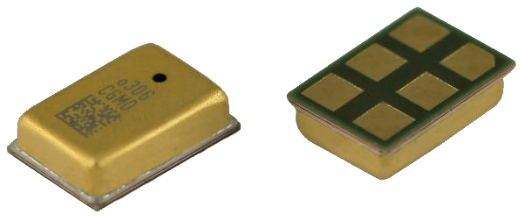
V2S Vibration Sensor: Speech pickup in consumer electronics applications through human body vibration. Application overview, integration and signal processing guidelines.



Nikolay Skovorodnikov, Senior Manager, Applications Engineering

Abstract

The V2S vibration sensor is a compact, low-noise device that offers full audible bandwidth frequency range. This document covers the primary application for the V2S Sensor: human body vibrations to enhance phone call clarity in noisy environments for wearable consumer electronics devices. It includes main design principles for using the V2S Sensor, mechanical and electrical integration guidelines as well as signal post-processing ideas.

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

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

Table of Contents

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

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

Introduction: Problem Definition and General Mechanism Idea

Phone call quality is a critical feature of all kinds of consumer electronic devices. Whether it is a mobile phone, wireless headset, or any other communication device – it is critical to pick up a clear intelligible speech signal in any environment. Some use cases are relatively easy to accomplish, such as a quiet conference room. However, modern consumer expectations are to stay connected while walking on a windy street or taking a call in a loud restaurant or busy traffic situation. Such loud environments are challenging. Conventional microphones pick up any audio input signal. This includes the target user speech signal as well as any other unwanted environmental noise. As a result, the person on the other end of the phone call will have hard time understanding the speech. We have all been there – trying to talk to someone and they interrupt us by saying, *"I can't understand the word you're saying, go find a quiet place and call me back!"* The Syntiant Voice Vibration Sensor (V2S) is designed to overcome this problem and help make communication more reliable.

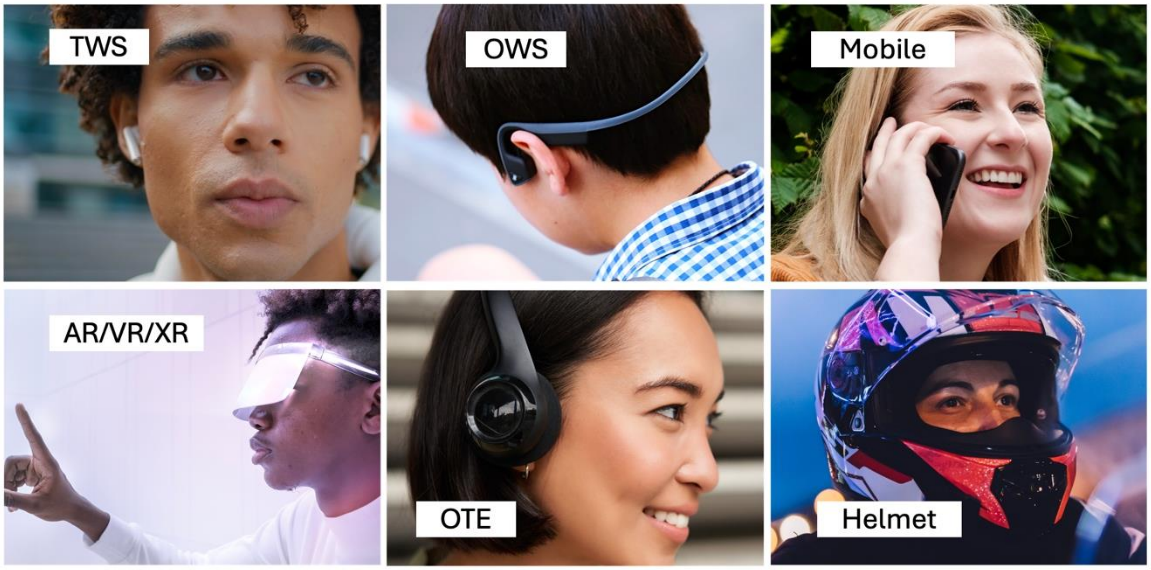
Clear phone call experiences can be achieved by taking advantage of the voice vibration of the human body. When people talk, the vibration signal associated with speech originates in the throat and mouth and propagates through the surrounding tissue. A wearable device in contact with the user's head will vibrate accordingly. The V2S vibration sensor integrated inside the device can record the target vibration signal. Unlike microphones that capture both useful speech signals and environmental noise simultaneously, the V2S vibration sensor will only pick up the target signal and is generally immune to environmental sounds.

