery life of the device is evident for the minimum settings of 65 dB for an Acoustic threshold and 7 seconds for the hold time. Battery life as high as 7x compared to alternate listening solutions could be achieved by proper selection of design parameters. Above all, integration of VM1010 into any always listening system only requires two additional GPIO pins and an external resistor to set the acoustic threshold.

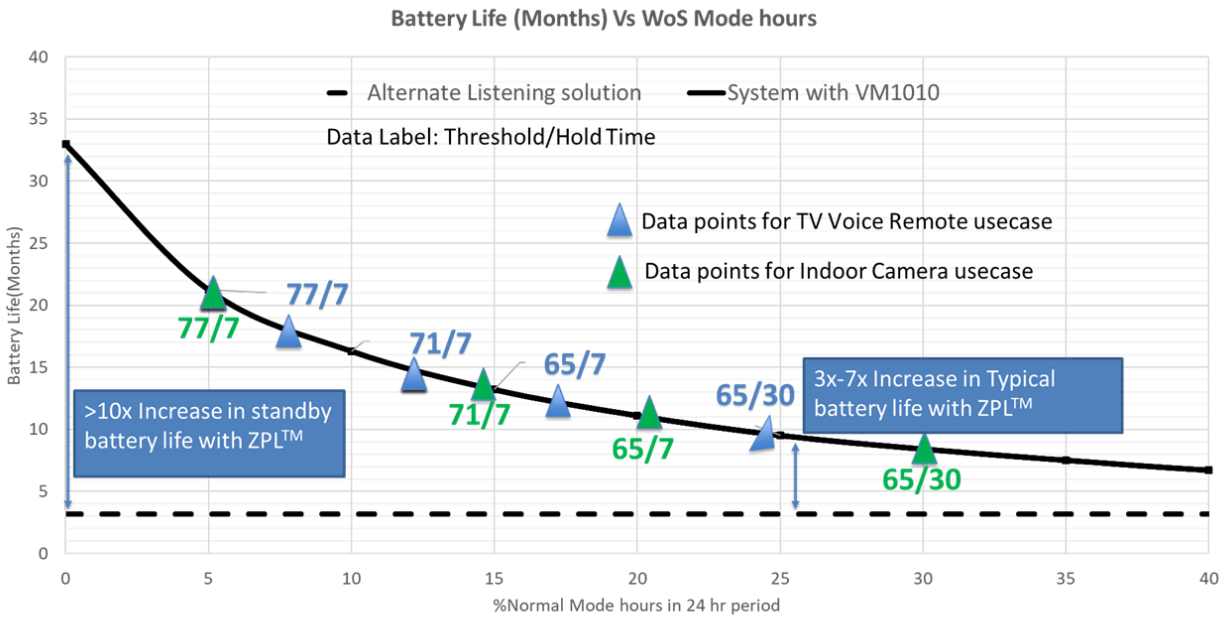

Figure 7: Living Room Usecases: Battery Life (months) vs. % Normal mode

ZeroPower Listening™ – Frequently Asked Questions

Does TV playback volume increase false positives?

One might possibly imagine that the TV playback itself can increase false wakeups on ZPL™ technology in a living room, thereby compensating for the power savings obtained with the technology. But our household data which includes false wakeups from active television time and other background activity in the living room indicates that the normal mode time falls only between 15-25% of the time in a living room. This is because VM1010 activates on a peak detect level instead of an RMS level. In addition, the average volume setting on the TV and the distance of the voice remote from the TV would automatically compensate for the peak sound pressure level at the WoS microphone of the ZPL™ device. In our experiment on sound levels for the various programming content on television for a 65 dB Acoustic threshold on VM1010, the average normal mode hours fall around 30% across the 4 different genres on television. This is shown in Figure 8 below.

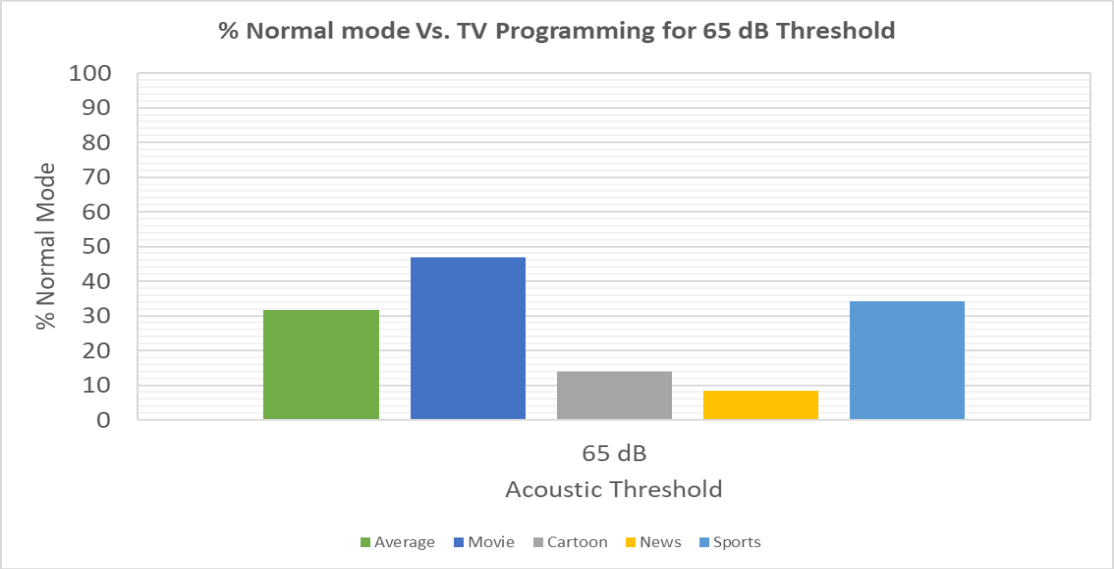

Figure 8: Effect of TV Programming content on WoS Microphone

What about other extreme Scenarios?

As with any energy harvesting technologies, power optimization with ZPL™ very well depends on the usecase and varies with extreme scenarios such as number of people, pets or kids in the household that can contribute to more wakeups. Figure 9 shows two such extreme scenarios from the 700 hours of data logged in households. Even with the extreme cases, ZPL™ still provides atleast 60% of WoS mode hours. While the design parameters may not necessarily be tuned for these extreme cases, it is obvious that there is a huge battery life advantage with ZPL™ compared to alternate listening solutions.

Figure 9: ZPLTM with Extreme scenarios

For additional details on datalogger study or to discuss your application, please contact info@vespermems.com