Syntiant V2S vibration sensor has a unique set of specifications that make it a great fit for such a use case. It has small size and low power optimized for space- and battery-constrained wearable applications. High SNR help secure a clear signal even from small vibration input. Most importantly, it can pick up vibration signals up to 10kHz. Wide bandwidth separates the sensor from other solutions on the market. It is able to cover the entire frequency range of human speech for a clear phone call use case.

It is important to mention that there are multiple techniques used to clean up the signal in loud environments and improve phone call quality. Some common examples include multi-microphone algos such as beamforming or software postprocessing methods like noise suppression. Using the V2S sensor offers an alternative or additional raw signal that can be post-processed for further improvement.

The value of V2S voice vibration sensor extends beyond phone applications and provides low-frequency vibration information that can be substituted for a conventional accelerometer. Events like footsteps or basic gestures can be recognized. Additionally, the wide-bandwidth vibration sensor can also help authenticate the user. Signals in the vibration domain guarantees that the device is in physical contact with the source of the signal and is not remotely transmitted.

The V2S vibration sensor has demonstrated value across different types of consumer electronic devices. The most common market category is ear applications. Various form-factors of wireless earphones include in-ear devices (TWS or True Wireless Stereo) with open fit and closed fit (usually has rubber ear tip). There is also an emergent category of open-ear wireless buds (OWS). Larger form-factor on-ear headphones with a band over the head are also in this category (OTE). Another product class is everything related to smart glasses. This includes AR, VR headsets, smart glasses and connected helmets. These devices have a different balance of vibration signals compared to ear applications and need to be considered separately. Last but not least, mobile phones may also be placed in the body contact category for vibration sensor applications. The classic phone call use case, when a smartphone receiver is pressed against user's ear, can also provide a clear vibration signal and bring value to the user.



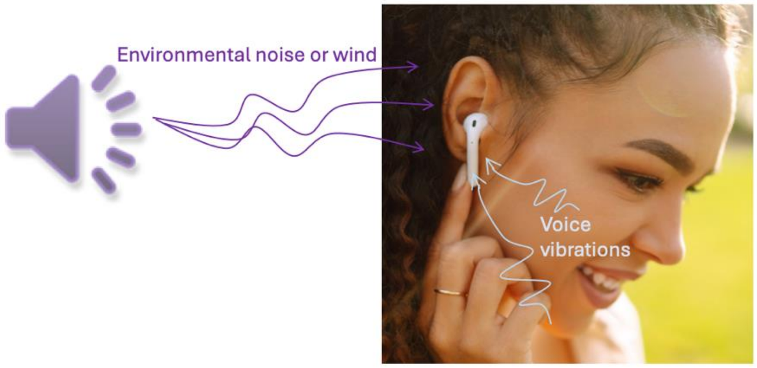
Categories of consumer electronics devices that can enable clear phone call feature by using V2S vibration sensor to pick up voice vibrations via body contact.

Beyond common consumer electronics use cases, one can consider using the high quality V2S vibration sensor to record useful information about the user's body. An example is an electronic stethoscope. Syntiant has demonstrated a prototype achieving a detailed heart sound recording of a heartbeat signal. State-of-the-art electronic stethoscopes today use quite complex mechanical designs based on an acoustic microphone. A design based on a V2S vibration sensor opens a path to a simpler construction with comparable signal quality.

Signal Balance for Successful V2S Application

In order to improve a phone call signal in a loud environment, a particular balance of signals needs to be achieved. Key integration principles will be discussed in this chapter. A general approach is common across all body-contact consumer applications but the details vary according to device type.

The overall principle is to have a high signal-to-noise ratio of wanted signal to unwanted noise. Vibration speech picked up by the V2S sensor is the target signal that needs to be maximized. External environmental acoustic signals will couple into the vibration domain. This unwanted noise signal path needs to be minimized. Additionally, there might be other sources of vibration that could introduce unwanted input signals, such as speaker playback, which can cause shaking, or noise generated from touching or handling the device.



Body contact voice vibration signal above vibration coupled from outside loud environment.

Signal Pickup

In this chapter, speech vibration coupling path is discussed. Design goal is to maximize this signal for the best use case results.

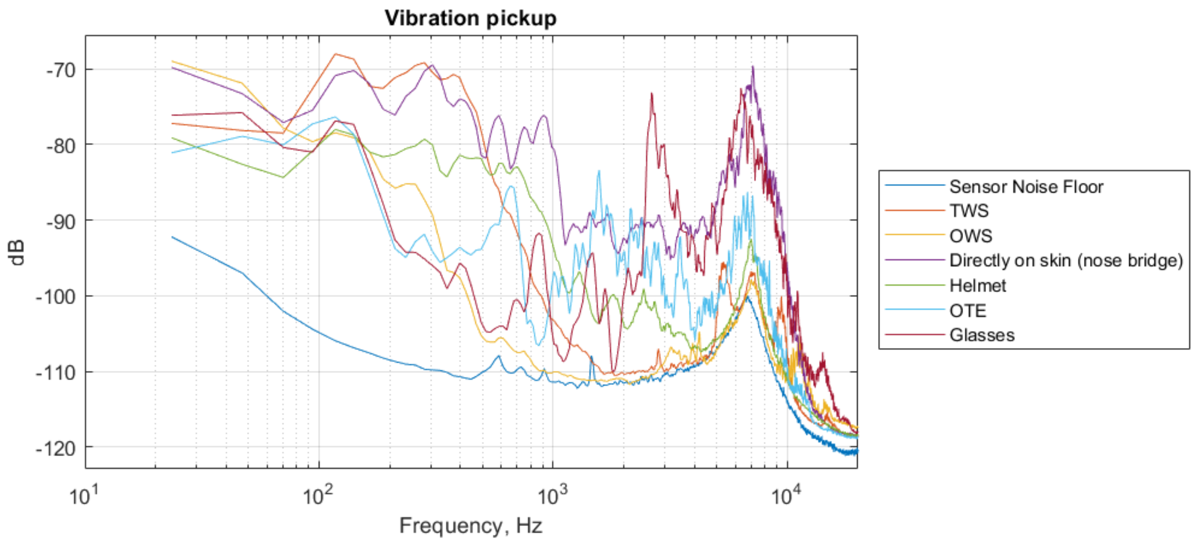
As mentioned, speech vibration signal originates in the throat and mouth. Specifically, vowel sounds come from the vocal cords and consonants are determined by the teeth, tongue and lips. It is important to mention that this is true for normal speech volume and shouting. Whispering has a different mechanism of producing sound and there is little to no vibration associated with it.

Speech vibration propagates through body tissue. Consumer devices with a V2S vibration sensor in contact with the skin surface can pick up that signal. Denser parts of the body conduct vibration much better. This why voice vibration sensors are sometimes referred to “bone sensors.” Hard/bony parts of the head or face will have a lot better vibration signal compared to soft tissue. It has been demonstrated that the V2S vibration sensor, when attached to the skin, can generate clear vibration signals that are comparable to microphone recordings within the full human audible bandwidth. This performance is achieved when the sensor is placed in areas such as the ear canal, forehead, or the bridge of the nose.

Next, speech vibration signal couples into the body of the consumer device. At this point, it is important to consider the difference between various device classes. From a physics perspective, the vibration signal is preserved in absence of impedance boundary. This means that a consumer device needs to be as hard and dense as human bone or cartilage. The target signal can be fully harvested if there is reliable string contact with human body and consumer device is made out of hard plastic. This tends to be rather uncomfortable for the user. Thus, there is a tradeoff between comfort and vibration pickup. In ear applications, a rubber tip on closed-fit TWS headphones is the source of high-frequency speech vibration attenuation. Harder rubber can be considered to capture more signal. For open fit TWS or OWS devices, it is recommended to have a secure and snug connection with the body as close to the skull as possible. For face-worn devices like smart glasses, it is easier to capture full audible bandwidth vibration signal as long as there is not too much cushioning/soft/elastic material which would dampen the vibrations. Image below shows a vibration signal pickup in various form factors.

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

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



Speech pickup using V2S sensor on various consumer devices.

The position of the V2S sensor inside a consumer device needs to be carefully selected. The V2S voice vibration sensor is a wide-bandwidth single axis accelerometer which is only sensitive to the vibration in the direction orthogonal to the PCB base. It is recommended to experiment with various locations inside consumer devices to determine the most optimal placement. Generic guidance such as when people talk, most of the vibration occurs in the left-right direction, along the axis connecting the ears. Placing the sensor orthogonal to the ear axis is expected to give the best results.

As long as construction of the consumer device is solid, it is expected to vibrate as a single piece. Thus, once orientation of the V2S sensor is determined, exact positioning is not critical. Usually, placing the V2S sensor on the main circuit board alongside other components is sufficient and does not require dedicated flex circuits. It was demonstrated that having a V2S sensor placed in the middle of a consumer device as opposed to just below the surface leads to some vibration response differences in high frequencies, although it is not expected to have a major impact on speech pickup overall quality.

Sound Isolation

Subject of this chapter is the coupling of unwanted vibration signals from environmental sounds. Design target is to minimize such signal path. It would lead to the best results for the use case of clear phone call in loud environments.

The V2S vibration sensor operation is based on a vacuum-sealed MEMS element. The sensor itself is perfectly isolated from external acoustic sounds. As long as the sensor does not move, it will have no output no matter how loud the environment. This was demonstrated on a special bench engineered to minimize vibrations. Datasheet specifications for acoustic isolation of V2S sensor itself is based on such experiments.

Real-life applications of the vibration sensor in consumer electronic devices require deeper analysis. Sound propagates through the air in form of pressure waves. Impinging an acoustic wave can create vibration when it falls onto an object. If a wearable device experiences vibrations under the impact of an external sound field, such vibration will be detected by the V2S sensor. Thus, sound isolation is not a challenge for sensor improvement, rather a system-level mechanical optimization to minimize vibrations.

Wind noise needs to be discussed separately. Turbulent in nature, wind can have a devastating impact on the output signal of a conventional microphone. The vibration sensor can bring much better results in similar conditions. Some low-frequency vibration induced by wind may come through but gust turbulent artifacts are not present in the vibration domain.

Sound-induced vibration of consumer electronic devices was evaluated. The mechanism is complex and highly dependent on the details of mechanical design such as weight, sound direction and shape of the device. Size of the device is the dominant factor for picking up external environmental sounds. To illustrate this coupling path, one can think of the consumer device as a diaphragm of a large microphone. A larger form-factor will do a better job converting sound into vibration. Smaller devices will be less susceptible to external sounds. Frequency-domain consideration is also possible. Low frequency audio with large wavelength can effectively induce vibration on larger devices. Smaller devices lose vibration sensitivity to low-frequency audio while preserving high-frequency susceptibility. Small size wireless earphones like TWS and OWS provide great insulation in low- and mid-audio range and start picking up sound-induced vibrations in high frequencies. Larger devices like smart glasses and OTE headphones pick up sexternal sound across entire audio spectrum.

Feasibility of V2S vibration sensor application for clear phone call use depends on signal-to-noise ratio in target frequency range. As long as there is a more useful vibration signal from the user’s speech conducted through the body compared to environment sound coupled into the vibration domain due to device shaking, additional value can be given to the consumer. Fine balance between such signals needs to be evaluated individually given exact mechanical properties of the wearable device and external sound details.

Speaker Cross Talk and Other Considerations

In addition to maximizing user voice pickup and minimizing environmental sound pickup, there are other important considerations during V2S sensor integration. Every device described above has speakers. Moving coil inside the speaker creates vibration that can be registered by the V2S sensor. Such crosstalk coupling needs to be minimized for best results in loud environments.

Vibrational speaker crosstalk can be mitigated by optimized mechanical placement. Ideally, speaker vibration needs to happen orthogonally to V2S sensor pickup axis. This way vibration can be minimized. Also, separating speaker from sensor module with any sift material can help attenuate unwanted vibrations. Depending on exact mechanical design of the consumer device there might be more or less flexibility for V2S sensor placement with respect to the speaker.

Speaker signal crosstalk is not a new problem for voice applications. Very similar phenomena takes place in the audio domain. Microphones pick up the sound of speaker playback. If not mitigated, users start hearing their own voice during the phone call, phenomenon often described as an echo. There are software solutions for mitigation. Echo cancelling algorithms rely on the fact that speaker playback signal is known inside the device. The algorithm analyses the signal from the microphone or V2S vibration sensor and removes the features that match with the speaker playback signal. Only parts of the signal that are not present in the playback are sent via the uplink channel.

The combination of mechanical isolation from the speaker and the echo-cancelling algo can fix the problem of speaker vibration crosstalk for the V2S sensor.

Sensor Integration

Electrical integration

The V2S sensor output format is Pulse Density Modulation (PDM). This is similar to the widely adopted digital MEMS microphone output format. Electrical integration of the V2S sensor is equivalent to a PDM microphone. The V2S sensor requires a PDM input on an audio CODEC. Any chip made for digital mic interfaces would also work with a V2S sensor. PDM interface consists of two lines: CLK and DATA. Similar to a MEMS microphone, V2S also has a pin which can be used to select one of two available channels on a PDM bus. V2S sensors can share a PDM bus with another V2S sensor or digital MEMS microphone.

Similar to PDM microphones, electrical integration on a given circuit board can use passive components like resistors or chokes to protect from RF interference. Also, there is a limit to the amount of DATA line capacitance V2S sensor can drive. Particularly large devices with long electrical traces inside might need to use buffers to preserve PDM signal waveform quality.

Barometric Relief Hole

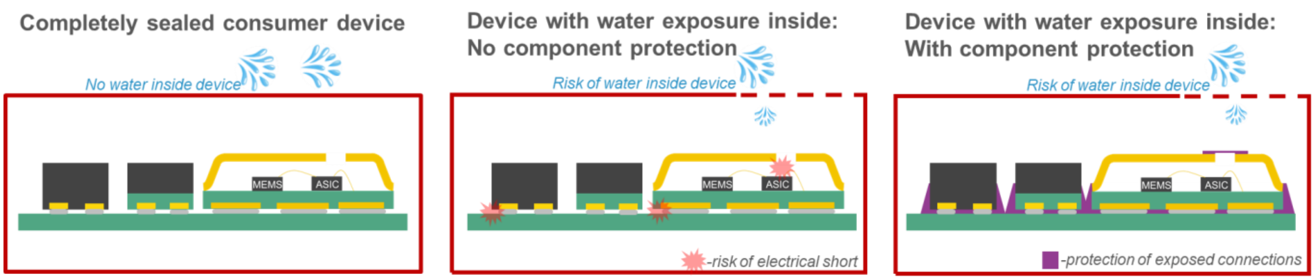
The V2S sensor is a SMT component with MEMS element and ASIC chip. Sensor packaging includes base PCB and a metal can. The can of the sensor has a small round hole. Unlike MEMS microphones, this is not an acoustic port. MEMS element of V2S sensor is completely sealed, sensitive to acceleration and does not require access to the outside to properly operate. However, the space under the metal can is filled with air. During the reflow process, the temperature can reach 260 degrees Celsius. A barometric relief hole is required to avoid pressure build up in a sealed package. Thus, the sole purpose of a hole is to survive the reflow process. The hole must be open when the sensor is undergoing the high temperature reflow process. After the sensor was attached to the PCB, there is no impact to performance of the sensor weather the barometric relief hole is open or closed. It is recommended to have the hole covered if there is risk of water intrusion. See waterproofing chapter for more details.

Waterproofing

Robustness against environmental contaminants is a major advantage of the V2S vibration sensor over conventional MEMS microphones. Unlike an audio mic, it does not need a sound port open to the outside world. At the same time, details need to be clarified.

The V2S vibration sensor should be treated just like any other SMT electronic component with exposed electrical pads (i.e., passive resistor). A conductive particle or dirty water with ions present can potentially create and electrical short and damage the circuit. Exposure to deionized (DI) water is acceptable. V2S sensor has two potentially vulnerable areas: 1) external electrical reflow pads; and 2) internal bond wires that could be accessible through the barometric relief hole.

Imagine V2S sensor is placed inside the system on a PCB or flex alongside other electrical components. If the external case of the consumer device is waterproof and there is no risk of water getting inside, then no additional waterproofing is needed for the V2S sensor. However, if the sensor is potentially exposed to the outside (i.e., through the speak output nozzle of a TWS device), then special mitigation is needed similar to any other SMT electrical component. Epoxy is recommended on the PCB for protection of the mounting pads and membrane is needed to close the barometric relief hole.



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

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

SYNTIANT®

Post-processing

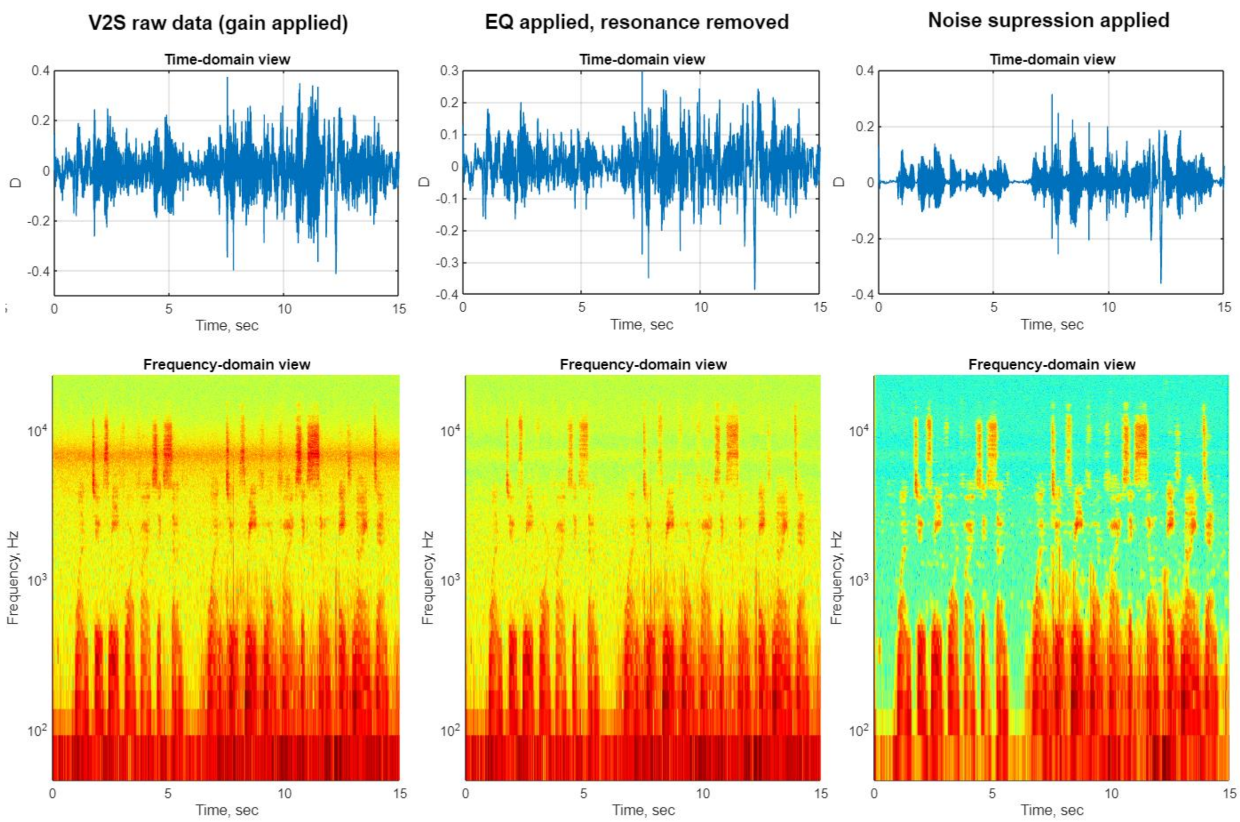
There are many options for creative signal processing with a V2S vibration sensor. Full audible signal pickup bandwidth gives an opportunity to potentially bring microphone-like experiences to the user with a clean signal in loud environments.

Depending on device design, it is important to identify frequency range where user speech signal coupled via body vibration is higher than externally coupled environmental sounds. This becomes the working range for the V2S sensor. Devices with strong body contact such as helmets tend to cover entire audible frequency range. Devices with softer contact such as TWS headphones can only offer a vibration range with signals above the noise only in low and mid-range frequencies. The V2S sensor signal can be used by itself or combined with a microphone signal. A common technique for wind noise mitigation is to use vibration data below approximately 1kHz and use high frequency data from the microphone on TWS headphones. As a result, wind noise artifacts are removed since they are only concentrated in the low-frequency range. Simultaneously, full range audio fidelity is achieved with high-frequency microphone data.

The V2S sensor has a resonance at approximately 7kHz. This feature is defined by mechanical design of the MEMS sensing spring-mass oscillator. Additional sensitivity in frequency range around the resonance can be unpleasant listening, while the treble is overrepresented. However, noise density of the sensor follows similar spectral shape as a magnitude response. Thus, equalization spectral mask can be applied to compensate for the resonance. The same attenuation would act on noise and magnitude. Signal-to-noise ratio in each frequency bin would be preserved while response of the sensor becomes flat and provides more natural listening.

In some cases, one can face a need to reduce the noise floor of the V2S sensor. There are some tricks to improve the specs beyond the datasheet values. Multiple V2S sensors can be used. If placed right next to each other, sensors will have identical input vibration signal. A simple summation algorithm would provide improved signal-to-noise ratio. Identical input signals will assure constructive interference and algorithm output would be proportional to the number of sensors. At the same time, noise is completely uncorrelated, which makes noise power proportional to the number of sensors. Mathematically, total SNR improves by 3dB with every doubling of the number of sensors.

Digital signal processing techniques can be used to improve the signal-to-noise ratio in the V2S channel. If sensor self-noise is noticeable in the recording, it is recommended to consider a noise suppression algorithm. There are several variations of this method, but the underlying principle remains consistent: the algorithm evaluates each segment of the signal in both the time and frequency domains to determine whether it represents the target useful signal or unwanted noise. Noise bins are attenuated, and overall output signal is cleaned up. Noise suppression works very well with the V2S sensor because sensor noise floor is known and very stable. The algorithm decision process is easier compared to guessing environmental noise sounds.



Example of V2S post-processing on OWS prototype recording. Raw data, EQ applied for resonance compensation, noise suppression applied.